

Chapitre 3

Estimation

3.I Ergodicité

Soit $x = (x_t, t = 1, \dots, N)$ une série temporelle. Modéliser cette série par un processus stationnaire $X = (X_t, t \in \mathbb{Z})$, c'est supposer qu'elle en est une réalisation "typique", grâce à laquelle on espère en déduire la loi de X . S'il s'agit d'un bruit blanc, ce problème est du ressort de la loi forte des grands nombres. Pour des processus plus généraux, il faut faire appel aux théorèmes dits ergodiques.

Définition : Le processus stationnaire $(X_t, t \in \mathbb{Z})$ est dit **ergodique** si

$$\forall k \in \mathbb{N}^*, \quad \forall \phi : \mathbb{R}^k \rightarrow \mathbb{R} \text{ mesurable bornée}$$
$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{p=0}^{N-1} \phi(X_{p+1}, \dots, X_{p+k}) = \mathbb{E}[\phi(X_1, \dots, X_k)] \quad \mathbb{P}\text{-p.s.}$$

Pour qu'un processus stationnaire X soit ergodique, il suffit essentiellement que X_t et X_s soient asymptotiquement indépendants quand $|t - s|$ tend vers l'infini. C'est évidemment le cas si X est une famille de variables aléatoires i.i.d. Un contre-exemple simple est donné par le processus $(Y_n = \varepsilon_1)_{n \in \mathbb{N}}$: il est stationnaire, mais

$$\begin{aligned} & \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{p=1}^N \phi(Y_p(\omega), \dots, Y_{p+n}(\omega)) \\ &= \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{p=1}^N \phi(\varepsilon_1(\omega), \dots, \varepsilon_1(\omega)) \\ &= \phi(\varepsilon_1(\omega), \dots, \varepsilon_1(\omega)) \neq \mathbb{E}[\phi(\varepsilon_1, \dots, \varepsilon_1)] \end{aligned}$$

Théorème 3.I.1 *Si X est un processus gaussien stationnaire, et si sa fonction d'autocorrélation tend vers 0 en l'infini, alors sa partie déterministe est constante et il est ergodique.*

3.II Estimation des caractéristiques d'un processus faiblement stationnaire

Considérons une série temporelle $x = (x_t, t \in \{1, \dots, N\})$ que l'on suppose être une réalisation d'un processus faiblement stationnaire $X = (X_t, t \in \mathbb{Z})$. On se pose ici le problème de l'estimation des paramètres

“du second ordre” décrivant ce processus, à savoir

1. sa moyenne $\mu = \mathbb{E}[X_t]$ et sa variance $\gamma_X(0)$
2. sa fonction d'autocorrélation ρ_X
3. sa fonction d'autocorrélation partielle r_X
4. les coefficients de sa décomposition de Wold $(\psi_u, u \in \mathbb{N})$
5. la variance du bruit blanc d'innovation σ^2

Les estimateurs présentés ici sont construits grâce aux méthodes de substitution. Il s'agit d'une estimation préliminaire. Si l'on suppose de plus que X est un processus ARMA, d'autres techniques d'estimation peuvent être utilisées. Elles seront vues ultérieurement.

Notations : Si $T_X = T(X_1, \dots, X_N)$ désigne un estimateur, alors on notera $T_x = T(x_1, \dots, x_N)$ la valeur prise par cette variable aléatoire au point $(X_1 = x_1, \dots, X_N = x_N)$.

3.II.A Estimation de la moyenne du processus

L'estimateur usuel de la moyenne μ_X du processus X est la moyenne empirique

$$\bar{X}_N = \frac{1}{N} \sum_{t=1}^N X_t$$

C'est la variable aléatoire qui minimise

$$Y \mapsto \mathbb{E} \left[\sum_{t=1}^N (X_t - Y)^2 \right]$$

Si X est un bruit blanc gaussien décentré, \bar{X}_N est l'estimateur UVMB de μ_X , c'est aussi l'estimateur du maximum de vraisemblance.

Théorème 3.II.1 *Soit X un processus faiblement stationnaire de moyenne μ_X . Alors*

1. \bar{X}_N est un estimateur sans biais de μ_X .
2. si $\lim_{t \rightarrow +\infty} \gamma_X(t) = 0$, alors $\lim_{N \rightarrow +\infty} \text{var}(\bar{X}_N) = 0$; il en résulte que \bar{X}_N converge vers μ_X au sens de la norme hilbertienne et qu'il est un estimateur faiblement consistant.
3. si $\sum_{t=0}^{+\infty} |\gamma_X(t)| < +\infty$, alors

$$\lim_{N \rightarrow +\infty} N \text{var}(\bar{X}_N) = \gamma_X(0) + 2 \sum_{t=1}^{+\infty} \gamma_X(t).$$

Démonstration du théorème 3.II.1 Le point 1 est évident. Le second est une conséquence de l'égalité suivante :

$$\begin{aligned} \text{var}(\bar{X}_N) &= \frac{1}{N^2} \sum_{s,t=1}^N \gamma_X(|s-t|) \\ &= \frac{1}{N^2} \sum_{u=-N+1}^{N-1} (N-|u|) \gamma_X(|u|) \\ &\leq \frac{2}{N} \sum_{u=0}^{N-1} |\gamma_X(u)| \end{aligned}$$

Or un résultat classique sur les moyennes de Césaro nous apprend que si $|\gamma_X(N)|$ tend vers 0 quand N tend vers l'infini, alors $\frac{1}{N} \sum_{u=0}^{N-1} |\gamma_X(u)|$ converge aussi vers 0. Cela achève de démontrer le point 2. Montrons le point 3 : on a

$$\frac{1}{N} \text{var} \left(\sum_{u=1}^N X_u \right) = \frac{1}{N} \sum_{m,n=1}^N \gamma_X(|m-n|) = \sum_{p=-N+1}^{N-1} \frac{N-p}{N} \gamma_X(|p|) = \sum_{p=-\infty}^{+\infty} \alpha_{p,N} \gamma_X(|p|)$$

avec $\alpha_{p,N} = \frac{N-p}{N}$ si $|p| < N$ et $\alpha_{p,N} = 0$ sinon. Pour tout N , on a $|\alpha_{p,N} \gamma_X(|p|)| \leq |\gamma_X(|p|)|$ et $\sum_{t=0}^{+\infty} |\gamma_X(t)| < +\infty$. On peut donc appliquer le théorème de convergence dominée, et en déduire

$$\begin{aligned} \lim_{N \rightarrow +\infty} N \text{var}(\bar{X}_N) &= \lim_{N \rightarrow +\infty} \sum_{p=-\infty}^{+\infty} \alpha_{p,N} \gamma_X(|p|) \\ &= \sum_{p=-\infty}^{+\infty} \lim_{N \rightarrow +\infty} \alpha_{p,N} \gamma_X(|p|) \\ &= \sum_{p=-\infty}^{+\infty} \gamma_X(|p|) \end{aligned}$$

□

Corollaire 3.II.2 *Théorème central limite. Soit X un processus fortement stationnaire vérifiant*

$$X = \mu_X + \psi(B)(\varepsilon)$$

avec ε bruit blanc fort de variance σ^2 et $\psi(B)$ filtre linéaire. Alors $\sqrt{N}(\bar{X}_N - \mu_X)$ converge en loi vers une gaussienne centrée de variance $\gamma_X(0) + 2 \sum_{t=1}^{+\infty} \gamma_X(t)$.

Démonstration du corollaire 3.II.2 dans le cas où ε est un bruit blanc gaussien. Tout processus X de la forme $\mu_X + \psi(B)(\varepsilon)$ vérifie l'hypothèse $\sum_{t=0}^{+\infty} |\gamma_X(t)| < +\infty$. Il résulte alors du théorème 3.II.1 que $\sqrt{N}(\bar{X}_N - \mu_X)$ est une gaussienne centrée dont la variance converge quand N tend vers l'infini. Un résultat très classique nous permet d'en conclure que la suite de gaussiennes $(\sqrt{N}(\bar{X}_N - \mu_X))$ converge en loi quand N tend vers l'infini, vers une gaussienne centrée de variance la limite des variances (il suffit, pour s'en convaincre, d'étudier la convergence des fonctions caractéristiques). □

Les résultats précédents permettent de construire des statistiques de tests asymptotiques, pour tester si μ_X est nul ou non.

3.II.B Autocovariance et autocorrélation empiriques

Plusieurs estimateurs empiriques peuvent être définis pour l'autocovariance d'un processus X faiblement stationnaire. Le plus usuel est

$$\hat{\gamma}_X^{(N)}(h) = \begin{cases} \frac{1}{N} \sum_{t=1}^{N-h} (X_t - \bar{X}_N)(X_{t+h} - \bar{X}_N) & \text{si } 0 \leq h \leq N \\ 0 & \text{sinon} \end{cases}$$

En pratique, on évite de calculer $\hat{\gamma}_X^{(N)}(h)$ pour des valeurs de h supérieures à $N/4$ (certains préconisent plutôt $10 \ln_{10} N$).

On définit une variante de cet estimateur par

$$\hat{\gamma}_X^{(N)'}(h) = \begin{cases} \frac{1}{N-h} \sum_{t=1}^{N-h} (X_t - \bar{X}_N)(X_{t+h} - \bar{X}_N) & \text{si } 0 \leq h \leq N \\ 0 & \text{sinon} \end{cases}$$

Théorème 3.II.3 *Soit X un processus faiblement stationnaire.*

1. $(\hat{\gamma}_X^{(N)}(|h|), h \in \mathbb{Z})$ est de type positif.

2. Si de plus $\sum_{t=0}^{+\infty} |\gamma_X(t)| < +\infty$, alors

$$\lim_{N \rightarrow +\infty} N \mathbb{E}[\hat{\gamma}_X^{(N)}(h) - \gamma_X(h)] = - \left(\gamma_X(0) + 2 \sum_{t=1}^{+\infty} \gamma_X(t) \right) - h \gamma_X(h).$$

Même résultat pour $\hat{\gamma}_X^{(N)'}(h)$.

Démonstration du théorème 3.II.3 Pour montrer le point 1, définissons, pour tout $t \in \mathbb{Z}$, $X'_t = X_t$ si $t = 1, \dots, N$, et $X'_t = \bar{X}_N$ sinon. Soient $p \geq 1$, $t_1, \dots, t_p \in \mathbb{Z}$ et $\alpha_1, \dots, \alpha_p \in \mathbb{C}$ quelconques; il s'agit de montrer que $\sum_{i,j=1}^p \alpha_i \bar{\alpha}_j \hat{\gamma}_X^{(N)}(t_i - t_j)$ est positif :

$$\begin{aligned} \sum_{i,j=1}^p \alpha_i \bar{\alpha}_j \hat{\gamma}_X^{(N)}(|t_i - t_j|) &= \sum_{i,j=1}^p \frac{\alpha_i \bar{\alpha}_j}{N} \left(\sum_{t=1}^{N-|t_i-t_j|} (X_t - \bar{X}_N)(X_{t+|t_i-t_j|} - \bar{X}_N) \right) \\ &= \sum_{i,j=1}^p \frac{\alpha_i \bar{\alpha}_j}{N} \left(\sum_{t \in \mathbb{Z}} (X'_t - \bar{X}_N)(X'_{t+|t_i-t_j|} - \bar{X}_N) \right) \\ &= \sum_{i,j=1}^p \frac{\alpha_i \bar{\alpha}_j}{N} \left(\sum_{t \in \mathbb{Z}} (X'_{t+t_j} - \bar{X}_N)(X'_{t+t_i} - \bar{X}_N) \right) \\ &= \frac{1}{N} \sum_{t \in \mathbb{Z}} \left(\sum_{i,j=1}^p \alpha_i \bar{\alpha}_j (X'_{t+t_j} - \bar{X}_N)(X'_{t+t_i} - \bar{X}_N) \right) \\ &= \frac{1}{N} \sum_{t \in \mathbb{Z}} \left| \sum_{i=1}^p \alpha_i (X'_{t+t_i} - \bar{X}_N) \right|^2 \geq 0 \end{aligned}$$

