MIMM7 - Séries Chronologiques

Correction de quelques exercices du TD2

1. **(**)** Soit (ε_n) un bruit blanc gaussien de variance σ^2. On considère les processus X et Y, où:

$$X_n = A\cos(\pi n/3) + B\sin(\pi n/3) + Y_n \quad \text{et} \quad Y_n = \varepsilon_n + 2.5\varepsilon_{n-1} \quad \text{pour } n \in \mathbb{N},$$

où A et B sont deux variables normales centrées réduites indépendantes et indépendantes des ε_n.

(a) Quel type de processus est Y? Tracer la densité spectrale et le corrélogramme (théorique) de Y. Déterminer la loi de Y_n pour $n \in \mathbb{N}$.

(b) X a-t-il une tendance? une saisonnalité?

(c) Montrer que X est stationnaire et déterminer sa densité spectrale et son corrélogramme.

(d) Déterminer la loi de X_n pour $n \in \mathbb{N}$.

Correction : (a) Y est un MA(1). Sa densité est $f_Y(\lambda) = \dfrac{\sigma^2}{2\pi}\left|1+2.5e^{i\lambda}\right|^2$, soit $f_Y(\lambda) = \dfrac{\sigma^2}{2\pi}(7.25+5\cos\lambda)$. Sa covariance vérifie : $r(0) = 7.25\sigma^2$, $r(1) = 2.5\sigma^2$, et $r(k) = 0$ pour $|k| \geq 2$. Ainsi la corrélation théorique de Y est $\rho(0) = 1$, $\rho(1) = 10/29$ et $\rho(k) = 0$ pour $|k| \geq 2$. Y_n est une variable aléatoire gaussienne centrée car ε est un bruit centré gaussien, donc de distribution $\mathcal{N}(0, 7.25\sigma^2)$.
(b) On a $\mathbb{E}X_n = 0$ pour $n\imath\mathbb{Z}$, car $\mathbb{E}A = \mathbb{E}B = 0$. Par suite, X ne possède ni tendance ni saisonnalité.
(c) Pour montrer la stationnarité de X, il nous faut juste montrer que $\mathbb{E}X_nX_{n+k}$ ne dépend que de k, car X est une série chronologique gaussienne centrée. Ainsi,

$$\begin{aligned}
\mathbb{E}X_nX_{n+k} &= \cos\left(\frac{\pi n}{3}\right)\cos\left(\frac{\pi(n+k)}{3}\right) + \sin\left(\frac{\pi n}{3}\right)\sin\left(\frac{\pi(n+k)}{3}\right) + \mathbb{E}Y_nY_{n+k} && (1)\\
&= \cos\left(\frac{\pi k}{3}\right) + r_Y(k), && (2)
\end{aligned}$$

car Y est bien un processus stationnaire. X est bien un processus stationnaire. On remarque en revanche que r_X ne tend pas vers zéro en $+\infty$. Donc $(r_X(k))_k$ n'est pas une suite sommable, donc on ne peut définir de densité spectrale. On peut cependant déterminer son corrélogramme :

$$\rho_X(k) = \frac{1}{1+7.25\sigma^2}\left(\cos\left(\frac{\pi k}{3}\right) + r_Y(k)\right).$$

(d) X est une variable aléatoire gaussienne centrée et de variance $1 + 7.25\sigma^2$.

2. (*) Soit (ε_n) un bruit blanc de variance 1. On considère les processus X et Y, où:

$$X_n = X_{n-3} + \varepsilon_n \quad \text{et} \quad Y_n = 0.999Y_{n-3} + \varepsilon_n \quad \text{pour } n \in \mathbb{Z}.$$

   (a) Quel type de processus est Y? Tracer la densité spectrale et le corrélogramme (théorique) de Y.

   (b) Montrer que X n'est pas stationnaire. Calculer $\rho(n) = \mathbb{E}X_0X_n$ pour $n \in \mathbb{N}$ et tracer $\rho(n)$ en fonction de n. Que remarque-t-on?

   (c) Déterminer un filtre rendant X stationnaire (et non nul). Appliquer ce filtre à Y et déterminer la densité spectrale du processus filtré.

