Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2010 − 2011

Statistiques II

Contrôle continu n°2, avril 2011

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

1. **(Sur 15 points)** On considère une suite $(\varepsilon_i)_{i\in\mathbb{N}}$ de variables aléatoires indépendantes identiquement distribuées (vaiid) centrées et de variance commune σ^2 et de moment d'ordre 4, $\mu_4 = \mathbb{E}\varepsilon_0^4 < \infty$. On définit pour tout $i \in \mathbf{Z}$ la suite $(X_i)_{i\in\mathbf{Z}}$ telle qu'avec $a \in]-1,1[$:
$$X_{i+2} = a\,X_i + \varepsilon_{i+2} \qquad \text{pour tout } i \in \mathbf{Z}.$$
   (a) Rappeler, en justifiant, quel processus est $(X_i)_{i\in\mathbf{Z}}$ **(0.5pts)**. Est-il stationnaire **(0.5pts)**? Causal **(0.5pts)**?
   (b) Déterminer l'expression de X_i en fonction de $(\varepsilon_{i-k})_{k\in\mathbf{N}}$ **(1pt)**.
   (c) Déterminer l'expression (sans nombre complexe) de la densité spectrale de (X_i) **(1pt)**.
   (d) Montrer que $r(k)$, l'autocovariance de $(X_i)_{i\in\mathbf{Z}}$ vérifie $r(2n) = \mu a^n$ et $r(2n+1) = 0$ pour $n \in \mathbf{N}$, où μ est un réel que l'on pourra exprimer en fonction de σ^2 et a **(2.5pts)**.
   (e) On suppose que $(X_1, \cdots, X_N)$ est observé, avec N "grand", les ε_i, a et σ^2 étant inconnus. On désire estimer a et σ^2. On définit $\widehat{a}_N = \text{Argmin}_{\{\alpha,\, 0\le\alpha<1\}} Q_N(\alpha)$ où $Q_N(\alpha) = \frac{1}{N-2}\sum_{k=3}^N (X_k - \alpha\, X_{k-2})^2$. Montrer que $\widehat{a}_N$ est unique et déterminer son expression exacte en fonction de $(X_1, \cdots, X_N)$ **(2pts)**.
   (f) Montrer que $Q_N(a)$ vérifie une loi forte des grands nombres et un théorème central limite **(1pt)**.
   (g) Plus généralement, on peut montrer (ne pas le faire) que pour tout $|\alpha| < 1$, $Q_N(\alpha) \xrightarrow[n\to+\infty]{p.s.} Q(\alpha)$ avec $Q(\alpha) = \lim_{N\to\infty} \mathbb{E} Q_N(\alpha)$. Calculer $Q(\alpha)$ **(1pt)**et montrer pour tout $\alpha \neq a$, $Q(\alpha) > Q(a)$ **(1pt)**. En déduire que $\widehat{\alpha}_N \xrightarrow[N\to+\infty]{p.s.} a$ **(3pts)**.
   (h) Montrer alors que $\widehat{\sigma}^2_N = Q_N(\widehat{a}_N)$ est un estimateur convergeant presque sûrement vers σ^2 **(1pt)**.

