MIMM7 - Séries Chronologiques

Correction de quelques exercices du TD1

1. **(**)** On suppose que la série $X = \{X_t, t \in \mathbb{N}\}$ est de la forme $X_t = P(t) + \varepsilon_t$ avec P un polynôme de degré 1 et ε un bruit blanc gaussien de variance σ^2 inconnue. On suppose connu $(X_1, \cdots, X_N)$. Déterminer l'expression exacte de l'estimation de P par moindres carrés. Quelle est sa vitesse de convergence?

   Correction : On écrit donc $P(t) = at + b$, soit, avec les notations du cours, une matrice
$$Z = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ . & . \\ N & 1 \end{pmatrix}$$
   On obtient ainsi que ${}^tZ.Z = N \begin{pmatrix} (N+1)(2N+1)/6 & (N+1)/2 \\ (N+1)/2 & 1 \end{pmatrix}$, soit
   $({}^tZ.Z)^{-1} = \dfrac{1}{N(N-1)} \begin{pmatrix} 12/(2N+1) & -6 \\ -6 & 4N+2 \end{pmatrix}$. Après calculs, on trouve que le vecteur colonne $\hat{P} = Z({}^tZ.Z)^{-1}.{}^tZ.X$, soit pour $i = 1, \cdots, N$,
$$\hat{P}(i) = \frac{1}{N(N-1)} \left[\left(\frac{12i}{N+1} - 6 \right) \sum_{j=1}^{N} jX_j + (4N+2-6i) \sum_{j=1}^{N} X_j \right].$$
   Pour calculer la vitesse de convergence, on utilise le fait que $\mathbb{E}X_i = P(i)$ et que $\mathrm{var}X_i = \sigma^2$. Si X est la matrice colonne composée des X_i, pour $i = 1, \cdots, N$, alors $\mathrm{cov}(X) = \sigma^2 I_N$ où I_N est la matrice identité de taille N. En conséquence,
   $\mathrm{cov}(\hat{P}) = {}^t(Z({}^tZ.Z)^{-1}.{}^tZ)\mathrm{cov}(X)(Z({}^tZ.Z)^{-1}.{}^tZ) = \sigma^2 Z({}^tZ.Z)^{-1}.{}^tZ$, donc $\mathrm{cov}(\hat{P})$ est la matrice de terme général $\dfrac{\sigma^2}{N(N-1)} \left(\dfrac{12ij}{N+1} - 6j + 4N + 2 - 6i \right)_{1 \leq i,j \leq N}$. On en déduit que $\mathrm{cov}(\hat{P}) = \mathcal{O}(1/N)$, donc on a bien convergence du vecteur $\hat{P}$ vers P, et la vitesse de convergence est $\sqrt{N}$. ∎

2. **(***)** On suppose que la série $X = \{X_t, t \in \mathbb{N}\}$ est de la forme $X_t = s(t) + \varepsilon_t$ avec s une fonction de période 2 et ε un bruit blanc gaussien de variance σ^2 inconnue. On suppose connu $(X_1, \cdots, X_{2N})$.

   (a) Montrer que la variance empirique de X n'est pas un estimateur convergent vers la variance σ^2.

   (b) Déterminer l'expression exacte de l'estimation de s par une régression par moindres carrés. Quelle est la vitesse de convergence?

   (c) Déterminer alors une estimation de σ à partir de la série à laquelle on a retiré l'estimation de s. Quelle est alors la vitesse de convergence?

Correction : a/ On a $X_k = s(-1)^k + \varepsilon_k$ pour $k = 1, \cdots, 2N$. On a facilement $\bar{X}_{2N} = \bar{\varepsilon}_{2N}$ et ainsi, on montre que la variance empirique des X_i est reliée à la variance empirique des ε_i par la relation :

$$S_X^2(2N) = \frac{1}{2N}\sum_{i=1}^{2N}(X_i - \bar{X}_{2N})^2 = s^2 + \frac{1}{2N}\left(\sum_{i=1}^{2N}(\varepsilon_i - \bar{\varepsilon}_{2N})^2 + \sum_{p=1}^{N} 2s(\varepsilon_{2p} - \varepsilon_{2p+1})\right).$$

Par suite, $\mathbb{E}S_X^2(2N) = s^2 + \mathbb{E}S_\varepsilon^2(2N)$, donc $S_X^2(2N)$ est un estimateur biaisé de σ^2, dont le biais s^2 est inconnu. Ce ne peut donc par être un estimateur convergent.

b/ On considère ainsi la matrice colonne de taille $(2N, 1)$:

$$Z = \begin{pmatrix} 1 \\ -1 \\ 1 \\ . \quad . \\ -1 \end{pmatrix}$$

On alors ${}^tZ.Z = 2N$. Ainsi, on obtient une estimation de s,

$$\hat{s} = \frac{1}{2N}\sum_{i=1}^{N}(X_{2i} - X_{2i-1}) = \frac{1}{2N}\sum_{i=1}^{2N}(-1)^i X_{2i}.$$