Correction : (a) Y est un processus AR(3). La densité spectrale de Y est :

$$f_Y(\lambda) = \frac{1}{2\pi}\frac{1}{|1-0.999e^{3i\lambda}|^2} = \frac{1}{2\pi}\frac{1}{(0.999-\cos(3\lambda))^2+\sin^2(3\lambda)}.$$

On observe donc que lorsque $\lambda \to 0$ alors $f_Y(\lambda)$ est très grand.
En ce qui concerne le corrélogramme, on a directement $r(k) = (0.999)^3 r(k-3)$ pour $k \geq 3$. De plus, $r(0) = \mathbb{E}Y_0^2 = \mathbb{E}(0.999Y_{-3}+\varepsilon_n)^2 = 0.999^2 r(0) + 1$, d'où $r(0) \simeq 500$. Enfin, $r(1) = \mathbb{E}Y_0Y_1 = 0.999r(2)$ et $r(2) = \mathbb{E}Y_0Y_2 = 0.999r(1)$. Donc $r(1) = r(2) = 0$. En conséquence, on a

$$\rho(3n) = (0.999)^{3n} \quad \text{et} \quad \rho(3n+1) = \rho(3n+2) = 0 \quad \text{pour } n \in \mathbb{N},$$

donc le processus garde une très longue mémoire de son passé, car $\rho(k)$ est proche de 1 même pour $n = 60$.
(b) On a $\mathbb{E}X_n = 0$ pour tout $n \in \mathbb{Z}$, mais $\mathbb{E}X_n^2 = \mathbb{E}X_{n-3}^2 + 1$, donc la variance de X_n n'est pas une constante : X n'est pas un processus stationnaire.
On a $\rho(n) = \mathbb{E}X_0X_n = \mathbb{E}X_0X_{n-3}$, donc $\rho(n) = 0$ si n n'est pas un multiple de 3 et $\rho(n) = \mathbb{E}X_0^2 = \infty$ si n est un multiple de 3.
(c) X possède des accroissements de longueur 3 stationnaires. Pour rendre X stationnaire, il suffit d'utiliser le filtre $(1,0,0,-1)$, soit la série filtrée $X'_n = X_{n+3} - X_n$. On a évidemment $X'_n = \varepsilon_{n+3}$ pour $n \in \mathbb{Z}$.
Si on applique ce filtre à Y, on obtient $Y'_n = Y_{n+3} - Y_n = 0.999Y_n + \varepsilon_{n+3} - (0.999Y_{n-3} + \varepsilon_n)$ donc $Y'_n = 0.999Y'_{n-3} + \varepsilon_{n+3} - \varepsilon_n$. Par suite, (Y'_n) est un processus ARMA(3,3). La desnité spectrale de ce processus est donc :

$$f_{Y'}(\lambda) = \frac{1}{2\pi}\frac{\left|1-e^{3i\lambda}\right|^2}{|1-0.999e^{3i\lambda}|^2}.$$

3. (**) Soit (ε_n) un bruit blanc gaussien de variance σ^2. On considère le processus X, où:

$$X_n = \varepsilon_n - \theta\varepsilon_{n-1} \quad \text{pour } n \in \mathbb{N},$$

où $\theta \in]-1,1[$.

   (a) Quel type de processus est X? Déterminer sa densité spectrale et sa fonction de covariance (soit $r(k) = \mathbb{E}X_0X_k$ pour $k \in \mathbb{N}$).

(b) Calculer la log-vraisemblance exacte (c'est-à-dire $\log f_{(X_1,\ldots,X_n)}(x_1,..,x_n)$) en fonction de σ^2 et θ (on écrira matriciellement cette expression). On notera $\Sigma_n(\sigma,\theta)$ la matrice d'ordre n: $(r(j-i))_{1\leq i,j\leq n}$. Peut-on obtenir une expression simple de l'estimateur de (σ^2,θ) par maximum de vraisemblance?

(c) Écrire le contraste de Whittle pour les paramètres (σ^2,θ). Peut-on obtenir une expression simple de l'estimateur de (σ^2,θ) par minimum de contraste?

(d) Répondre au deux questions précédentes en supposant $\theta = 0.5$ connu (le paramètre à estimer est alors seulement σ^2).

Correction : (a) X est un processus MA(1), de densité spectrale

$$f_X(\lambda) = \frac{\sigma^2}{2\pi}\left|1-\theta e^{i\lambda}\right|^2,$$

et de covariance $r(0) = \sigma^2(1+\theta^2)$, $r(1) = r(-1) = -\theta\sigma^2$ et $r(k) = 0$ pour $|k| \geq 2$.
(b) On définit une matrice $\Sigma_n(\sigma,\theta)$ de covariance de $(X_1,\cdots,X_n)$ et on a ainsi :

$$\Sigma_n(\sigma,\theta) = \sigma^2 \begin{pmatrix} 1+\theta^2 & -\theta & 0 & \cdots & 0 & 0 \\ -\theta & 1+\theta^2 & -\theta & \cdots & 0 & 0 \\ 0 & -\theta & 1+\theta^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & 1+\theta^2 & -\theta \\ 0 & 0 & 0 & \cdots & -\theta & 1+\theta^2 \end{pmatrix}.$$