Le point 2 est également purement calculatoire, et ne présente pas de mystère; pour simplifier les formules, supposons simplement que X est centré :

$$\begin{aligned} N \mathbb{E} \left[\hat{\gamma}_X^{(N)}(h) - \gamma_X(h) \right] &= \sum_{t=1}^{N-h} (\mathbb{E}[X_t X_{t+h}] - \mathbb{E}[\bar{X}_N X_t] - \mathbb{E}[\bar{X}_N X_{t+h}] + \mathbb{E}[\bar{X}_N^2]) - N \gamma_X(h) \\ &= (N-h) \gamma_X(h) - N \mathbb{E}[\bar{X}_N^2] \\ &\quad + \sum_{t=N-h+1}^N \mathbb{E}[\bar{X}_N X_t] - N \mathbb{E}[\bar{X}_N^2] + \sum_{t=1}^h \mathbb{E}[\bar{X}_N X_t] + (N-h) \mathbb{E}[\bar{X}_N^2] - N \gamma_X(h) \\ &= -h \gamma_X(h) - N \mathbb{E}[\bar{X}_N^2] + \left(\sum_{t=N-h+1}^N \mathbb{E}[\bar{X}_N X_t] + \sum_{t=1}^h \mathbb{E}[\bar{X}_N X_t] - h \mathbb{E}[\bar{X}_N^2] \right) \end{aligned}$$

On en déduit le résultat car $N\mathbb{E}[\bar{X}_N^2] = N \text{var}(\bar{X}_N)$ tend vers $\gamma_X(0) + 2 \sum_{t=1}^{+\infty} \gamma_X(t)$, et que le dernier terme est une somme finie, bornée, d'espérances qui tendent toutes vers 0 car $\text{var}(\bar{X}_N)$ tend vers 0. \square

Remarque : Si la moyenne μ_X du processus X est connue - en particulier si le processus est centré -, alors on définit plutôt $\hat{\gamma}_X^{(N)}$ et $\hat{\gamma}_X^{(N) \prime}$ par

$$\begin{aligned}\hat{\gamma}_X^{(N)}(h) &= \frac{1}{N} \sum_{t=1}^{N-h} (X_t - \mu_X)(X_{t+h} - \mu_X) \\ \hat{\gamma}_X^{(N) \prime}(h) &= \frac{1}{N-h} \sum_{t=1}^{N-h} (X_t - \mu_X)(X_{t+h} - \mu_X)\end{aligned}$$

Dans ce cas, $\hat{\gamma}_X^{(N) \prime}(h)$ est un estimateur sans biais de $\gamma_X(h)$.

Théorème 3.II.4 Soit X un processus faiblement stationnaire tel que

1. $\mathbb{E}[X_t^4] = cte < +\infty$;
2. $\text{cov}(X_t X_{t+k}, X_{t+m} X_{t+m+k})$ est indépendant de t ;
3. $\lim_{m \rightarrow +\infty} \text{cov}(X_t X_{t+k}, X_{t+m} X_{t+m+k}) = 0$;

alors $\hat{\gamma}_X^{(N)}(k)$ converge au sens hilbertien vers $\gamma_X(k)$ quand $N \rightarrow +\infty$. Si, en outre, X est un processus gaussien, et si $\sum_{t=0}^{+\infty} |\gamma_X(t)| < +\infty$, alors

$$\lim_{N \rightarrow +\infty} N \text{cov}(\hat{\gamma}_X^{(N)}(h), \hat{\gamma}_X^{(N)}(k)) = \sum_{u=-\infty}^{+\infty} (\gamma_X(u)\gamma_X(u+h-k) + \gamma_X(u-k)\gamma_X(u+h))$$

Remarque : Sous les hypothèses de la dernière assertion, on a en particulier que

$$\lim_{N \rightarrow +\infty} N \text{var}(\hat{\gamma}_X^{(N)}(h)) = \sum_{u=-\infty}^{+\infty} (\gamma_X(u)^2 + \gamma_X(u-h)\gamma_X(u+h))$$

ce qui implique la convergence en moyenne quadratique de $\hat{\gamma}_X^{(N)}(h)$ vers $\gamma_X(h)$.

L'estimateur empirique de l'autocorrélation est

$$\hat{\rho}_X^{(N)}(k) = \frac{\hat{\gamma}_X^{(N)}(k)}{\hat{\gamma}_X^{(N)}(0)}$$

L'estimateur $r_X^{(N)}$ de la fonction d'autocorrélation partielle est défini à partir des équations de Yule-Walker, où l'on a remplacé la fonction d'autocorrélation par son estimateur.

Théorème 3.II.5 Soit ε un bruit blanc (au sens fort) tel que $\mathbb{E}[\varepsilon_t^4] = cte < +\infty$. Considérons $\phi(B)$ un filtre linéaire, μ_X une constante, et X le processus stationnaire défini par $X = \mu_X + \phi(B)(\varepsilon)$. Alors

- si X est un MA(q), $\sqrt{N}\hat{\rho}_X^{(N)}(k)$ converge en loi vers une gaussienne centrée de variance $1 + 2 \sum_{u=1}^q \rho_X^2(u)$ pour tout $k > q$;

- si X est un $AR(p)$, $\sqrt{N} \hat{r}_X^{(N)}(k)$ converge en loi vers une gaussienne centrée de variance 1 pour tout $k > p$.

On déduit de ces comportements asymptotiques différents tests sur les paramètres du processus X .

Exercice 1. 1) Soit $X = (X_t, t \in \mathbb{Z})$ un processus stationnaire vérifiant

$$X_t = \varepsilon_t + \phi X_{t-12}$$

avec ε bruit blanc. Déterminer sa fonction d'autocorrélation. (On pourra établir une relation de récurrence entre $\rho_X(t)$ et $\rho_X(t+12)$, et la résoudre.)

2) Soit $x = (x_1, \dots, x_N)$ une série temporelle dont la fonction d'autocorrélation estimée a l'allure suivante :

$$\hat{\rho}_x(12j) \approx (0,7)^j, \quad j = 0, 1, \dots \quad (3.1)$$

$$\hat{\rho}_x(12j \pm 1) \approx 0,4 \times (0,7)^j, \quad j = 0, 1, \dots \quad (3.2)$$

$$\hat{\rho}_x(h) \approx 0 \text{ sinon.} \quad (3.3)$$

Proposer, en le justifiant, un modèle $ARMA(12,1)$, pour cette série.

3.II.C Algorithme des innovations

Si X est un processus faiblement stationnaire centré et régulier, sa décomposition de Wold s'écrit

$$X_t = \varepsilon_t + \sum_{u=1}^{+\infty} \psi_u \varepsilon_{t-u}$$

avec ε bruit blanc d'innovation. Les paramètres de cette décomposition sont les coefficients $(\psi_u, u \geq 1)$ et la variance σ^2 de ε_t . Pour les déterminer, il semble naturel de calculer $\varepsilon_t = X_t - P_{X_{t-1}, X_{t-2}, \dots}^\perp(X_t)$, et d'en déduire

$$\begin{aligned} \sigma^2 &= \text{var}(\varepsilon_t) \\ \psi_i &= \text{cov}(X_{t+i}, \varepsilon_t) / \sigma^2 \end{aligned}$$

Or on ne peut calculer $P_{X_{t-1}, X_{t-2}, \dots}^\perp(X_t)$, mais seulement les approximations $P_{X_{t-1}, \dots, X_1}^\perp(X_t)$ (grâce aux équations de Yule-Walker par exemple). Notons $\tilde{\varepsilon}_t = X_t - P_{X_{t-1}, \dots, X_1}^\perp(X_t)$ et $\tilde{\sigma}_t^2 = \text{var}(\tilde{\varepsilon}_t)$. La famille $(\tilde{\varepsilon}_t / \tilde{\sigma}_t, t = 1, \dots, N)$ n'est autre que la base orthonormée de $V^2(X_t, t = 1, \dots, N)$ construite par le procédé de Schmidt. En particulier, on a la propriété :

$$\forall t = 1, \dots, N, \quad V^2(X_1, \dots, X_t) = V^2(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_t).$$

C'est l'analogie fini-dimensionnel de la propriété classique vérifiée par le bruit blanc d'innovation d'un processus $ARMA$ (ou d'un processus faiblement stationnaire régulier) :

$$\forall t \in \mathbb{Z}, \quad V^2(X_s, s \leq t) = V^2(\varepsilon_s, s \leq t).$$

Posons pour tout $t = 1, \dots, N$

$$\psi_{t,i} = \frac{\text{cov}(X_t, \tilde{\varepsilon}_{t-i})}{\tilde{\sigma}_{t-i}^2} = \frac{\langle X_t, \tilde{\varepsilon}_{t-i} \rangle}{\|\tilde{\varepsilon}_{t-i}\|_2^2};$$

alors $\psi_{t,0} = 1$, et

$$X_t = \tilde{\varepsilon}_t + \sum_{i=1}^{t-1} \psi_{t,i} \tilde{\varepsilon}_{t-i}.$$

Rappelons que dans le cas d'un processus ARMA (ou d'un processus faiblement stationnaire régulier), de bruit blanc d'innovation ε , on a

$$X_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i},$$

avec

$$\psi_i = \frac{\text{cov}(X_t, \varepsilon_{t-i})}{\sigma^2}.$$

Remarque. La famille de variables aléatoires $(\tilde{\varepsilon}_t, t \geq 1)$ peut être vue comme le processus d'innovation associé au processus $(Y_t, t \in \mathbb{Z})$ avec $Y_t = 0$ si $t \leq 0$, et $Y_t = X_t$ si $t \geq 0$. Le processus Y n'est pas faiblement stationnaire, ce qui explique pourquoi $\tilde{\text{tild}}e\varepsilon$ n'est pas un bruit blanc.

Théorème 3.II.6 Avec les notations et les hypothèses précédentes, on a

$$\begin{aligned} \lim_{N \rightarrow +\infty} \|\varepsilon_N - \tilde{\varepsilon}_N\|_2 &= 0 \\ \lim_{N \rightarrow +\infty} \psi_{N,i} &= \psi_i \\ \lim_{N \rightarrow +\infty} \tilde{\sigma}_N^2 &= \sigma^2 \end{aligned}$$

Démonstration du théorème 3.II.6 En vertu de la stationnarité faible de X , il est facile de vérifier l'égalité :

$$\begin{aligned} \|\varepsilon_N - \tilde{\varepsilon}_N\|_2 &= \left\| P_{V^2(X_s, s < N)}^\perp(X_N) - P_{V^2(X_s, 1 \leq s < N)}^\perp(X_N) \right\|_2 \\ &= \left\| P_{V^2(X_s, s < 0)}^\perp(X_0) - P_{V^2(X_s, -N \leq s < 0)}^\perp(X_0) \right\|_2 \end{aligned}$$

Or $V^2(X_s, s < 0) = \bigcup_{N > 0} V^2(X_s, -N \leq s < 0)$. D'après le corollaire 5.II.2, on en déduit que $P_{V^2(X_s, -N \leq s < 0)}^\perp(X_0)$ converge au sens de la norme hilbertienne vers $P_{V^2(X_s, s < 0)}^\perp(X_0)$ quand N tend vers l'infini. On en déduit le premier point, à savoir

$$\lim_{N \rightarrow +\infty} \|\varepsilon_N - \tilde{\varepsilon}_N\|_2 = 0$$

Les deux autres limites en découlent, car $\tilde{\sigma}_N^2 = \|\tilde{\varepsilon}_N\|_2^2$ a même limite que $\|\varepsilon_N\|_2^2 = \sigma^2$; d'autre part,

$$|\langle X_N, \varepsilon_N \rangle - \langle X_N, \tilde{\varepsilon}_N \rangle| \leq \|X_N\|_2 \|\varepsilon_N - \tilde{\varepsilon}_N\| \rightarrow_{N \rightarrow +\infty} 0$$

d'où il résulte que

$$\psi_{N,i} = \frac{\langle X_N, \tilde{\varepsilon}_N \rangle}{\tilde{\sigma}_{N-i}}$$

converge vers

$$\frac{\langle X_N, \varepsilon_N \rangle}{\sigma} = \psi_i$$

□

En pratique, le procédé d'orthonormalisation de Schmidt permet le calcul par récurrence de $(\tilde{\sigma}_t^2, t = 1, \dots, N)$ et de $(\psi_{t,i}, 0 \leq i \leq t \leq N)$:

$$\begin{aligned} \psi_{t,t-j} &= \frac{1}{\tilde{\sigma}_j^2} \left(\gamma_X(t-j) - \sum_{u=1}^{j-1} \psi_{t,t-u} \psi_{j,j-u} \tilde{\sigma}_u^2 \right), \quad j = 1, \dots, t \\ \tilde{\sigma}_t^2 &= \gamma_X(0) - \sum_{u=1}^{t-1} \psi_{t,t-u}^2 \tilde{\sigma}_u^2 \end{aligned}$$

avec $\tilde{\sigma}_1^2 = \gamma_X(0)$. Cette méthode porte en statistique des processus le nom d'**algorithme des innovations**. En remplaçant dans l'algorithme précédent la fonction d'autocovariance γ_X par son estimateur $\hat{\gamma}_X^{(N)}$, on définit les estimateurs $\hat{\psi}_{t,i}^{(N)}$ et $\hat{\sigma}_t^{2(N)}$.