Proof. (a) X est un AR[2] stationnaire causal...
(b) Il est clair que $X_n = \sum_{i=0}^\infty a^i \varepsilon_{n-2i}$.
(c) On a $f(\lambda) = \frac{\sigma^2}{2\pi}\left|1 - a e^{2i\lambda}\right|^{-2} = \frac{\sigma^2}{2\pi}\left|1 + a^2 - 2a\cos(2\lambda)\right|^{-1}$.
(d) On a clairement $r(k+2) = a r(k)$ pour tout $k \geq 0$. De ceci on déduit bien que $r(2n) = \mu a^n$, où μ est un réel. De plus, de par l'expression précédente de X_n en fonction des ε_i, il est clair que $\mathbb{E}X_0 X_{2n+1} = 0$ donc $r(2n+1) = 0$. Il reste à déterminer $r(0)$ et $r(1)$. Mais $r(0)(1+a^2) - 2ar(2) = \sigma^2$ (avec $\mathbb{E}(X_{n+2} - aX_n)^2 = \mathbb{E}\varepsilon^2_{n+2}$) et avec $r(1) = 0$ on a $r(0) = \frac{\sigma^2}{1-a^2}$ d'où $r(2n) = a^n r(0)$ et $r(2n+1) = 0$ pour tout $n \in \mathbf{N}$.
(e) Soit $Y^N = (X_3, X_4, \cdots, X_N)$ et $Z^N = (X_1, X_2, \cdots, X_{N-2})$, qui sont deux vecteurs de $\mathbf{R}^{N-2}$. On a clairement $(N-2)Q_N(\alpha)$ qui est la norme au carré pour le produit scalaire classique sur $\mathbf{R}^{N-2}$ entre Y^N et aZ^N donc comme on est en dimension finie le minimum de la distance représente la distance entre Y^N et le sev de dim engendré par Z^N: ce minimum existe bien et est unique. On obtient donc $\widehat{a}_N = ({}^t Z^N Z^N)^{-1}\, {}^t Z^N Y^N = \left(\sum_{i=1}^{N-2} X_{i+2}X_i\right)\left(\sum_{i=1}^{N-2} X_i^2\right)^{-1}$.
(f) On a $Q_N(a) = \frac{1}{n-2}\sum_{k=3}^n \varepsilon_k^2$. Comme les (ε_k^2) sont des vaiid d'espérance finie valant σ^2, d'après la loi forte des grands nombres, $Q_N(a) \xrightarrow[n\to+\infty]{p.s.} \sigma^2$. Comme les (ε_k^2) sont également des vaiid de variance finie valant $\mu^4 - \sigma^4$, d'après le théorème central limite, on obtient $\sqrt{N-2}\big(Q_N(a) - \sigma^2\big) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0, \mu_4 - \sigma^4)$. (g) Du fait de la stationarité de X, on a
$$\begin{aligned}\mathbb{E}Q_N(\alpha) &= \mathbb{E}(X_2 - \alpha X_0)^2 = (1+\alpha^2)r(0) - 2\alpha r(2)\\ &= \frac{(1+\alpha^2) - 2\alpha a}{1-a^2}\sigma^2 = \frac{(\alpha-a)^2 + 1 - a^2}{1-a^2}\sigma^2.\end{aligned}$$
On voit donc bien que $Q(\alpha)$ est minimum quand $\alpha = a$ et donc $Q(\alpha) > Q(a)$ pour $\alpha \neq a$.
Le minimum de la fonction de Q est donc a. Or $Q_N(\alpha)$ tend presque sûrement vers $Q(\alpha)$ donc le minimum de $Q_N(\alpha)$, soit $Q_N(\widehat{a}_N)$ tend presque sûrement vers le mimimum des $Q(\alpha)$ donc vers $Q(a)$. Or la fonction $\alpha \mapsto Q_N(\alpha)$ est continue presque sûrement, donc si $Q_N(\widehat{a}_N) \xrightarrow[N\to+\infty]{p.s.} Q(a)$ on a bien $\widehat{a}_N \xrightarrow[N\to+\infty]{p.s.} a$.
(h) D'après ce qui précède $Q_N(\widehat{a}_N) \xrightarrow[N\to+\infty]{p.s.} Q(a)$, donc $Q_N(\widehat{a}_N) \xrightarrow[N\to+\infty]{p.s.} \sigma^2$. □

2. **(Sur 14 points)** Soit une chaîne de Markov $X = (X_t)_{t\in\mathbf{N}}$ homogène prenant les valeurs 1 ou -1 et telle que $\mathbb{P}(X_1 = 1 \mid X_0 = 1) = p$ et $\mathbb{P}(X_1 = -1 \mid X_0 = -1) = q$ où p et q sont des réels de $[0,1]$. On suppose également que X_0 suit une loi quelconque inconnue.

   (a) Déterminer la matrice de transition de X **(0.5pts)**. A quelles conditions sur p et q cette chaîne est-elle irréductible **(0.5pts)**? On se placera désormais sous ces conditions.

   (b) Déterminer la mesure invariante μ de X **(1pt)**. On suppose désormais que X_0 suit la loi μ.

   (c) Montrer que $\mathbb{E}X_n = \dfrac{p-q}{2-p-q}$ **(0.5pts)** et calculer $\mathrm{var}X_n$ pour $n \in \mathbf{N}$ **(0.5pts)**.

   (d) Montrer que $\mathrm{cov}(X_0, X_1) = 4\dfrac{(1-q)(1-p)(p+q-1)}{(2-p-q)^2}$ **(3pts)**. A quelle condition sur p et q X est-il un bruit blanc **(1pt)**?

   (e) On considère le processus $Y = (Y_n)_{n\in\mathbf{N}}$ modélisant par exemple défini par $Y_n = a\,X_n + b + \varepsilon_n$ pour $n \in \mathbf{N}$, où ε_i est un bruit blanc fort gaussien indépendant de (X_n) et a, b deux réels inconnus. Déterminer la tendance, la saisonnalité et la partie bruit $u = (u_n)_{n\in\mathbf{N}}$ de Y **(0.5pts)**. Ce bruit est-il stationnaire **(1pt)**? Un bruit blanc **(1pt)**?

   (f) On suppose que l'on observe $(X_1, \cdots, X_N)$ et $(Y_1, \cdots, Y_N)$ et l'on désire estimer a et b par régression linéaire. Donner l'expression d'un estimateur $(\widehat{a}_N, \widehat{b}_N)$ de (a,b) (on pourra utiliser les notations $\overline{X}_N = \frac{1}{N}\sum_{i=1}^N X_i$, $\overline{Y}_N = \frac{1}{N}\sum_{i=1}^N Y_i$, $\overline{\sigma}_X^2 = \frac{1}{N}\sum_{i=1}^N X_i^2 - \overline{X}_N^2$ et $\overline{\sigma}_{XY} = \frac{1}{N}\sum_{i=1}^N X_iY_i - \overline{X}_N\overline{Y}_N$) **(2pts)**. Montrer que cet estimateur converge **(2.5pts)**.

