Université Paris I, Panthéon - Sorbonne

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

Statistiques II

Correction de l'examen final, mai 2012

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

1. **(18 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
$$X_{n+1}-\alpha X_n=\varepsilon_{n+1}-2\alpha\varepsilon_n \quad \text{pour} \quad n\in\mathbf{Z},$$
avec $|\alpha|<1$ un réel inconnu.

   (a) Quel processus est X? Est-il centré? stationnaire? gaussien? (justifier) **(2 pts)**

   (b) Montrer que pour tout $n\in\mathbf{Z}$,
$$X_n=\varepsilon_n-\sum_{i=1}^{\infty}\alpha^i\,\varepsilon_{n-i}. \qquad \textbf{(2 pts)}$$

   (c) Déduire l'expression de l'autocovariance $r_X(k)$ de X pour tout $k\in\mathbf{Z}$ (on traitera à part le cas $k=0$) **(2.5 pts)**. Montrer que pour tout $p\in\mathbf{N}$, $\lim_{n\to\infty}n^p\,r(n)=0$ **(0.5 pts)**. Que se passe-t-il lorsque $\alpha=\pm\frac{1}{\sqrt{2}}$ **(0.5 pts)**?

   (d) Déterminer l'expression de la densité spectrale f de X **(1 pt)**. Determiner la loi asymptotique de $\sqrt{n}\,\overline{X}_n$ lorsque $n\to\infty$, où $\overline{X}_n:=\frac{1}{n}\big(X_1+\cdots+X_n\big)$ **(1 pt)**.

   (e) Soit $(X_1,\ldots,X_N)$ une trajectoire observée de X. Déduire de ce qui précède la limite en probabilité de $N\big(\overline{X}_N\big)^2$ quand $N\to\infty$ **(1.5 pts)**.

   (f) Soit $\widehat{\sigma}_N^2:=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}X_i^2$ et $\widehat{r}_N(1):=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N-1}X_iX_{i+1}$. Montrer que le vecteur $\big(\widehat{\sigma}_N^2,\widehat{r}_N(1)\big)$ suit un théorème de la limite centrale (on écrira la matrice de covariance asymptotique à l'aide de f et sans calculer les intégrales) **(1 pt)**.

   (g) Déduire de ce qui précède un estimateur $(\widehat{\alpha},\widehat{\sigma}_\varepsilon^2)$ de $(\alpha,\sigma_\varepsilon^2)$ **(1 pt)** et expliquer pourquoi cet estimateur est convergent **(1 pt)**.

   (h) On veut obtenir une prédiction $\widehat{X}_{N+1}$ de X_{N+1}, la trajectoire $(X_1,\ldots,X_N)$ étant connue. Sous une condition portant sur α que l'on précisera, donner l'expression de ε_N en fonction de $(X_{N-k})_{k\in\mathbf{N}}$ **(2 pts)**. En déduire une expression d'un prédicteur $\widehat{X}_{N+1}$ obtenu uniquement à partir de $(X_1,\ldots,X_N)$ **(2 pts)**.

Proof. (a) X est un ARMA$(1,1)$, centré, stationnaire (car $|\alpha|<1$) et gaussien.
(b) $P(B)^{-1}=\sum_{i=0}^{\infty}\alpha^iB^i$ et $Q(B)=1-2\alpha B$, d'où $P(B)^{-1}Q(B)=I-\sum_{i=1}^{\infty}\alpha^iB^i$, d'où le résultat.
(c) On a $r_X(0)=\mathbb{E}\big[\big(\varepsilon_n-\sum_{i=1}^{\infty}\alpha^i\,\varepsilon_{n-i}\big)\big(\varepsilon_n-\sum_{j=1}^{\infty}\alpha^j\,\varepsilon_{n-j}\big)\big]=\sigma_\varepsilon^2(1+\sum_{i=1}^{\infty}\alpha^{2i})=(1-\alpha^2)^{-1}\sigma_\varepsilon^2$.
Pour $k>0$, $r_X(k)=\mathbb{E}\big[\big(\varepsilon_n-\sum_{i=1}^{\infty}\alpha^i\,\varepsilon_{n-i}\big)\big(\varepsilon_{n+k}-\sum_{j=1}^{\infty}\alpha^j\,\varepsilon_{n+k-j}\big)\big]=\sigma_\varepsilon^2(-\alpha^k+\sum_{i=1}^{\infty}\alpha^i\alpha^{i+k})=\alpha^k\sigma_\varepsilon^2(2\alpha^2-1)(1-\alpha^2)^{-1}$.
Comme r_X est une fonction paire, on a $r_X(|k|)=\alpha^k\sigma_\varepsilon^2(2\alpha^2-1)(1-\alpha^2)^{-1}$.
On a $\big|n^p\alpha^n\big|=e^{p\log n+n\log|\alpha|}\to 0$ quand $n\to\infty$ car $|\alpha|<1$.
Si $\alpha=\pm\frac{1}{\sqrt{2}}$ alors $r_X(k)=0$ si $k\neq 0$: X se comporte comme un bruit blanc gaussien.
(d) D'après le cours, $f_X(\lambda)=\frac{\sigma_\varepsilon^2}{2\pi}\big|1-2\alpha e^{i\lambda}\big|^2\times\big|1-\alpha e^{i\lambda}\big|^{-2}$.
D'après le cours, comme X est un ARMA gaussien, $\sqrt{n}\,\overline{X}_n\ \xrightarrow[n\to\infty]{\mathcal{L}}\ \mathcal{N}(0,2\pi\,f_X(0))=\mathcal{N}\big(0,\sigma_\varepsilon^2\,|1-2\alpha|^2\times|1-\alpha|^{-2}\big)$.
(e) On déduit facilement de ce qui précède que $N\big(\overline{X}_N\big)^2\ \xrightarrow[N\to\infty]{\mathcal{L}}\ \sigma_\varepsilon^2\,|1-2\alpha|^2\times|1-\alpha|^{-2}\chi^2(1)$.
(f) D'après le cours, $\sqrt{N}\big(\widehat{\sigma}_N^2,\widehat{r}_N(1)\big)-\big(r_X(0),r_X(1)\big)\ \xrightarrow[N\to\infty]{\mathcal{L}}\ \mathcal{N}\Big(0\,,\,\big(4\pi\displaystyle\int_{-\pi}^{\pi}f_X^2(\lambda)\cos((i-1)\lambda)\cos((j-1)\lambda)d\lambda\big)_{1\le i,j\le 2}\Big)$.
(g) On a $(r_X(0),r_X(1))=f(\sigma_\varepsilon^2,\alpha)$, donc $(\widehat{\sigma}_\varepsilon^2,\widehat{\alpha})=f^{-1}(\widehat{\sigma}_N^2,\widehat{r}_N(1))$ est un estimateur de $(\alpha,\sigma_\varepsilon^2)$. D'après ce qui précède et la