Théorème 3.II.7 Si X est un processus ARMA, dont l'innovation ε est un bruit blanc au sens fort, possédant un moment d'ordre 4, alors, en reprenant les notations précédentes, pour tout $k \geq 1$ et pour toute suite d'entiers positifs $(m_N, N \geq 1)$ telle que $m_N < N$, $\lim_{N \rightarrow +\infty} m_N = +\infty$ et $\lim_{N \rightarrow +\infty} \frac{m_N}{N^{1/3}} = 0$, le vecteur aléatoire

$$\sqrt{N}^t (\hat{\psi}_{m_N,1}^{(N)} - \psi_1, \dots, \hat{\psi}_{m_N,k}^{(N)} - \psi_k)$$

converge en loi vers un vecteur gaussien centré de matrice de covariance $A = (a_{i,j})_{i,j=1,\dots,k}$ telle que $a_{i,j} = \sum_{u=1}^{\min(i,j)} \psi_{i-u} \psi_{j-u}$. De plus, $\hat{\sigma}_{m_N}^{2(N)}$ converge en probabilité vers σ^2 .

Exercice 2. Soit $X = (X_t, t \in \mathbb{Z})$ un processus ARMA(1,1), vérifiant $X_t + \phi X_{t-1} = \varepsilon_t + \theta \varepsilon_{t-1}$, avec ε bruit blanc de variance σ_ε^2 , et $|\theta|, |\phi| < 1$.

1. On pose $\tilde{\varepsilon}_1 = X_1$ et $\tilde{\varepsilon}_t = X_t - P_{V^2(1, X_1, \dots, X_{t-1})}^\perp(X_t)$ quel que soit $t > 1$. Montrer que pour tout $t > 1$ il existe $\theta_{t,1}$ tel que

$$X_t + \phi X_{t-1} = \tilde{\varepsilon}_t + \theta_{t,1} \tilde{\varepsilon}_{t-1}$$

2. En déduire la relation de récurrence

$$\begin{cases} \delta_1 = \frac{\gamma_X(0)}{\sigma_\varepsilon^2} \\ \delta_{t+1} = 1 + \theta^2 - \frac{\theta^2}{\delta_t} \end{cases}$$

avec $\delta_t = \mathbb{E}[\tilde{\varepsilon}_t^2] / \sigma_\varepsilon^2$.

3. Etablir une relation de récurrence simple définissant la suite $(\tilde{\varepsilon}_i)_{1 \leq i \leq N}$.
4. (Question portant sur le chapitre suivant) Calculer formellement l'erreur de prédiction au pas h , puis en donner une approximation explicite en fonction de θ et ϕ .

Exercice 3. Soit $X = (X_t, t \in \mathbb{Z})$ un processus ARMA(p, q). On pose $\tilde{\varepsilon}_i = X_i - P_{V^2(1, X_1, \dots, X_{i-1})}^\perp(X_i)$.

- 1) On suppose $p = 0$. Soit $(\psi_{i,j})_{1 \leq j < i \leq N}$ les coefficients déterminés par l'algorithme des innovations, tels que

$$X_i = \sum_{j=0}^{i-1} \psi_{i,j} \tilde{\varepsilon}_{i-j}$$

Montrer que $\psi_{i,j} = 0$ si $j > q$.

- 2) On suppose maintenant $p \neq 0$. On pose pour tout $i \in \{1, \dots, N\}$

$$Y_i = \begin{cases} X_i & \text{si } i \leq \max(p, q) \\ \phi(B)(X)_i & \text{sinon.} \end{cases}$$

- a) Montrer que $V^2(1, X_1, \dots, X_i) = V^2(1, Y_1, \dots, Y_i)$ quel que soit $i \in \{1, \dots, N\}$.
- b) Montrer que $\tilde{\varepsilon}_i = Y_i - P_{V^2(1, Y_1, \dots, Y_{i-1})}^\perp(Y_i)$ quel que soit $i \in \{1, \dots, N\}$.
- c) En déduire une récursion simple pour le calcul des $\tilde{\varepsilon}_i$.

3.III Estimation des paramètres d'un modèle ARMA

Soit $x = (x_t, t \in \{1, \dots, N\})$ une série temporelle qu'on souhaite modéliser non seulement par un processus faiblement stationnaire, mais plus précisément par un processus ARMA(p, q) caractérisé par

$$\phi(B)(X) = \theta(B)(\varepsilon)$$

avec

$$\begin{aligned}\theta(z) &= 1 + \theta_1 z + \dots + \theta_q z^q \\ \phi(z) &= 1 + \phi_1 z + \dots + \phi_p z^p \\ \varepsilon &\text{ bruit blanc de variance } \sigma_\varepsilon^2\end{aligned}$$

L'étude consiste à déterminer d'abord les couples (p, q) possibles, pour chacun d'entre eux à estimer les paramètres du modèle ARMA associé, à tester les modèles obtenus, et parmi ceux retenus, à choisir celui qui a le meilleur pouvoir prédictif ou informatif.

3.III.A Estimation de p et q

Cas d'un processus MA(q)

Soit $X = (X_t, t \in \mathbb{Z})$ un processus MA(q). D'après le théorème 3.II.5, pour tout $r > q$, alors l'autocorrélation empirique $\rho_X^{(N)}(r)$ est approximativement une gaussienne centrée de variance

$$\frac{1}{N} \left(1 + \sum_{u=1}^q \rho_X(u)^2 \right) = \frac{1}{N} \left(1 + \sum_{u=1}^r \rho_X(u)^2 \right)$$

En pratique, on fait donc l'approximation

$$\forall r > q, \quad \sqrt{N} \frac{\rho_X^{(N)}(r)}{\sqrt{1 + \sum_{u=1}^r \rho_X^{(N)}(u)^2}} \approx \mathcal{N}(0, 1).$$

On en déduit un test (pratique mais peu rigoureux) de $H_0 : X$ est un MA(q) ; si pour (presque) tout $r > q$ et $r < N/4$, on a

$$\left| \rho_X^{(N)}(r) \right| < 1,96 \sqrt{\frac{1 + \sum_{u=1}^r \rho_X^{(N)}(u)^2}{N}},$$

alors on accepte H_0 . Si l'on se limite à $r < N/4$, c'est que l'estimation de $\rho_X^{(N)}(r)$ devient de moins en moins précise quand r croît vers N .

Cas d'un processus AR(p)

Soit $X = (X_t, t \in \mathbb{Z})$ un processus AR(p). D'après le théorème 3.II.5, pour tout $r > q$, alors l'autocorrélation partielle empirique $r_X^{(N)}(r)$ est approximativement une gaussienne centrée de variance $1/N$. On en déduit comme précédemment un test de $H_0 : X$ est un AR(p) ; si pour (presque) tout $r > q$ et $r < N/4$, on a

$$\left| r_X^{(N)}(r) \right| < 1,96 \sqrt{\frac{1}{N}},$$

alors on accepte H_0 .

Cas d'un processus ARMA(p, q)

On pourrait s'inspirer du critère du coin pour construire un test sur les paramètres p et q . En pratique, on procède différemment et plus simplement en vertu de la remarque suivante : si X un processus ARMA(p, q) vérifiant

$$\phi(B)(X) = \theta(B)(\varepsilon)$$

avec $\theta(z)$ et $\phi(z)$ sans racine commune, et ε bruit blanc d'innovation de variance σ^2 , alors il peut se mettre sous forme MA(∞)

$$X_t = \varepsilon_t + \sum_{u \geq 1} \psi_u \varepsilon_{t-u}$$

ou AR(∞)

$$X_t + \sum_{v \geq 1} \eta_v X_{t-v} = \varepsilon_t$$

Comme on a les inégalités

$$\mathbb{E} \left[\left(X_t - \varepsilon_t - \sum_{u=1}^{q'} \psi_u \varepsilon_{t-u} \right)^2 \right] \leq \sigma^2 \left(\sum_{u \geq q'+1} |\psi_u| \right)^2$$

$$\mathbb{E} \left[\left(X_t - \varepsilon_t + \sum_{v \geq 1} \eta_v X_{t-v} \right)^2 \right] \leq \gamma_X(0) \left(\sum_{u \geq p'+1} |\eta_u| \right)^2$$

et que $\sum_{u \geq 0} |\psi_u|$ comme $\sum_{u \geq 0} |\eta_u|$ sont finis, on en déduit qu'il est possible d'approcher d'aussi près qu'on le souhaite X par un processus de type MA(q') ou AR(p') avec $p' \geq p$ et $q' \geq q$. En pratique, on choisit pour valeurs de p' et q' celles que l'on détermine par les méthodes vues précédemment, en supposant le processus AR puis MA. Cela fournit une première modélisation de X comme processus ARMA(p', q'), avec en général plusieurs couples (p', q') possibles. On verra par la suite comment améliorer cette première estimation-majoration des paramètres de "taille" du processus ARMA.

3.III.B Estimation des paramètres du filtre ARMA et de la variance du bruit blanc

Une fois que l'on a décidé de chercher pour la série temporelle x un modèle ARMA(p, q), reste à estimer les coefficients qui le caractérisent. On procède en deux temps, une première estimation empirique étant ensuite affinée par des techniques fondées sur le maximum de vraisemblance.

Estimation préliminaire

Estimation de la partie autorégressive. Si X est un processus ARMA(p, q) causal et inversible, vérifiant $\phi(B)(X) = \theta(B)(\varepsilon)$, alors on a vu que sa fonction d'autocovariance vérifie la relation de récurrence suivante :

$$\forall t \geq q+1 \quad \gamma_X(t) + \phi_1 \gamma_X(t-1) + \dots + \phi_p \gamma_X(t-p) = 0,$$

et que l'on peut en déduire le système d'équations dit aussi de Yule-Walker

$$\begin{pmatrix} \rho_X(q) & \dots & \rho_X(q-p+1) \\ \vdots & & \vdots \\ \rho_X(q+p-1) & \dots & \rho_X(q) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} = - \begin{pmatrix} \rho_X(q+1) \\ \vdots \\ \rho_X(q+p) \end{pmatrix}$$

Les estimateurs de Yule-Walker de ϕ_1, \dots, ϕ_p sont donc naturellement définis par

$$\begin{pmatrix} \hat{\rho}_X^{(N)}(q) & \dots & \hat{\rho}_X^{(N)}(q-p+1) \\ \vdots & & \vdots \\ \hat{\rho}_X^{(N)}(q+p-1) & \dots & \hat{\rho}_X^{(N)}(q) \end{pmatrix} \begin{pmatrix} \hat{\phi}_1^{(N)} \\ \vdots \\ \hat{\phi}_p^{(N)} \end{pmatrix} = - \begin{pmatrix} \hat{\rho}_X^{(N)}(q+1) \\ \vdots \\ \hat{\rho}_X^{(N)}(q+p) \end{pmatrix}$$

Estimation de la partie moyenne mobile et de la variance du bruit blanc. On s'appuie sur la remarque simple suivante : $Y = \phi(B)(X)$ est un processus MA(q), de paramètres $\theta_1, \dots, \theta_q$.

Une fois calculé $\hat{\phi}^{(N)}(z) = 1 + \hat{\phi}_1^{(N)}z + \dots + \hat{\phi}_p^{(N)}z^p$, on pose

$$Y = \hat{\phi}^{(N)}(B)(X),$$

et l'on détermine les estimateurs $\hat{\theta}_1^{(N)}, \dots, \hat{\theta}_q^{(N)}$ de $\theta_1, \dots, \theta_q$ en appliquant à Y l'algorithme des innovations. On en déduit également l'estimateur $\sigma^{(N)2}$ de la variance du bruit blanc σ^2 .

Estimateurs des moindres carrés et du maximum de vraisemblance

Les estimateurs précédents sont assez grossiers. Ils servent essentiellement à initialiser les procédés de calcul numérique (par descente de gradient et approximations successives) des estimateurs plus élaborés qui sont vus maintenant.

Notations. Soit $X = (X_t, t \in \mathbb{Z})$ un processus ARMA(p, q) tel que

$$\phi(B)(X) = \theta(B)(\varepsilon)$$

avec ε bruit blanc gaussien de variance σ_ε^2 . Notons $\phi = (\phi_1, \dots, \phi_p)$ et $(\theta_1, \dots, \theta_q)$. La fonction d'autocorrélation de X est une fonction de ϕ et θ , qu'on notera $\rho_{(\theta, \phi)}$. De même, la fonction d'autocorrélation partielle de X , qui se calcule à partir de la fonction d'autocorrélation, est-elle une fonction de ϕ et θ , qu'on notera $r_{(\theta, \phi)}$. La matrice de corrélation d'ordre k sera notée $R_{(\theta, \phi)}^{(k)} = (\rho_{(\theta, \phi)}(|i-j|))_{i,j=1}^k$, et la matrice de covariance par $\Gamma_{(\theta, \phi, \sigma^2)}^{(k)} = \gamma_X(0)R_{(\theta, \phi)}^{(k)}$.

