Université Paris I, Panthéon - Sorbonne

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

Statistiques II

Contrôle continu n°2, avril 2014

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

1. **(Sur 10 points)** Soit $(\delta_n)_{n\in\mathbf{N}}$ une chaîne de Markov prenant ses valeurs dans $\{0,1\}$, telle que pour tout $n \in \mathbf{N}$, $\mathbb{P}(\delta_{n+1} = 1 \mid \delta_n = 1) = \mathbb{P}(\delta_{n+1} = 0 \mid \delta_n = 0) = \theta$ avec $\theta \in [0,1]$.

   (a) Donner la matrice de transition de $(\delta_n)_{n\in\mathbf{N}}$ **(0.5pts)**. Est-ce une chaîne homogène **(0.5pts)**?

   (b) Déterminer les valeurs de θ pour lesquelles (δ_n) est irréductible, et celles pour lesquelles elle ne l'est pas **(1pt)**.

   (c) Déterminer les mesures invariantes de (δ_n) suivant θ **(1pt)**.

   (d) On suppose que la loi de δ_0 est $\mathbb{P}(\delta_0 = 1) = p_0$ avec $p_0 \in [0,1]$. Montrer que $\mathbb{P}(\delta_n = 0) = \frac{1}{2}\big(1 + (2\theta - 1)^n(2p_0 - 1)\big)$ pour tout $n \in \mathbf{N}$ **(2pts)**. En déduire l'espérance et la variance de δ_n **(1pt)**. A quelles conditions sur p_0 et θ la suite (δ_n) est-elle stationnaire **(1pt)**?

   (e) Suivant les valeurs de p_0 et θ, déterminer la loi limite de δ_n quand $n \to \infty$ **(1.5pts)**.

   (f) Si $\theta = p_0 = 1/2$, montrer que $(\delta_n)_{n\in\mathbf{N}}$ est une suite de variables aléatoires indépendantes identiquement distribuées suivant une loi de Bernoulli dont on précisera le paramètre **(1.5pts)**.

2. **(Sur 10 points)** Soit $\varepsilon = (\varepsilon_t)_{t\in\mathbf{Z}}$ un bruit blanc fort centré de variance $\sigma^2 > 0$, indépendant de $(\delta_n)_{n\in\mathbf{N}}$. On définit le processus $X = (X_n)_{n\in\mathbf{N}}$ tel que

$$X_n = \delta_n(\varepsilon_n - a\varepsilon_{n-2}) + (1-\delta_n)\varepsilon_n \qquad \text{pour} \quad n \in \mathbf{N}.$$

   avec $a \in \mathbf{R}$, où (δ_n) est la chaîne de Markov précèdente.

   (a) Lorsque $\theta = 1$ et $p_0 = 1$, montrer que pour tout $n \in \mathbf{N}$, $X_n = \varepsilon_n - a\varepsilon_{n-2}$ **(0.5pts)**. Quel processus est alors (X_n) **(0.5pts)**? Déterminer son espérance **(0.5pts)** et sa fonction d'autocovariance **(1pt)**.

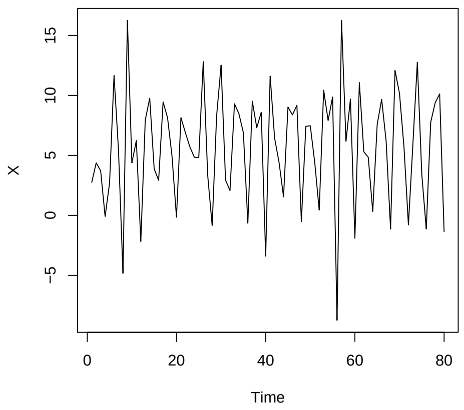
   (b) On considère maintenant le cas general $\theta \in]0,1/2[\cup]1/2,1[$ avec $p_0 = 1/2$ ((δ_n) est alors stationnaire). Montrer que (X_n) est stationnaire **(2pts)**. Déterminer l'espérance **(0.5pts)** et l'autocovariance de (X_n) **(1.5pts)**. Montrer que $\mathbb{E}\big[X_n^2 X_{n+3}^2\big] \neq \mathbb{E}\big[X_n^2\big]\mathbb{E}\big[X_{n+3}^2\big]$ **(2.5pts)**. En déduire que (X_n) n'est pas un processus MA(2) **(1pt)**.

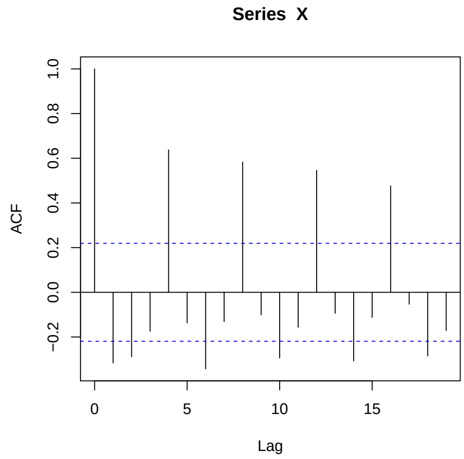
3. **(Sur 8 points)** Voici des simulations effectuées avec le logiciel R.