(Attention, si on avait posé 2 valeurs s_1 et s_2, on aurait trouvé 2 estimateurs distincts de s car on n'aurait pas utiliser dans l'approximation la liaison $s_1 + s_2 = 0$). Posons maintenant $Z_i = X_{2i} - X_{2i-1}$. De l'expression de X_i, on sait que $\mathbb{E}Z_i = 2s$ et $\mathrm{var}Z_i = 2\sigma_2$. De plus, (Z_i) est une suite de v.a. gaussiennes indépendantes. Par conséquent, on obtient le comportement de $\hat{s}$:

$$\sqrt{N}(\hat{s} - s) \underset{N\to+\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0, \sigma^2/2),$$

soit une vitesse de convergence en $\sqrt{N}$.

c/ On considère donc maintenant $Y_i = X_i - (-1)^i\hat{s} = \varepsilon_i + (-1)^i(s - \hat{s})$.
On peut alors estimer σ^2 à partir de la variance empirique des Y_i. On a alors $\bar{Y}_{2N} = \frac{1}{2N}\sum_{j=1}^{2N} Y_j = \frac{1}{2N}\sum_{j=1}^{2N}\varepsilon_j = \bar{\varepsilon}_{2N}$, d'où $S_Y^2(2N) = \frac{1}{2N}\sum_{i=1}^{2N}(\varepsilon_i + (-1)^i(s - \hat{s}) - \bar{\varepsilon}_{2N})^2$, ou encore
$S_Y^2(2N) = S_\varepsilon^2(2N) + (s - \hat{s})^2 + \frac{1}{N}(s - \hat{s})\sum_{i=1}^{2N}(-1)^i(\varepsilon_i - \bar{\varepsilon}_{2N}) = S_\varepsilon^2(2N) - (s - \hat{s})^2$. Or, $S_\varepsilon^2(2N)$ converge à la vitesse $\sqrt{N}$ vers σ^2 et $(s - \hat{s})$ converge à la vitesse $\sqrt{N}$ vers 0. Donc, la variance empirique des Y converge vers σ^2 à la vitesse $\sqrt{N}$.

3. **(**)** Soit (ε_t) un bruit blanc gaussien de variance σ^2. On considère $X_t = (3t + \varepsilon_t)^2 + (-1)^t$ pour $t \in \mathbb{N}$.

   (a) Calculer $\mathbb{E}X_t$. En déduire la tendance et la saisonnalité (et la période) de X_t.

   (b) Calculer $\mathrm{var}X_t$.

   (c) On considère la série détendancialisée et désaisonnalisée. Est-elle stationnaire?

Correction : a/ On a $\mathbb{E}X_t = (-1)^t + 9t^2 + \mathbb{E}(\varepsilon_t^2) = (-1)^t + 9t^2 + \sigma^2$. Donc, x a une saisonnalité $(-1)^t$ de période 2, et une tendance $9t^2 + \sigma^2$, polynôme du second degré.

b/ On a $\mathrm{var}X_t = \mathrm{var}(\varepsilon_t^2 + 6t\varepsilon) = 2\sigma^4 + 36t^2\sigma^2$.

c/ La série détendancialisée et désaisonnalisée est $\varepsilon_t^2 - \sigma^2 + 6t\varepsilon$. Ce n'est pas une série stationnaire (voir le calcul de la variance précédent).

4. **(**)** On considère la série (X_t), où $X_t = P(t) + s(t) + \varepsilon_t$ avec $P(t)$ un polynôme de degré 3, $s(t)$ une fonction périodique de période 4, telle que $s(1) + s(2) + s(3) + s(4) = 0$, et (ε_t) est un bruit banc gaussien de variance σ^2.

   (a) Déterminer un filtre linéaire éliminant la tendance, puis un filtre linéaire éliminant la saisonnalité.

   (b) Quel est le filtre composé de ces deux filtres?

   (c) Déterminer la loi de la série filtrée. Est-ce un bruit blanc?

Correction : a/ D'après le cours un filtre éliminant la tendance est obtenu à partir des coefficients du polynôme $(1 - X)^4$, soit le filtre $a_i = (-1)^i C_i^4$ pour $i = 0, 1, \cdots, 4$. Pour éliminer une saisonnalité de période 4, on peut choisir le filtre $b_4 = 1$, $b_1 = b_2 = b_3 = 0$ et $b_0 = -1$, ou bien le filtre $c_1 = c_2 = c_3 = c_4 = 1/4$.

b/ Par le premier filtre (a), on obtient la série $Y_i = X_{i+4} - 4X_{i+3} + 6X_{i+2} - 4X_{i+1} + X_i$, puis par le filtre (b) on obtient $Z_i = Y_{i+4} - Y_i = X_{i+8} - 4X_{i+7} + 6X_{i+6} - 4X_{i+5} + 4X_{i+3} - 6X_{i+2} + 4X_{i+1} - X_i$, soit le filtre (d) correspondant.

c/ La série Z ainsi filtrée est donc stationnaire comme filtre linéaire des ε_i, gaussienne, centrée et de variance : $138\sigma^2$. Ce n'est pas un bruit blanc, car, par exemple, $\mathrm{cov}(Z_i, Z_{i+8}) = -\sigma^2$.