Soit $(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N)$ la base orthogonale de $V^2(X_1, \dots, X_N)$ construite par l'algorithme des innovations :

$$\tilde{\varepsilon}_t = X_t - P_{V^2(X_1, \dots, X_{t-1})}^\perp(X_t)$$

Il existe des coefficients $\psi_{i,j}$ et $\eta_{i,j}$ tels que pour tout $t = 1, \dots, N$ et $i = 1, \dots, t-1$, on ait

$$\begin{aligned} X_t &= \tilde{\varepsilon}_t + \sum_{i=1}^{t-1} \psi_{t,i} \tilde{\varepsilon}_{t-i} \\ \tilde{\varepsilon}_t &= X_t + \sum_{i=1}^{t-1} \eta_{t,i} X_{t-i} \end{aligned}$$

Là encore, ces coefficients ne dépendent que de θ et ϕ et on les notera plutôt $\psi_{i,j}^{(\theta, \phi)}$ et $\eta_{i,j}^{(\theta, \phi)}$. En effet,

$$P_{V^2(X_1, \dots, X_{t-1})}^\perp(X_t) = - \sum_{i=1}^{t-1} \eta_{t,i} X_{t-i},$$

ce qui implique que les coefficients $\eta_{t,i}$ sont déterminés par les équations de Yule-Walker

$$R_{(\theta, \phi)}^{(t-1)} \begin{pmatrix} -\eta_{t,1} \\ \vdots \\ -\eta_{t,t-1} \end{pmatrix} = \begin{pmatrix} \rho_{(\theta, \phi)}(1) \\ \vdots \\ \rho_{(\theta, \phi)}(t) \end{pmatrix}$$

Les coefficients $\eta_{i,j}$ sont donc bien fonction de θ et φ uniquement. Quant aux coefficients $\psi_{i,j}$, ils peuvent être calculés à partir des $\eta_{i,j}$. En effet, notons $S_{(\theta, \phi)}^{(N)}$ et $T_{(\theta, \phi)}^{(N)}$ les matrices triangulaires supérieures

$$S_{(\theta, \phi)}^{(N)} = \begin{pmatrix} 1 & \psi_{2,1}^{(\theta, \phi)} & \psi_{3,2}^{(\theta, \phi)} & \dots & \psi_{N,N-1}^{(\theta, \phi)} \\ & 1 & \psi_{3,1}^{(\theta, \phi)} & & \psi_{N,N-2}^{(\theta, \phi)} \\ & & \ddots & & \vdots \\ & & & 1 & \psi_{N,1}^{(\theta, \phi)} \\ & & & & 1 \end{pmatrix}$$

et

$$T_{(\theta,\phi)}^{(N)} = \begin{pmatrix} 1 & \eta_{2,1}^{(\theta,\phi)} & \eta_{3,2}^{(\theta,\phi)} & \cdots & \eta_{N,N-1}^{(\theta,\phi)} \\ & 1 & \eta_{3,1}^{(\theta,\phi)} & & \eta_{N,N-2}^{(\theta,\phi)} \\ & & \ddots & & \vdots \\ & & & 1 & \eta_{N,1}^{(\theta,\phi)} \\ & & & & 1 \end{pmatrix}.$$

Il s'agit bien sûr des matrices de passage entre les deux bases (X_1, \dots, X_N) et $(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N)$:

$$(X_1, \dots, X_N) = (\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N) S_{(\theta,\phi)}^{(N)}$$

et inversement

$$(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N) = (X_1, \dots, X_N) T_{(\theta,\phi)}^{(N)}$$

En particulier, $S_{(\theta,\phi)}^{(N)} = \left(T_{(\theta,\phi)}^{(N)}\right)^{-1}$. Les coefficients $\psi_{i,j}$ sont donc bien fonction des $\eta_{i,j}$ et donc de θ et ϕ . On en déduit aussi que

$$\begin{aligned} \Gamma_{(\theta,\phi,\sigma^2)}^{(N)} &= \mathbb{E} [{}^t(X_1, \dots, X_N) (X_1, \dots, X_N)] \\ &= {}^t S_{(\theta,\phi)}^{(N)} \mathbb{E} [{}^t(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N) (\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N)] S_{(\theta,\phi)}^{(N)} \\ &= {}^t S_{(\theta,\phi)}^{(N)} \cdot \sigma^2 \Delta_{(\theta,\phi)}^{(N)} S_{(\theta,\phi)}^{(N)} \end{aligned}$$

avec $\Delta_{(\theta,\phi)}^{(N)}$ la matrice diagonale

$$\Delta_{(\theta,\phi)}^{(N)} = \begin{pmatrix} \delta_1^{(\theta,\phi)} & & & \\ & \ddots & & \\ & & \delta_N^{(\theta,\phi)} & \\ & & & \delta_N^{(\theta,\phi)} \end{pmatrix} \text{ avec } \delta_i^{(\theta,\phi)} = \frac{\text{var}(\tilde{\varepsilon}_i)}{\sigma^2}.$$

Enfin on notera $\psi_i^{(\theta,\phi)}$ et $\eta_i^{(\theta,\phi)}$ les coefficients apparaissant dans les représentations MA(∞) et AR(∞) de X :

$$\frac{\theta(z)}{\phi(z)} = 1 + \sum_{i=1}^{+\infty} \psi_i^{(\theta,\phi)} z^i = \frac{1}{1 + \sum_{i=1}^{+\infty} \eta_i^{(\theta,\phi)} z^i}$$

Méthode des moindres carrés ordinaires Soit donc $x = (x_1, \dots, x_N)$ une série temporelle qu'on suppose modélisable par X un processus ARMA(p, q) dont on cherche les autres coefficients. La méthode des moindres carrés consiste à choisir les valeurs de θ et ϕ qui minimisent la variance σ^2 du bruit blanc

d'innovation. Or $\left(\frac{\tilde{\varepsilon}_t}{\sqrt{\delta_t^{(\theta,\phi)}}}, t = 1, \dots, N\right)$ est justement un bruit blanc de variance σ^2 . L'estimateur le plus naturel de σ^2 est donc

$$\frac{1}{N} \sum_{t=1}^N \frac{\tilde{\varepsilon}_t^2}{\delta_t^{(\theta,\phi)}}.$$

On pose

$$\tilde{\varepsilon}_t^{(\theta,\phi)}(x) = x_t + \sum_{i=1}^{t-1} \eta_{t,i}^{(\theta,\phi)} x_{t-i},$$

et on choisit comme valeurs de θ et ϕ celles qui minimisent la fonction

$$(\theta, \phi) \mapsto \frac{1}{N} \sum_{t=1}^N \frac{\tilde{\varepsilon}_t^{(\theta,\phi)}(x)^2}{\delta_t^{(\theta,\phi)}} = \frac{1}{N} x \cdot \left(R_{(\theta,\phi)}^{(N)}\right)^{-1} \cdot {}^t x \quad (3.4)$$

Notons $\hat{\theta}$ et $\hat{\phi}$ ces estimateurs, et $\hat{\varepsilon}_t = X_t + \sum_{i=1}^{t-1} \eta_{t,i}^{(\hat{\theta}, \hat{\phi})} X_{t-i}$. L'estimateur de σ^2 est la variable aléatoire

$$\frac{1}{N-p-q} \sum_{t=1}^N \frac{\tilde{\varepsilon}_t^2}{\delta_t^{(\hat{\theta}, \hat{\phi})}}.$$

Notez que la renormalisation est en $1/N-p-q$ et non $1/N$, pour corriger l'écart qui existe entre $\tilde{\varepsilon}_t$ et $\hat{\varepsilon}_t$, et entre $\delta_t^{(\theta, \phi)}$ et $\delta_t^{(\hat{\theta}, \hat{\phi})}$.

Méthode du maximum de vraisemblance Si on suppose de plus que X est gaussien, alors la log-vraisemblance du modèle statistique associé à (X_1, \dots, X_N) au point $x = (x_1, \dots, x_N)$ est égale à

$$\begin{aligned} \ln L(\theta, \phi, \sigma^2; x) &= \ln \left(\frac{1}{\sqrt{(2\pi)^N \det(\Gamma_{(\theta, \phi, \sigma^2)}^{(N)})}} \exp \left(-\frac{1}{2} x \cdot \left(\Gamma_{(\theta, \phi, \sigma^2)}^{(N)} \right)^{-1} \cdot x \right) \right) \\ &= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2} \ln \left(\det \left(R_{(\theta, \phi)}^{(N)} \right) \right) - \frac{1}{2\sigma^2} x \cdot \left(R_{(\theta, \phi)}^{(N)} \right)^{-1} \cdot x \end{aligned}$$

Or

$$\begin{aligned} x \cdot \left(R_{(\theta, \phi)}^{(N)} \right)^{-1} \cdot x &= x \cdot \left({}^t S_{(\theta, \phi)}^{(N)} \cdot \Delta_{(\theta, \phi)}^{(N)} \cdot S_{(\theta, \phi)}^{(N)} \right)^{-1} \cdot x \\ &= x \cdot \left(S_{(\theta, \phi)}^{(N)} \right)^{-1} \cdot \left(\Delta_{(\theta, \phi)}^{(N)} \right)^{-1} \cdot \left({}^t S_{(\theta, \phi)}^{(N)} \right)^{-1} \cdot x \\ &= x \cdot T_{(\theta, \phi)}^{(N)} \cdot \begin{pmatrix} 1/\delta_1^{(\theta, \phi)} & & \\ & \ddots & \\ & & 1/\delta_{(N)}^{(\theta, \phi)} \end{pmatrix} \cdot {}^t T_{(\theta, \phi)}^{(N)} \cdot x \\ &= \sum_{t=1}^N \frac{\tilde{\varepsilon}_t^{(\theta, \phi)}(x)^2}{\delta_t^{(\theta, \phi)}} \\ \det \left(R_{(\theta, \phi)}^{(N)} \right) &= \det \left({}^t S_{(\theta, \phi)}^{(N)} \right) \det \left(\Delta_{(\theta, \phi)}^{(N)} \right) \det \left(S_{(\theta, \phi)}^{(N)} \right) \\ &= \det \left(\Delta_{(\theta, \phi)}^{(N)} \right) = \prod_{t=1}^N \delta_t^{(\theta, \phi)} \end{aligned}$$

La log-vraisemblance est donc égale à

$$\ln L(\theta, \phi, \sigma^2; x) = -\frac{1}{2\sigma^2} \sum_{t=1}^N \frac{\tilde{\varepsilon}_t^{(\theta, \phi)}(x)^2}{\delta_t^{(\theta, \phi)}} - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2} \sum_{t=1}^N \ln \left(\delta_t^{(\theta, \phi)} \right) - \frac{N}{2} \ln(2\pi)$$

On choisit comme valeurs de θ , ϕ et σ^2 celles qui maximisent la fonction

$$(\theta, \phi, \sigma^2) \mapsto \ln L(\theta, \phi, \sigma^2; x).$$

Il ressort facilement des équations de vraisemblance qu'en ce maximum on a

$$\sigma^2 = \frac{1}{N} \sum_{t=1}^N \frac{\tilde{\varepsilon}_t^{(\theta, \phi)}(x)^2}{\delta_t^{(\theta, \phi)}},$$

et que l'on détermine θ et ϕ en maximisant

$$-\frac{N}{2} \ln \left(\sum_{t=1}^N \frac{\tilde{\varepsilon}_t^{(\theta, \phi)}(x)^2}{\delta_t^{(\theta, \phi)}} \right) - \frac{1}{2} \sum_{t=1}^N \ln \left(\delta_t^{(\theta, \phi)} \right);$$

autrement dit, en minimisant

$$\frac{1}{N} \sum_{t=1}^N \frac{\tilde{e}_t^{(\theta, \phi)}(x)^2}{\delta_t^{(\theta, \phi)}} \times \left(\prod_{t=1}^N \delta_t^{(\theta, \phi)} \right)^{\frac{1}{N}}. \quad (3.5)$$

Exercice 4. Soit $X = (X_t, t \in \mathbb{Z})$ un processus stationnaire de type $AR(1)$, vérifiant l'équation

$$X_t = \phi X_{t-1} + \varepsilon_t$$

avec ε un bruit blanc gaussien de variance σ_ε^2 , et $|\phi| < 1$.