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

```
eps=2*rnorm(81)
X=6-6/c(1:80)+rep(c(3,4,0,-7),20)+eps[2:81]-0.8*eps[1:80]
ts.plot(X)
acf(X)
```

   Voici les deux graphes obtenus:





Questions: Quel est le processus simulé par le vecteur X (écrire le processus formellement en détaillant ses éventuelles tendance et saisonnalité) **(1pt)***? Que peut-on déduire de la commande* `ACF` **(0.5pts)***?*

(b) Voici les commandes tapées ensuite:

```
Z1=c(1:80)
Z2=rep(c(1,0,0,-1),20)
Z3=rep(c(0,1,0,-1),20)
Z4=rep(c(0,0,1,-1),20)
reg=lm(X~Z1+Z2+Z3+Z4)
summary(reg)
```

Voici les résultats obtenus:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.88262    0.64595   7.559 8.21e-11 ***
Z1           0.01901    0.01386   1.372    0.174
Z2           3.29274    0.55407   5.943 8.22e-08 ***
Z3           2.85634    0.55372   5.158 1.97e-06 ***
Z4           0.31241    0.55372   0.564    0.574

Residual standard error: 2.859 on 75 degrees of freedom
Multiple R-squared: 0.6656,     Adjusted R-squared: 0.6478
F-statistic: 37.32 on 4 and 75 DF,  p-value: < 2.2e-16
```

Questions: Qu'a-t-on fait par ces commandes **(1pt)***? Expliquer pourquoi la $p-value$ associée à Z1 est importante* **(0.5pts)**.

(c) Voici les commandes tapées ensuite:

```
Z11=1/c(1:80)
reg=lm(X~Z11+Z2+Z3+Z4)
summary(reg)
```

Voici les résultats obtenus:

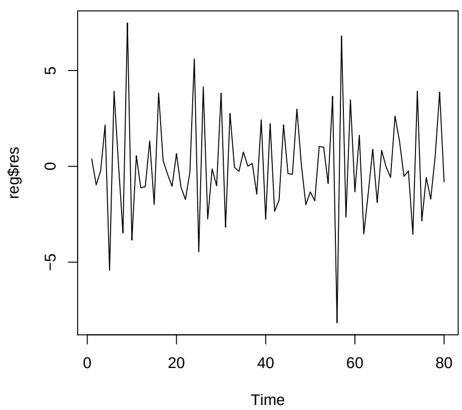
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.0998     0.3401  17.936  < 2e-16 ***
Z11          -7.2057     2.3984  -3.004  0.00362 **
Z2            3.4665     0.5339   6.493 8.15e-09 ***
Z3            2.8463     0.5296   5.374 8.35e-07 ***
Z4            0.2434     0.5303   0.459  0.64751

Residual standard error: 2.735 on 75 degrees of freedom
Multiple R-squared: 0.694,      Adjusted R-squared: 0.6777
F-statistic: 42.53 on 4 and 75 DF,  p-value: < 2.2e-16
```

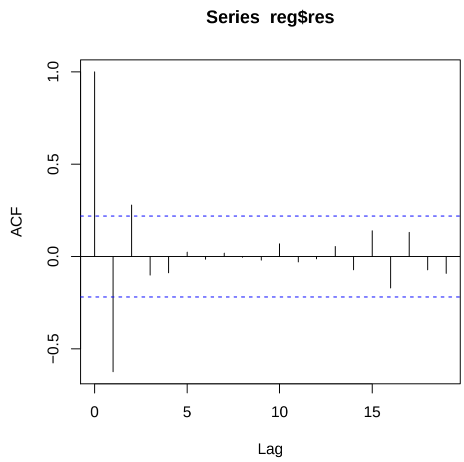
Questions: Qu'a-t-on fait par ces commandes par rapport à celles qui précèdent et a-t-on gagné quelque chose **(0.5pts)** *? Que représentent formellement les valeurs* 6.0998, −7.2057, 3.4665 **(0.5pts)** *? S'attendait-on approximativement à ces valeurs et pourquoi* **(0.5pts)** *? Quelle sont la tendance et la saisonnalité estimées* **(0.5pts)** *? Que peut-on conclure quant au modèle utilisé pour expliquer* (X_t) **(0.5pts)** *?*

(d) On tape enfin les commandes suivantes:

```
ts.plot(reg$res)
acf(reg$res)
```

Voici les deux graphes obtenus:





Questions: Que conclure de l'ACF? Les valeurs numériques des deux premières barres sont 1 *et* −0.64*. Pouvait-on s'attendre à cela numériquement* **(1pt)** *?*

(e) *Questions: Donner le code pour simuler une trajectoire de taille* 100 *du processus ARMA*$(2,1)$*:* $X_t - 0.3X_{t-1} - 0.1X_{t-2} = \varepsilon_t + 2\varepsilon_{t-1}$*, avec un bruit* (ε_t) *qui suit une loi normale de variance* 9 *(on prendra soin de vérifier auparavant que l'on a bien un ARMA causal...)* **(1.5pts)**