Delta-méthode, cet estimateur converge vers $(\alpha, \sigma_\varepsilon^2)$ suivant un Théorème de la limite centrale à la vitesse $\sqrt{N}$.
(h) Si $|2\alpha| < 1$, donc $|\alpha| < 1/2$, on a $Q(B)^{-1} = \sum_{i=0}^{\infty} (2\alpha)^i B^i$ et ainsi $\varepsilon = Q(B)^{-1} P(B) X$, donc $\varepsilon_N = X_N + \frac{1}{2} \sum_{i=1}^{\infty} (2\alpha)^i X_{N-i}$. On en déduit que $\widehat{X}_{N+1} = \mathbb{E}\big[X_{N+1} \mid (X_{N-k})_{k\geq 0}\big] = \mathbb{E}\big[\varepsilon_{N+1} \mid (X_{N-k})_{k\geq 0}\big] - 2\alpha\big(X_N + \frac{1}{2} \sum_{i=1}^{\infty} (2\alpha)^i X_{N-i}\big)$ soit encore $\widehat{X}_{N+1} = -2X_N + \sum_{i=1}^{\infty} (2\alpha)^i X_{N-i}$, d'après l'indépendance de ε_{N+1} et $(X_{N-k})_{k\geq 0}$. On en déduit naturellement une prédiction $\widehat{X}_{N+1} = -2X_N + \sum_{i=1}^{N-1} (2\widehat{\alpha})^i X_{N-i}$ avec $\widehat{\alpha}$ obtenu dans la question (g).

□

2. **(8.5 points)** A partir de la liste de commandes et de résultats qui suivent:
   (a) Commenter chacune des commandes écrites, en expliquant ce que ces commandes sont supposées faire **(3 pts)**.
   (b) Donner la légende des différentes figures et commenter ces figures **(1 pt)**.
   (c) Pour le résultat de la commande `summary(X.BIC)`, expliquer ce que signifie mathématiquement `Pr(>|t|)` et ce que l'on conclue contrétement **(1 pt)**.
   (d) Que peut-on conclure de la figure issue de la commande `acf(X.BIC$res)` puis de la commande `Box.test(X.BIC$res, lag = 5, type="Ljung")` (test portemanteau) **(1 pt)**?
   (e) Sans faire le calcul, mais en précisant explicitement les constantes, donner exactement le modèle obtenu pour X_t lorsque $t = 2000 + 2/12$ **(2 pts)**.
   (f) Que représente le résultat 23.39663 dans le résultat issue de la toute dernière commande `Xpred` **(0.5 pts)**?

Proof. (a) On désaisonnalise la série, puis on effectue une régression polynomiale dont le degré est choisi par le critère BIC, puis on modélise le bruit de la série détendancialisée et désaisonnalisée par un processus ARMA(1,1) après avoir observé la non-indépendance de ce bruit. Enfin, on effectue une prédiction pour les 12 mois de 2009.
(b) On représente la série, puis l'estimation de la saisonnalité, la régression polynomiale chosie et les résidus obtenus.
(c) La commande `Pr(>|t|)` est relative à des tests de student de nullité pour les coefficients de la régression choisie par le critère BIC. La régression est validée.
(d) Le corrélogramme montre la non-indépendance de ces résidus, ce que confirme un test du portemanteau.
(e) Le modèle obtenu pour le résidu est un ARMA(1,1) en plus de la saisonnalité (mois de mars) et de la régression obtenue précédemment. On obtient ainsi:

$$X_t = 3.8918 + 19 + 0.0009521 * 74^2 - 0.000004133 * 74^3 + \widehat{\varepsilon}_t.$$

(f) C'est la prédiction du chiffre d'affaire pour le mois de mai 2009. □