1) Déterminer la matrice de covariance du vecteur aléatoire (X_1, X_2) , en fonction de ϕ et σ_ε^2 . En déduire la densité de la loi de (X_1, X_2) .

On rappelle que si $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \neq 0$, alors $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

2) On suppose connues deux observations x_1 et x_2 du processus X aux instants 1 et 2. Déterminer les estimateurs du maximum de vraisemblance de ϕ et σ_ε^2 en fonction de x_1 et x_2 .

Exercice 5. Soit $X = (X_t, t \in \mathbb{Z})$ un processus $AR(p)$, vérifiant $\phi(B)(X)_t = \varepsilon_t$ avec ε bruit blanc de variance σ_ε^2 . On pose $\tilde{\varepsilon}_i = X_i - P_{V^2(1, X_1, \dots, X_{i-1})}^\perp(X_i)$. Soit $x = (x_1, \dots, x_N) \in \mathbb{R}^N$ une réalisation de (X_1, \dots, X_N) , $(\tilde{e}_1^{(\phi)}(x), \dots, \tilde{e}_N^{(\phi)}(x))$ la réalisation correspondante de $(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N)$. On note $\delta_i = \mathbb{E}[\tilde{\varepsilon}_i^2] / \sigma_\varepsilon^2$.

a) Montrer que pour tout $t > p$, $\delta_t = 1$ et $\tilde{e}_t^{(\phi)}(x) = \phi(B)(x)_t$.

b) En déduire une expression simple de la vraisemblance de x dans ce modèle $AR(p)$.

Exercice 6. Soit $X = (X_t, t \in \mathbb{Z})$ un processus $AR(2)$, vérifiant $X_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} = \varepsilon_t$ avec ε bruit blanc. On note $\tilde{\phi}_1^{(N)}$ et $\tilde{\phi}_2^{(N)}$ les estimateurs du maximum de vraisemblance de ϕ_1 et ϕ_2 connaissant (X_1, \dots, X_N) . Quand N tend vers l'infini, $\sqrt{N}(\tilde{\phi}_1^{(N)} - \phi_1, \tilde{\phi}_2^{(N)} - \phi_2)$ tend en loi vers un vecteur gaussien centré (U, V) de matrice de covariance

$$\begin{pmatrix} 1 - \phi_2^2 & \phi_1(1 - \phi_2) \\ \phi_1(1 - \phi_2) & 1 - \phi_2^2 \end{pmatrix}$$

En déduire des domaines de confiance au niveau α pour ϕ_1 , ϕ_2 , puis pour le couple (ϕ_1, ϕ_2) .

Comparaison des deux méthodes, et résultats asymptotiques.

Lemme 3.III.1 Sous les notations précédentes, X étant un processus $ARMA$ causal et inversible,

$$\lim_{t \rightarrow +\infty} \delta_t^{(\theta, \phi)} = 1$$

C'est une conséquence du théorème 3.II.6. On en déduit immédiatement le corollaire

Corollaire 3.III.2

$$\frac{1}{N} \sum_{t=1}^N \frac{\tilde{e}_t^{(\theta, \phi)}(x)^2}{\delta_t^{(\theta, \phi)}} \times \left(\prod_{t=1}^N \delta_t^{(\theta, \phi)} \right)^{\frac{1}{N}} \underset{N \rightarrow +\infty}{\sim} \frac{1}{N} \sum_{t=1}^N \frac{\tilde{e}_t^{(\theta, \phi)}(x)^2}{\delta_t^{(\theta, \phi)}}$$

Preuve : La limite de $\ln \left(\delta_t^{(\theta, \phi)} \right)$ est 0 quand t tend vers $+\infty$; la moyenne de Césaro $\frac{1}{N} \sum_{t=1}^N \ln \left(\delta_t^{(\theta, \phi)} \right)$ converge aussi vers 0 quand N tend vers $+\infty$. On en déduit

$$\lim_{N \rightarrow +\infty} \left(\prod_{t=1}^N \delta_t^{(\theta, \phi)} \right)^{\frac{1}{N}} = \exp 0 = 1.$$

□

Remarque : Les estimateurs des moindres carrés se distinguent nettement des estimateurs du maximum de vraisemblance lorsque $\mathbb{E}[(X_N - \bar{X}_N)^2]$ ne tend que lentement vers σ_ε^2 quand N tend vers $+\infty$, ce qui correspond au cas où le polynôme $\theta(z)$ possède une racine proche de 1.

Théorème 3.III.3 Soit $x = (x_1, \dots, x_N)$ une réalisation d'un processus ARMA causal et inversible X , caractérisé par les polynômes $\phi(z) = 1 + \phi_1 z + \dots + \phi_p z^p$ et $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$, et par la variance σ_ε^2 du bruit blanc d'innovation. Posons $\alpha = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)$. Soit $\hat{\alpha}$ l'estimateur du maximum de vraisemblance. Alors le vecteur $\sqrt{N}(\hat{\alpha} - \alpha)$ tend en loi vers un vecteur gaussien centré, de matrice de covariance de la forme $\begin{pmatrix} A & 0 \\ 0 & b \end{pmatrix}$ avec

$$\begin{aligned} A &= -\lim_{N \rightarrow +\infty} \frac{1}{N} \left(\frac{\partial^2 L_T}{\partial \alpha_i \partial \alpha_j}(\alpha) \right)_{1 \leq i, j \leq p+q} \\ b &= -\lim_{N \rightarrow +\infty} \frac{1}{N} \frac{\partial^2 L_T}{\partial (\sigma_\varepsilon^2)^2}(\alpha) \end{aligned}$$

Ce théorème sert en particulier à élaborer des tests (asymptotiques) de nullité des coefficients.

Remarque : On a le même résultat de convergence pour les estimateurs des moindres carrés, avec la même matrice de covariance asymptotique.

Méthode des moindres carrés conditionnels Les deux méthodes, des moindres carrés ordinaires et du maximum de vraisemblance, sont comme ont l'a vu très proches l'une de l'autre. Elles le sont en particulier en ceci qu'elles nécessitent à chaque évaluation de la fonction à minimiser le calcul de $N(N-1)/2$ coefficients obtenus par l'algorithme des innovations (ce nombre peut néanmoins être bien plus faible dans le cas d'un processus ARMA - voir exercices). Cette complexité peut devenir problématique, notamment pour de grands échantillons. La méthode des moindres carrés conditionnels s'avère alors plus adaptée. Elle est assez similaire aux précédentes, car elle consiste aussi à minimiser les erreurs de prédiction, mais en supposant qu'avant l'instant initial les valeurs inconnues prises par la série x sont nulles, c-à-d en conditionnant par $\{X_i = 0, \forall i \leq 0\}$. L'erreur de prédiction à l'instant t sachant tout le passé n'est autre que l'innovation :

$$\varepsilon_t = X_t - P_{V^2(X_s, s < t)}^\perp(X_t) = X_t + \sum_{i=1}^{+\infty} \eta_i^{(\theta, \phi)} X_{t-i}$$

Conditionnellement à l'événement $\{X_i = 0, \forall i \leq 0\}$, on a

$$\varepsilon_t = X_t + \sum_{i=1}^{t-1} \eta_i^{(\theta, \phi)} X_{t-i}.$$

On pose donc

$$e_t^{(\theta, \phi)}(x) = x_t + \sum_{i=1}^{t-1} \eta_i^{(\theta, \phi)} x_{t-i},$$

et on choisit comme valeurs de θ et ϕ celles qui minimisent la fonction

$$(\theta, \phi) \mapsto \sum_{t=1}^N e_t^{(\theta, \phi)}(x)^2 \quad (3.6)$$

L'évaluation de cette fonction ne nécessite plus que le calcul de N coefficients, qui s'opère par récurrence :

$$\eta_t^{(\theta, \phi)} = \phi_t - \sum_{u=1}^t \theta_u \eta_{t-u}^{(\theta, \phi)}$$

avec $\eta_0^{(\theta, \phi)} = 1$; cette formule est la conséquence de la relation existant entre les transformées en z

$$\theta(z)\eta^{(\theta, \phi)}(z) = \phi(z)$$

3.IV Test d'un modèle ARMA

Une fois le modèle ARMA (p, q) estimé, il convient ensuite de tester si la suite des résidus $(\hat{\varepsilon}_t^{(\hat{\theta}, \hat{\phi})} / \sqrt{\delta_t^{(\hat{\theta}, \hat{\phi})}}, t = 1, \dots, N)$ répond effectivement aux caractéristiques d'un bruit blanc (déjà vu au premier chapitre).

3.V Choix d'un modèle et critères d'information

Une fois passés les tests, il peut rester plusieurs modèles valides. On détermine le choix de l'un d'entre eux par d'autres critères, en particulier les critères d'information.

L'estimation par maximum de vraisemblance correspond à la minimisation d'une certaine distance entre la loi théorique et celle qu'on peut modéliser. Cette distance est appelée **information de Kullback** ou encore **entropie**. Il est donc assez naturel de choisir comme modèle celui qui minimise cette information.

Définition : Soient f et g deux densités de probabilité sur \mathbb{R}^m . On appelle **information de Kullback** de f par rapport à g la valeur suivante :

$$K(f, g) = \begin{cases} 2 \int_{\mathbb{R}^m} f(x_1, \dots, x_m) \ln \left(\frac{f(x_1, \dots, x_m)}{g(x_1, \dots, x_m)} \right) dx_1 \dots dx_m & \text{si } f \ll g \\ +\infty & \text{sinon.} \end{cases}$$

Propriété 3.V.1 1. $K(f, g)$ est positive.

2. $K(f, g) = 0$ si et seulement si $f = g$.

Soit $x = (x_1, \dots, x_N)$ une réalisation d'un processus X ARMA (p, q) gaussien, vérifiant $\phi(B)(X) = \theta(B)(\varepsilon)$, et on note $\hat{\phi}$, $\hat{\theta}$ et $\hat{\sigma}^2$ les estimateurs du maximum de vraisemblance (on suppose correctement estimés p et q). Soit Y le processus ARMA vérifiant $\hat{\phi}(B)(Y) = \hat{\theta}(B)(\varepsilon')$, avec ε' bruit blanc de variance $\hat{\sigma}^2$.

La densité de probabilité du vecteur (X_1, \dots, X_N) est donnée par $L(\theta, \phi, \sigma^2; \cdot)$ tandis que celle de (Y_1, \dots, Y_N) est donnée par $L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2; \cdot)$. La distance de Kullback entre le modèle exact et le modèle estimé vaut donc

$$\begin{aligned} & K(L(\theta, \phi, \sigma^2; \cdot), L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2; \cdot)) \\ &= 2\mathbb{E}[\ln(L(\theta, \phi, \sigma^2; X))] - 2\mathbb{E}[\ln(L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2; X))] \end{aligned}$$

Minimiser cette distance revient à minimiser

$$\begin{aligned} & -2\mathbb{E}\left[\ln(L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2; X))\right] \\ & = -2\ln(L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2; x)) + \mathbb{E}\left[\frac{1}{\hat{\sigma}^2} \sum_{t=1}^N \frac{\hat{e}_t^{(\hat{\theta}, \hat{\phi})}(X)^2}{\delta_t^{(\hat{\theta}, \hat{\phi})}}\right] - N \end{aligned}$$

Or on a l'estimation

$$\mathbb{E}\left[\frac{1}{\hat{\sigma}^2} \sum_{t=1}^N \frac{\hat{e}_t^{(\hat{\theta}, \hat{\phi})}(X)^2}{\delta_t^{(\hat{\theta}, \hat{\phi})}}\right] \sim_{N \rightarrow +\infty} 2(p + q + 1)$$

Akaïke a donc proposé de choisir comme modèle celui qui minimise

$$AIC(p, q, \theta, \phi, \sigma^2) = -2\ln L(\theta, \phi, \sigma^2; x) + 2(p + q + 1)$$

Ce critère a tendance à conduire à une surestimation des valeurs de p et q . Akaïke en a donc proposé une modification, le critère bayésien de Schwartz :

$$SBC \equiv BIC(p, q, \theta, \phi, \sigma^2) = -2\ln L(\theta, \phi, \sigma^2; x) + 2(p + q + 1) \ln N$$

Les estimateurs de p et q obtenus par minimisation de ce critère ont le mérite d'être convergents.