On peut obtenir le déterminant de $\Sigma_n(\sigma,\theta)$ par récurrence, mais il est plus délicat (et même) impossible) de donner une formulation simple de $\Sigma_n^{-1}(\sigma,\theta)$. Aussi doit se contenter d'écrire que :

$$\log f_{(X_1,\cdots,X_n)}(x_1,\cdots,x_n) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(det(\Sigma_n(\sigma,\theta)) - \frac{1}{2}(x_1,\cdots,x_n)\Sigma_n^{-1}(\sigma,\theta)^t(x_1,\cdots,x_n),$$

et pour déterminer une estimation des paramètres (σ,θ), il faut maximiser cette vraisemblance, ce qui peut se faire numériquement en calculant cette vraisemblance pour des valeurs précises de (σ,θ) et en cherchant l'argument maximal. En revanche, il est très difficile d'obtenir une expression exacte des estimateurs à partir d'une dérivation de la vraisemblance.
(c) Le contraste de Whittle s'écrit :

$$U_n(\sigma,\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\log(f_X(\lambda)) + \frac{I_n(\lambda)}{f_X(\lambda)}\right)d\lambda,$$

avec $f_X(\lambda)$ la densité spectrale (voir plus haut) et $I_n(\lambda)$ le péridogramme dont on peut donner les expressions suivantes :

$$f_X(\lambda) = \frac{\sigma^2}{2\pi}(1+\theta^2-2\theta\cos\lambda) \quad \text{et} \quad I_n(\lambda) = \frac{1}{\pi n}\sum_{1\leq k,k'\leq n} X_k X_{k'}\cos(\lambda(k-k')).$$

On peut alors écrire le contraste de Whittle sous la forme plus expicite :

$$U_n(\sigma,\theta) = \frac{1}{2}\log\left(\frac{\sigma^2}{2\pi}\right) \quad + \quad \frac{1}{4\pi}\int_{-\pi}^{\pi}\log(1+\theta^2-2\theta\cos\lambda)d\lambda + \ldots$$

$$.. + \frac{1}{2\pi n\sigma^2}\sum_{1\leq k,k'\leq n} X_k X_{k'}\int_{-\pi}^{\pi}\frac{\cos(\lambda(k-k'))}{1+\theta^2-2\theta\cos\lambda}d\lambda.$$

On obtient bien une fonction dépendant de σ et de θ que l'on peut numériquement minimiser en approchant les intégrales par des sommes de Riemann. Pour le faire théoriquement, ceci est plus délicat car la seconde intégrale n'a pas une expression simple.
(d) Si on sait que $\theta = 0.5$, alors :

- Pour la méthode du maximum de vraisemblance, la matrice $\Sigma_n(\sigma, \theta)$ s'écrit comme $\sigma^2.A_n$ où A_n est une matrice ne dépendant que de n. Alors, on peut maximiser la vraisemblance en dérivant par rapport à σ^2 et on obtient :

  $$\frac{\partial \log f_{(X_1,\cdots,X_n)}(x_1,\cdots,x_n)}{\partial \sigma^2} = -\frac{1}{2}\frac{1}{\sigma^2}\log(det\ A_n) + \frac{1}{2}\frac{1}{\sigma^4}(x_1,\cdots,x_n)A_n^{-1t}(x_1,\cdots,x_n),$$

  donc on trouve pour estimateur de σ^2,

  $$\widehat{\sigma^2} = \frac{(X_1,\cdots,X_n)A_n^{-1t}(X_1,\cdots,X_n)}{\log(det\ A_n)}.$$

- Pour la méthode du minimumn de contraste, on peut alors dériver le contraste par rapport à σ^2 et :

  $$\frac{\partial U_n(\sigma)}{\partial \sigma^2} = \frac{1}{2\sigma^2} - \frac{1}{2\pi n \sigma^4}\sum_{1\leq k,k'\leq n} X_k X_{k'}\int_{-\pi}^{\pi}\frac{\cos(\lambda(k-k'))}{1.25-\cos\lambda}d\lambda.$$

  donc on trouve pour estimateur de σ^2,

  $$\widehat{\sigma^2} = \frac{1}{n\pi}\sum_{1\leq k,k'\leq n} X_k X_{k'}\int_{-\pi}^{\pi}\frac{\cos(\lambda(k-k'))}{1.25-\cos\lambda}d\lambda.$$