Proof. (a) La matrice de transition de X est $P = \begin{pmatrix} q & 1-q \\ 1-p & p \end{pmatrix}$. La chaîne de Markov est irréductible si tous les états peuvent être visités donc si $0 \le p, q < 1$.
(b) La mesure invariante $\mu = (\mu_{-1}, \mu_1)$ est telle que $\mu P = \mu$ soit $\mu_{-1}(q-1) + (1-p)\mu_1 = 0$ et $\mu_{-1} + \mu_1 = 1$. On trouve donc que $\mu_{-1} = \dfrac{1-p}{2-p-q}$ et $\mu_1 = \dfrac{1-q}{2-p-q}$.
(c) Comme X_0 a pour loi μ, X_n a pour loi μ pour tout $n \in \mathbf{N}$. D'où pour tout $n \in \mathbf{N}$, on obtient que $\mathbb{E}X_n = \dfrac{p-q}{2-p-q}$ et $\mathrm{var}X_n = 1 - (\mathbb{E}X_n)^2 = \dfrac{(2-p-q)^2 - (p-q)^2}{(2-p-q)^2} = 4\dfrac{(1-p)(1-q)}{(2-p-q)^2}$.
(d) $\mathbb{E}X_0X_1 = (P(X_0 = 1 \cap X_1 = 1) + P(X_0 = -1 \cap X_1 = -1) - (P(X_0 = 1 \cap X_1 = -1) + P(X_0 = -1 \cap X_1 = 1)$, soit $\mathbb{E}X_0X_1 = P(X_0 = 1)(2p-1) + P(X_0 = -1)(2q-1) = \dfrac{(1-q)(2p-1) + (1-p)(2q-1)}{2-p-q} = \dfrac{3(p+q) - 4pq - 2}{2-p-q}$. Ainsi on en déduit que $\mathrm{cov}(X_0, X_1) = \dfrac{(3(p+q) - 4pq - 2)(2-p-q) - (p-q)^2}{(2-p-q)^2} = 4\dfrac{(1-q)(1-p)(p+q-1)}{(2-p-q)^2}$. Ainsi $\mathrm{cov}(X_0, X_1) = 0$ si et seulement si $p + q = 1$. Mais si $p + q = 1$, on a également $P(X_1 = \pm 1 | X_0 = \pm 1) = P(X_1 = \pm 1)$ donc X_1 et indépendant de X_0. Comme X_2 ne dépend en loi que de X_1 on en déduit que X_2 est aussi indépendant de X_0 et par itération les X_i sont tous indépendants: X est alors un bruit blanc fort lorsque $\mathbb{E}X_0 = 0$ soit $p = q = 1/2$.
(e) On a $\mathbb{E}Y_n = a\frac{p-q}{2-p-q} + b$ donc Y adment pour tendance $a\frac{p-q}{2-p-q} + b$ et pas de saisonnalité. Le bruit u vérifie $u_n = a(X_n - \mathbb{E}X_n) + \varepsilon_n$ donc comme X et ε sont indépendants, comme $X - \mathbb{E}X_0$ et ε sont stationnaires, on en déduit que u est stationnaire. En revanche, u est un bruit blanc si et seulement si $p + q = 1$.
(f) Après de simples calculs, on obtient $\begin{pmatrix} \widehat{a}_N \\ \widehat{b}_N \end{pmatrix} = \begin{pmatrix} \overline{\sigma}_{XY}/\overline{\sigma}_X^2 \\ \overline{Y}_N - \overline{X}_N\overline{\sigma}_{XY}/\overline{\sigma}_X^2 \end{pmatrix}$.
On a $\overline{X}_N \xrightarrow[N\to+\infty]{p.s.} \mathbb{E}X_0$ d'après le Théorème ergodique, $\overline{Y}_N = a\overline{X}_N + b + \overline{\varepsilon}_N \xrightarrow[N\to+\infty]{p.s.} a\mathbb{E}X_0 + b$ d'après le Théorème ergodique et la LFGN pour (ε_i), $\overline{\sigma}_X^2 \xrightarrow[N\to+\infty]{p.s.} \mathrm{var}(X_0)$ d'après le Théorème ergodique et enfin $\overline{\sigma}_{XY} = a\overline{\sigma}_X^2 + \overline{X\varepsilon}_N \xrightarrow[N\to+\infty]{p.s.} \mathrm{var}(X_0)$ d'après le Théorème ergodique et la LFGN. Donc $\widehat{a}_N \xrightarrow[N\to+\infty]{p.s.} a$ et $\widehat{b}_N \xrightarrow[N\to+\infty]{p.s.} b$. □