Remarques :

- $-2\ln L(\hat{\theta}, \hat{\phi}, \hat{\sigma}^2; x)$ est approximativement égal à $N \ln(\hat{\sigma}^2) + N \ln(2\pi) + N$. Ce qui permet de donner des versions plus simples des critères AIC et BIC.
- Dans un modèle, si l'estimation de certains des coefficients n'est pas significativement différente de 0, on les fixe arbitrairement nuls. Il faut dans ce cas remplacer dans l'évaluation des critères AIC et BIC $p + q$ par le nombre de paramètres effectivement estimés.
- L'un des buts de la modélisation est de permettre la prévision. L'erreur qui est alors commise est somme de deux types d'erreur :
 1. L'erreur de prévision pure, erreur que l'on commettrait si les paramètres du modèle étaient connus exactement (ainsi que toutes les valeurs passées du processus), erreur qui est proportionnelle à la variance du bruit blanc du modèle.
 2. L'erreur due à l'estimation des paramètres, qui est d'autant plus faible que les paramètres sont peu nombreux (et le processus connu sur une longue durée).

Cette distinction a déjà été observée au premier chapitre. Pour avoir l'erreur de prévision totale la plus faible, on peut donc être tenté de diminuer le nombre de paramètres du modèle, ce qui diminuerait l'erreur d'estimation, mais augmenterait la variance du bruit blanc; inversement, chercher l'erreur de prévision pure la plus petite conduit à privilégier des modèles à beaucoup de paramètres, dont l'erreur d'estimation sera forte.

Les critères d'information qui viennent d'être vus proposent en fait une évaluation simultanée et pondérée de la variance du bruit blanc et du nombre de paramètres. Ils invitent à choisir un modèle réalisant un compromis entre erreur de prévision pure et erreur d'estimation. Cela permet d'espérer une erreur de prévision totale la plus mince possible.

Chapitre 4

Modélisation et Prévision

4.I Modèles linéaires avec bruit ARMA

Au chapitre 1, l'étude de la série temporelle $x = (x_t, t \in T = \{1869, \dots, 1972\})$ associée au niveau du lac Huron sur une centaine d'années avait conduit à proposer le modèle linéaire suivant :

$$\forall t \in T, X_t = a + bt + ct^2 + \varepsilon_t$$

avec ε bruit blanc. Or, après estimation, les tests sur les résidus $(x_t - \hat{a} - \hat{b}t, t \in T)$ ont conduit à rejeter cette dernière hypothèse. On est donc amené à envisager un modèle plus général de la forme

$$\forall t \in T, X_t = a + bt + ct^2 + Y_t$$

avec Y processus ARMA. Reste alors à estimer à la fois les paramètres du modèle linéaire et ceux de Y .

Définition : Soit $x = (x_t, t \in T)$ une série temporelle. On appelle **modèle linéaire avec erreur ARMA** toute modélisation de x comme réalisation d'un processus X vérifiant

$$\forall t \in T, X_t = f(t, \beta) + Y_t$$

avec f fonction linéaire du paramètre $\beta \in \mathbb{R}^d$ et Y processus ARMA. Si f est une fonction constante de t , alors on dit simplement que X est un **processus ARMA décentré**.

L'estimation simultanée des différents paramètres de X est complexe. Rappelons les résultats de l'exercice 1.IV.C : si $X = (X_t, t \in \{1, \dots, N\})$ vérifie $X = A\beta + Y$ avec A matrice de dimensions $N \times m$ et de rang m , $Y = (Y_t, t \in \{1, \dots, N\})$ processus ARMA(p, q) de matrice de covariance $\sigma^2 R_{(\theta, \phi)}^{(N)}$, avec θ, ϕ connus, et $(\beta, \sigma^2) \in \mathbb{R}^m \times \mathbb{R}_+^*$ paramètres, alors l'estimateur des moindres carrés de β - i.e. celui qui minimise σ^2 -, qui est aussi l'estimateur du maximum de vraisemblance à supposer que Y soit gaussien, est

$$\hat{\beta} = \left({}^t A \left(R_{(\theta, \phi)}^{(N)} \right)^{-1} A \right)^{-1} {}^t A \left(R_{(\theta, \phi)}^{(N)} \right)^{-1} X. \quad (4.1)$$

Comme p, q, θ, ϕ ne sont pas connus mais doivent être estimés, on peut procéder de manière récurrente, par approximations successives :

- On suppose d'abord que $p = q = 0$, i.e. que Y est un bruit blanc. On en déduit l'estimation $\hat{\beta}_{(0)} = ({}^t A A)^{-1} {}^t A X$ de β à partir de la formule ordinaire.
- On calcule la série temporelle $y_{(0)} = x - A\hat{\beta}_{(0)}$, qu'on modélise par un processus ARMA($p_{(1)}, q_{(1)}$) de paramètres $\hat{\theta}_{(1)}$ et $\hat{\phi}_{(1)}$.
- On en déduit alors une nouvelle estimation

$$\hat{\beta}_{(1)} = \left({}^t A \left(R_{(\hat{\theta}_{(1)}, \hat{\phi}_{(1)})}^{(N)} \right)^{-1} A \right)^{-1} {}^t A \left(R_{(\hat{\theta}_{(1)}, \hat{\phi}_{(1)})}^{(N)} \right)^{-1} x$$

- On recalcule la série temporelle des erreurs $y_{(1)} = x - A.\hat{\beta}_{(1)}$ qu'on modélise par un processus ARMA($p_{(2)}, q_{(2)}$) de paramètres $\hat{\theta}_{(2)}$ et $\hat{\phi}_{(2)}$.
- On réitère le procédé jusqu'à ce que les valeurs prises par les paramètres ne se modifient sensiblement plus.

Exercice 1. Soit $X = (X_t, t \in \mathbb{Z})$ un processus de type MA(1) décentré :

$$\forall t \in \mathbb{Z} \quad X_t = \mu_X + \varepsilon_t + \theta\varepsilon_{t-1}$$

avec ε bruit blanc gaussien de variance σ^2 et $|\theta| < 1$. Déterminer l'estimateur du maximum de vraisemblance de μ_X et de θ , sachant (X_1, \dots, X_N) .

4.II Modèles "intégrés"

Dériver est une autre façon, plus souple, d'obtenir une série stationnaire. Cela consiste à faire passer la série initiale dans des filtres de type $I - B^n$.

4.II.A Modèles ARIMA

Définition : On dit que $X = (X_t, t \in \mathbb{Z})$ est un processus **ARIMA**(p, d, q) si $(I - B)^d(X)$ est un processus ARMA(p, q).

Suppression d'une tendance polynômiale

Si $x = (x_t, t \in T)$ est une série temporelle qui peut se mettre sous la forme

$$\forall t \in T, \quad x_t = at + b + y_t$$

avec $r = (r_t)_t$ une réalisation d'un processus stationnaire centré, alors la série $x' = (1 - B)(x)$ s'écrit

$$x'_t = a + r_t - r_{t-1} = a + y'_t$$

avec y' une réalisation d'un processus stationnaire centré. Partant de ce constat simple, on poursuit l'étude en modélisant y' par un processus Y' ARMA(p, q) décentré, vérifiant $\phi(B)(Y') = \mu + \theta(B)(\varepsilon)$, et on en déduit le modèle ARIMA($p, 1, q$) décentré $(I - B)\phi(B)(X) = \mu + \theta(B)(\varepsilon)$.

Si l'on dérive deux fois, autrement dit si l'on fait passer la série par le filtre $(I - B)^2$, on pourra éliminer une tendance binômiale. Plus généralement, pour supprimer une tendance polynômiale de degré d , il suffit de dériver d fois.

Suppression d'un comportement a-périodique non stationnaire

Une dérivation par $(1 - B)$ est aussi utile pour étudier des processus non stationnaires qui n'admettent pas pour autant de tendance simple.

Exemple : cas d'une marche aléatoire X

$$X_t = \varepsilon_t + X_{t-1}$$

avec ε bruit blanc. Alors le processus dérivé $Y = (1 - B)(X)$ est égal en fait à ε , et il est facile de l'identifier. D'autre part, on vérifie facilement que $\mathbb{E}[X_t] = \text{cte}$, et qu'il ne présente donc pas de tendance particulière.

Il peut être utile aussi de dériver pour obtenir une meilleure précision des paramètres du modèle. Supposons que l'on ait une série temporelle x que l'on modélise par un processus ARMA vérifiant

$$\phi(B)(X) = \theta(B)(\varepsilon)$$

avec $\phi(z)$ un polynôme possédant une racine proche de l'unité. Dans ce cas, l'intervalle de confiance pour l'estimation des coefficients de θ et ϕ sera grand. La raison en est que le processus X est très proche d'un

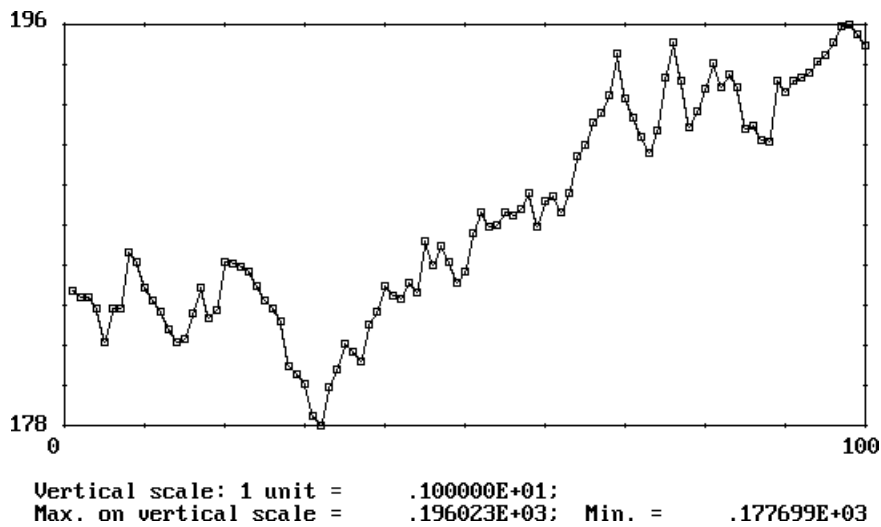


FIG. 4.1 – Graphe d’une réalisation de la marche aléatoire X pour $t \in \{1, \dots, 100\}$

processus non stationnaire, et sur l’intervalle de temps fini sur lequel on le considère, il peut même en être quasi indiscernable. Or les résultats de convergence des estimations ne sont valables que pour des processus stationnaires.

Dans cette situation, il peut être préférable de dériver x par $(I - B)$, et de modéliser la série résultante par un processus ARMA dont le polynôme θ associé ne possède pas de racine proche de l’unité, et pour lequel l’approximation des autres coefficients sera plus fine.

Les situations où il est utile de dériver sont donc plus nombreuses que celles où il s’agit simplement d’enlever une tendance polynômiale. Mais dans le cas où une tendance est présente qu’il s’agit de soustraire, notons que si la dérivation peut permettre de modéliser correctement la série temporelle, en revanche elle ne donne pas d’information sur la tendance elle-même.

Choix d’un modèle ARIMA

On différencie par le filtre $(I - B)$ si l’on repère une tendance générale non constante (en particulier monotone), ou si les premières valeurs de la fonction d’autocorrélation estimée sont sensiblement non nulles, et à décroissance lente.

Cette deuxième condition s’appuie sur le résultat suivant :

Proposition 4.II.1 *Soit X un processus ARMA tel que*

$$\phi(B)(X) = \theta(B)(\varepsilon)$$

Si α est la plus grande des racines (en valeur absolue) de $\theta(z)$, avec une multiplicité n , alors il existe une constante C telle que

$$\forall t \in \mathbb{N} \quad |\rho_X(t)| \leq Ct^{n-1}\alpha^t$$

On peut dériver plusieurs fois. On s’arrête dès que l’une ou l’autre des fonctions d’autocorrélation estimées décroît suffisamment vite vers 0, et qu’il n’existe pas de racine trop proche de 1 dans le polynôme θ du modèle

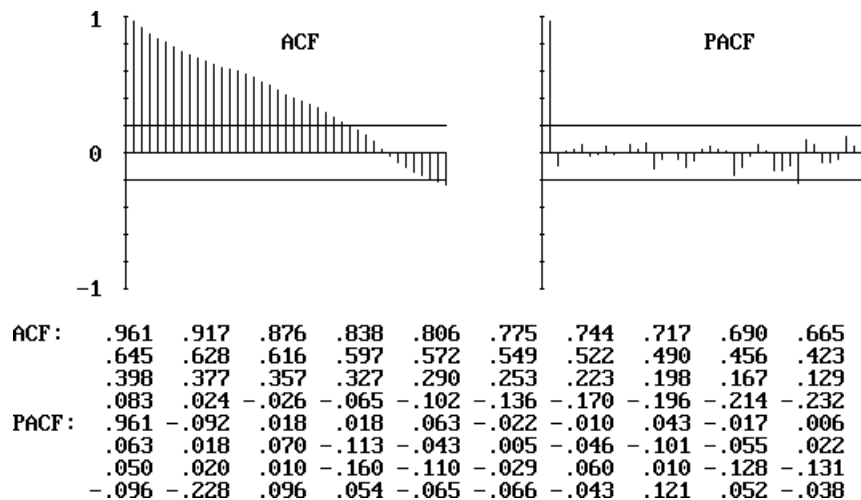


FIG. 4.2 – Autocorrélogramme associé

ARMA qu'on peut alors élaborer.

Exercice 2. Soit X un processus $ARIMA(0, 1, 1)$, tel que $X_0 = 0$. Calculer $\rho_X(s, s + h)$, puis montrer $\lim_{s \rightarrow +\infty} \rho_X(s, s + h) = 1$.

4.II.B Modèles SARIMA

Supposons que x soit une série temporelle qui se mette sous la forme $x_t = f_\sigma(t) + r_t$ avec f_σ périodique de période n . Alors $y = (I - B^n)(x)$ vérifie

$$\begin{aligned}
 y_t &= x_t - x_{t-n} \\
 &= f_\sigma(t) - f_\sigma(t-n) + r_t - r_{t-n} \\
 &= r_t - r_{t-n} = r_t''
 \end{aligned}$$

Ne reste plus alors qu'à établir pour y un modèle ARMA.

Définition : Un processus X tel que $(I - B^n)(X)$ soit un ARIMA est appelé un processus **SARIMA**.

En pratique, on filtre une série x par $(I - B^n)$ si l'on connaît a priori l'existence d'une périodicité de période n , ou si les valeurs de $\hat{\gamma}_x$ aux instants $n, 2n, 3n, \dots$ sont sensiblement non nuls, ou faiblement décroissantes.

Naturellement, il est possible de dériver plusieurs fois. La forme générale d'un processus "intégré" est donc la suivante :

$$\theta(B) \circ (I - B^{n_1})^{d_1} \circ \dots \circ (I - B^{n_k})^{d_k}(X) = \phi(B)(\varepsilon)$$

Remarques :

- Si l'on souhaite avoir le détail des variations saisonnières, ou isoler une tendance générale, une décomposition classique, utilisant les modèles linéaires ou les moyennes mobiles, peut être plus adaptée.
- Les filtres du type $(I - B^n)$ ne sont pas inversibles. Si $y = (I - B^n)(x)$, alors il n'est pas possible de récrire x comme $\psi(y)$ avec ψ filtre linéaire. Mais on dispose d'une formule de récurrence :

$$y_t = (I - B^n)(x)_t \iff x_t = y_t + x_{t-n}$$

Il suffit donc de disposer de n valeurs consécutives pour reconstruire la série initiale.

- Il ne faut pas abuser des dérivations, qui conduisent ensuite à des erreurs de prévision importantes. En particulier, on ne dérive pas un processus pour simplement le recentrer (on lui ôte plutôt sa moyenne empirique). De même, il ne faut pas oublier que dans $I - B^n$, il y a $I - B$ (composé avec $I + B + \dots + B^{n-1}$). Ainsi, si un processus paraît avoir une tendance affine et une saisonnalité de période n , une seule dérivation par $I - B^n$ suffira à supprimer la saisonnalité, et à se ramener à une tendance constante.
- Les processus ARMA, ARIMA et SARIMA vérifient tous une équation du type $\phi^*(B)(X) = \theta(B)(\varepsilon)$, avec $\phi^*(B)$ et $\theta(B)$ filtres linéaires ; mais le filtre $\phi^*(B)$ n'est inversible que dans le seul cas des processus ARMA. La distinction entre ces modèles est donc indispensable. Notons aussi qu'un processus ARIMA n'est pas stationnaire, et que la fonction d'autocorrélation d'un processus possédant une composante saisonnière ne tend vers 0 en l'infini : toutes choses erronées pour un processus ARMA.

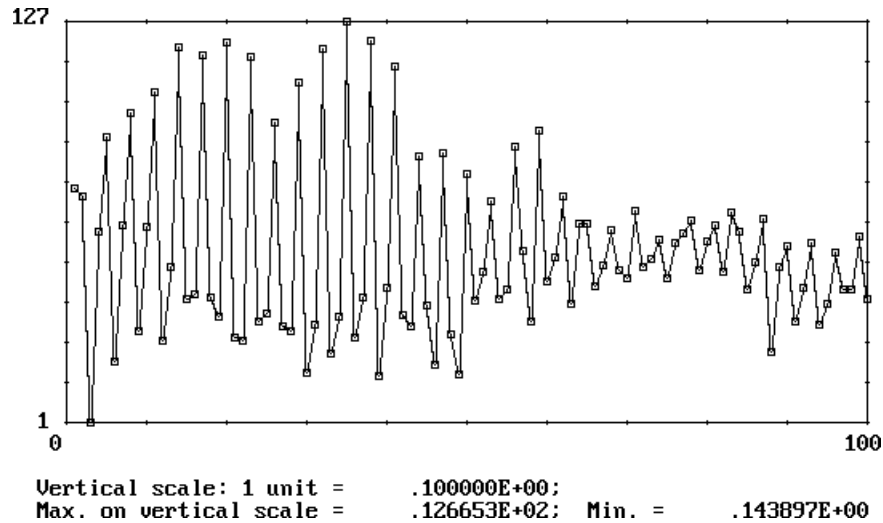


FIG. 4.3 – Graphe d'une réalisation $x = (x_t, t \in \{1, \dots, 100\})$ de $X_t = \varepsilon_t - 0,98X_{t-1}$

Exemple :

Exercice 3. Soit $X = (X_t, t \in \mathbb{N})$ un processus du second ordre vérifiant le modèle SARIMA(1,1,1)(0,1,0)₄ suivant :

$$\forall t \geq 6, (I - B)(I - B^4)(I + \phi B)(X)_t = (I + \theta B)(\varepsilon)_t$$

avec $|\phi|, |\theta| < 1$, et ε bruit blanc.

1. Déterminer une équation de récurrence vérifiée par la fonction moyenne du processus X .
2. En déduire le type de séries que l'on peut modéliser avec ce processus.
3. Mêmes questions en supposant cette fois que X est de modèle SARIMA(1,0,1)(0,2,0)₄ :

$$(I - B^4)^2(I + \phi B)(X)_t = (I + \theta B)(\varepsilon)_t$$

4.III Prévision

4.III.A Prédiction linéaire : cas général

Soit $x = (x_t, t \in T)$ une série temporelle qu'on modélise par un processus $X = (X_t, t \in S)$ avec $T \subset S$, et de loi paramétrée par $\beta \in \mathbb{R}^k$. Soit $s \in S \setminus T$ un indice de temps pour lequel on ne possède pas de relevé.

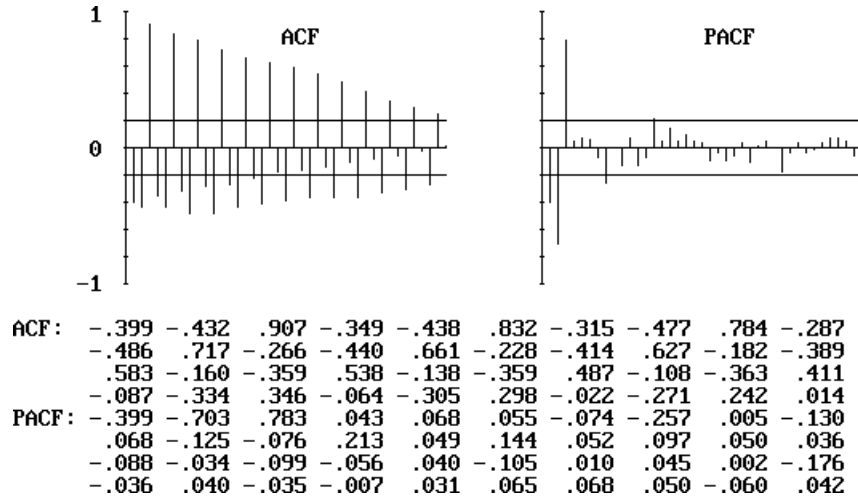


FIG. 4.4 – Autocorrélogramme de x

Prédire X_s , c'est déterminer une variable aléatoire fonction de $(X_t, t \in T)$ la plus proche possible de X_s . Si l'on mesure la distance entre deux variables par la norme hilbertienne, alors le meilleur prédicteur au sens des moindres carrés est

$$P_{L^2(X_t, t \in T)}^\perp(X_s) = \mathbb{E}[X_s | X_t, t \in T]$$

Cette fonction n'est pas en général calculable (sauf dans le cas gaussien), et l'on préfère se limiter aux fonctions affines de $(X_t, t \in T)$. On parle alors de meilleur prédicteur affine au sens des moindres carrés. Il s'agit de

$$\tilde{X}_T(s) = P_{V^2(1, X_t, t \in T)}^\perp(X_s) = a_0(s, \beta) + \sum_{t \in T} a_t(s, \beta) X_t$$

Rappelons que dans le cas d'un processus gaussien on a l'égalité entre $P_{V^2(1, X_t, t \in T)}^\perp(X_s)$ et $P_{L^2(X_t, t \in T)}^\perp(X_s)$.

Les coefficients a_i sont caractérisés par le système d'équations

$$\begin{aligned} a_0(s, \beta) + \sum_{t \in T} a_t(s, \beta) \mathbb{E}[X_t] &= \mathbb{E}[X_s] \\ \sum_{t \in T} a_t(s, \beta) \text{cov}(X_t, X_u) &= \text{cov}(X_s, X_u) \quad \forall u \in T \end{aligned}$$

Le problème évident est que la variable aléatoire $P_{V^2(1, X_t, t \in T)}^\perp(X_s)$ dépend du paramètre β en général inconnu, qu'on est donc amené à remplacer par son estimateur $\hat{\beta} = \hat{\beta}(X_t, t \in T)$. On obtient ainsi le prédicteur

$$\hat{X}_T(s) = a_0(s, \hat{\beta}) + \sum_{t \in T} a_t(s, \hat{\beta}) X_t$$

L'erreur quadratique commise vaut alors

$$\|X_T(s) - \hat{X}_T(s)\|_2 \leq \|X_s - \tilde{X}_T(s)\|_2 + \|\tilde{X}_T(s) - \hat{X}_T(s)\|_2$$

Le premier terme $\|X_s - \tilde{X}_T(s)\|_2$ correspond à l'erreur de prédiction pure, le second $\|\tilde{X}_T(s) - \hat{X}_T(s)\|_2$ à l'erreur d'estimation. Il est difficile de calculer exactement l'erreur totale, ni même l'erreur d'estimation, sauf dans le cas des modèles linéaires. L'erreur de prédiction pure est en revanche plus simple à déterminer. C'est aussi en général le terme dominant : en effet, les différents résultats de convergence montrent le plus souvent que $\|\beta - \hat{\beta}\|_2$ est de l'ordre de $1/\sqrt{\#T}$, dont il suit, par continuité, qu'il en est de même pour l'erreur d'estimation. Comme nous le verrons, il n'en est pas de même pour l'erreur de prédiction pure, que dans la pratique on est amené à confondre avec l'erreur totale commise.

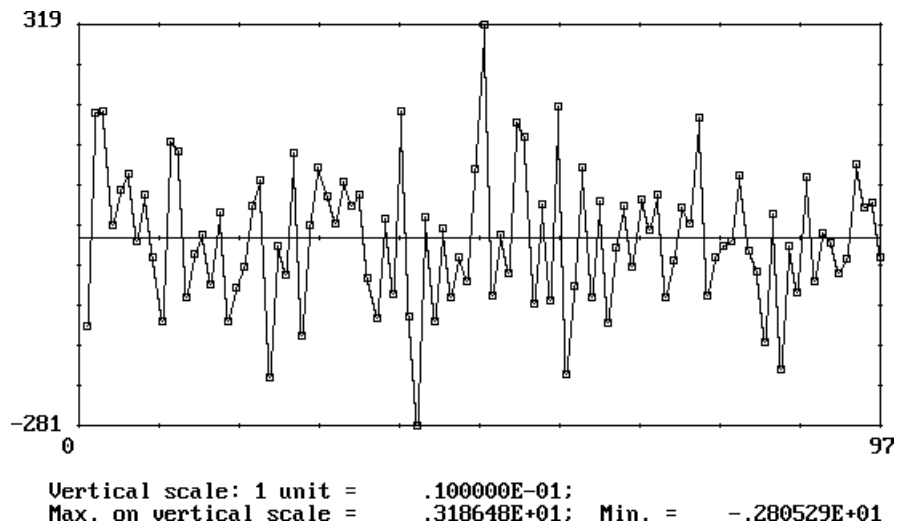


FIG. 4.5 – Graphe de $y = (1 - B)(x)$

4.III.B Prédiction pour un processus ARMA

Soit $x = (x_1, \dots, x_N)$ une série temporelle qu'on pense être une réalisation d'un processus X ARMA(p, q) vérifiant la représentation causale et inversible $\phi(B)(X) = \theta(B)(\varepsilon)$ avec ε bruit blanc de variance σ^2 .

Remarque. On supposera dans toute la suite les processus ARMA centrés. En conséquence, on pourra remplacer les espaces $V^2(1, X_1, \dots)$ par $V^2(X_1, \dots)$.

On pose comme à l'ordinaire

$$\begin{aligned} \tilde{\varepsilon}_t &= X_t - P_{V^2(X_1, \dots, X_t)}^\perp(X_t) \\ &= X_t + \sum_{u=1}^{t-1} \eta_{t,u} X_{t-u} \\ X_t &= \tilde{\varepsilon}_t + \sum_{u=1}^{t-1} \psi_{t,u} \tilde{\varepsilon}_{t-u} \\ \frac{\theta(z)}{\phi(z)} &= 1 + \sum_{u \geq 1} \psi_u z^u \end{aligned}$$

Et on note \hat{p} , \hat{q} , $\hat{\phi}$, $\hat{\theta}$, $\hat{\sigma}^2$, $\hat{\psi}_u$, $\hat{\psi}_{t,u}$ et $\hat{\eta}_{t,u}$ les estimateurs des différents paramètres du processus. On souhaite construire un prédicteur de X_{N+h} , avec $h \geq 1$, qu'on notera $\tilde{X}_N(h)$ (plutôt que $\tilde{X}_{\{1, \dots, N\}}(N+h)$) comme

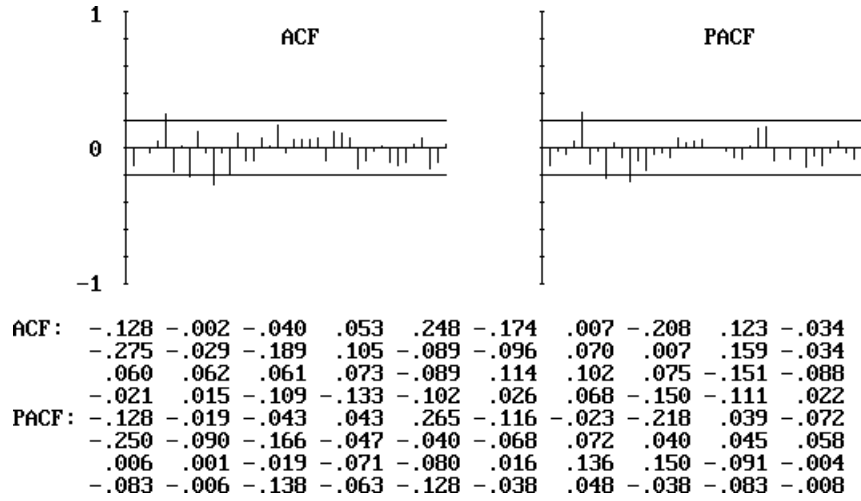


FIG. 4.6 – Autocorrélogramme de y

précédemment). Comme le processus est centré, on a

$$\begin{aligned}
\check{X}_N(h) &= P_{V^2(1, X_1, \dots, X_N)}^\perp(X_{N+h}) \\
&= P_{V^2(X_1, \dots, X_N)}^\perp(X_{N+h}) \\
&= P_{V^2(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N)}^\perp(X_{N+h}) \\
&= P_{V^2(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N)}^\perp\left(\tilde{\varepsilon}_{N+h} + \sum_{u=1}^{N+h-1} \psi_{N+h,u} \tilde{\varepsilon}_{N+h-u}\right) \\
&= \sum_{u=h}^{N+h-1} \psi_{N+h,u} \tilde{\varepsilon}_{N+h-u} \\
&= \sum_{u=h}^{N+h-1} \psi_{N+h,u} \left(X_{N+h-u} + \sum_{v=1}^{N+h-u-1} \eta_{N+h-u,v} X_{N+h-u-v}\right) \\
&= \sum_{t=0}^{N-1} X_{N-t} \left(\psi_{N+h,t} + \sum_{s=0}^{t-1} \psi_{N+h,s+h} \eta_{N-s,t-s}\right)
\end{aligned}$$

On en déduit le prédicteur linéaire

$$\hat{X}_N(h) = \sum_{t=0}^{N-1} X_{N-t} \left(\hat{\psi}_{N+h,t} + \sum_{s=0}^{t-1} \hat{\psi}_{N+h,s+h} \hat{\eta}_{N-s,t-s}\right)$$

L'erreur de prédiction vaut

$$\begin{aligned}
\|X_{N+h} - \check{X}_N(h)\|_2^2 &= \|\tilde{\varepsilon}_{N+h} + \sum_{u=1}^{h-1} \psi_{N+h,u} \tilde{\varepsilon}_{N+h-u}\|_2^2 \\
&= \sigma^2 \left(\delta_{N+h} + \sum_{u=1}^{h-1} \psi_{N+h,u}^2 \delta_{N+h-u}\right) \\
&\approx \sigma^2 \left(1 + \sum_{u=1}^{h-1} \psi_u^2\right)
\end{aligned}$$

A l'aide de l'inégalité de BienAymé-Tchebicheff, ou en supposant que le processus soit gaussien, on peut déduire de ce calcul d'erreur un intervalle de prévision (approximatif) pour X_{N+h} .

Remarques :

- L'erreur de prédiction est toujours supérieure à σ^2 (contrairement aux erreurs d'estimation que l'on peut faire "tendre" vers 0 si le nombre d'observations tend vers l'infini).
- $\lim_{h \rightarrow +\infty} \|X_{N+h} - \check{X}_N(h)\|_2 = \gamma_X(0)$.

Exercice 4. Soit X un MA(q). Montrer que $\check{X}_N(h) = P_{V^2(X_1, \dots, X_N)}^\perp(X_{N+h}) = 0$ si $h > q$.

Exercice 5. Soit X un processus autorégressif d'ordre p , vérifiant

$$\forall t \in \mathbb{Z}, \quad X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \varepsilon_t$$

avec ε bruit blanc d'innovation. On suppose connus (X_1, \dots, X_N) . Montrer que $\check{X}_N(1) = \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}$.

Exercice 6. Soit X un AR(1), vérifiant $X_t + \phi X_{t-1} = \varepsilon_t$ avec ε bruit blanc d'innovation associé à X . Montrer que $\check{X}_N(h) = (-\phi)^h X_N$ quel que soit $h \geq 0$. Calculer l'erreur de prédiction.

Exercice 7. Soit X un processus faiblement stationnaire de type MA(1), vérifiant :

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1} \quad \forall t \in \mathbb{Z}$$

avec ε bruit blanc faible de variance 1. On définit, à l'aide de l'algorithme des innovations, les variables aléatoires $\tilde{\varepsilon}_t$ et les coefficients $(\theta_{n,t})_{0 \leq n < t}$ tels que

$$\begin{aligned} \tilde{\varepsilon}_t &= X_t - P_{V^2(X_1, \dots, X_{t-1})}^\perp(X_t) \\ X_t &= \sum_{n=0}^{t-1} \theta_{n,t} \tilde{\varepsilon}_{t-n} \end{aligned}$$

1. Que vaut $\theta_{0,t}$? Que valent $\theta_{n,t}$ pour $n \geq 2$? Quelle est la limite de $\theta_{1,t}$, quand t tend vers l'infini ?
2. Déterminer en fonction de $\tilde{\varepsilon}$ et de $\theta_{1,t}$ les prédicteurs linéaires $\check{X}_N(h) = P_{V^2(X_1, \dots, X_N)}^\perp(X_{N+h})$, ainsi que la variance des erreurs de prédiction. En donner une approximation pour N grand.

4.III.C Prédiction pour un processus (S)AR(I)MA

Les processus SARIMA étant définis linéairement à partir des processus ARMA, il est facile de déduire du paragraphe précédent des formules de calcul de prédicteurs et d'erreurs de prédiction. Par exemple, si X est un ARIMA($p, 1, q$), autrement dit si $(I - B)(X) = Y$ avec X ARMA(p, q), alors quel que soit $t > 0$,

$$X_t = X_0 + Y_1 + \dots + Y_t.$$

En supposant X_0 orthogonal à $V^2(Y_t, t \geq 1)$ (plus généralement, on est toujours obligé de rajouter une hypothèse du type $V^2(X_s, s \geq 0) \perp V^2(Y_t, t \geq 1)$ lors de l'étude d'un processus (S)ARIMA), on en déduit que pour tout $h > 0$

$$\begin{aligned} P_{V^2(X_0, \dots, X_N)}^\perp(X_{N+h}) &= P_{V^2(X_0, Y_1, \dots, Y_N)}^\perp(X_0 + Y_1 + \dots + Y_{N+h}) \\ &= X_0 + Y_1 + \dots + Y_N + \check{Y}_N(1) + \dots + \check{Y}_N(h) \end{aligned}$$

On est ainsi ramené aux problèmes de prédiction pour un ARMA.

Mais il est intéressant de noter qu'on dispose en fait, tant pour les processus ARMA qu'ARIMA et SARIMA, de formules de récurrence qui simplifient nettement ces calculs.

Supposons que X soit un (S)AR(I)MA, vérifiant

$$(I - B)^d \circ (1 - B^{T_1})^{D_1} \circ \dots \circ (I - B^{T_n})^{D_n} \circ \phi(B)(X) = \theta(B)(\varepsilon)$$

avec ε bruit blanc de variance σ_ε^2 . On notera plus simplement $\phi^*(z) = (1 - z)^d (1 - z^{T_1})^{D_1} \dots (1 - z^{T_n})^{D_n} \phi(z)$; d'où l'écriture $\phi^*(B)(X) = \theta(B)(\varepsilon)$. Soit p^* le degré du polynôme $\phi^*(z)$, q celui de $\theta(z)$. On suppose $N > \max(p^*, q)$, et on souhaite calculer le prédicteur $\check{X}_N(h)$.

On pose

$$Y_t = \begin{cases} X_t & \text{si } t \leq \max(p^*, q) \\ \phi^*(B)(X)_t & \text{si } t > \max(p^*, q) \end{cases}$$

1. Quel que soit $t > 0$, $X_t - P_{V^2(X_1, \dots, X_{t-1})}^\perp(X_t) = Y_t - P_{V^2(Y_1, \dots, Y_{t-1})}^\perp(Y_t) = \tilde{\varepsilon}_t$.
2. Quel que soit $t > \max(p^*, q)$, $Y_t = \sum_{j=0}^q \theta_{t,j} \tilde{\varepsilon}_{t-j}$.

On en déduit que pour tout $N + h > \max(p^*, q)$,

$$\begin{aligned} X_{N+h} &= -\phi_1^* X_{N+h-1} - \dots - \phi_{p^*}^* X_{N+h-p^*} + \sum_{j=0}^q \theta_{N+h,j} \tilde{\varepsilon}_{N+h-j} \\ \check{X}_N(h) &= -\phi_1^* \check{X}_N(h-1) - \dots - \phi_{p^*}^* \check{X}_N(h-p^*) + \sum_{j=h}^q \theta_{N+h,j} \tilde{\varepsilon}_{N+h-j} \text{ si } h \leq q \\ &= -\phi_1^* \check{X}_N(h-1) - \dots - \phi_{p^*}^* \check{X}_N(h-p^*) \text{ si } h > q \end{aligned}$$

D'où un calcul par récurrence du meilleur prédicteur linéaire.

Fonction de prévision La suite $(\check{X}_N(h))_{h>q}$ vérifie la relation de récurrence

$$\check{X}_N(h) + \phi_1^* \check{X}_N(h-1) + \dots + \phi_{p^*}^* \check{X}_N(h-p^*) = 0$$

qui admet comme polynôme caractéristique

$$z^{p^*} + \phi_1^* z^{p^*-1} + \dots + \phi_{p^*}^* (= z^{p^*} \phi^*(1/z))$$

Les racines de ce polynôme sont $1/\alpha_1, \dots, 1/\alpha_n$ avec (α_i) les racines de $\phi^*(z)$. Soient d_1, \dots, d_n les multiplicités respectives de ces racines. Alors il existe P_1, \dots, P_n , polynômes de degrés respectifs $d_1 - 1, \dots, d_n - 1$ tels que, pour tout $h > q$,

$$\check{X}_N(h) = \frac{P_1(h)}{\alpha_1^h} + \dots + \frac{P_n(h)}{\alpha_n^h}$$

Les valeurs initiales $\check{X}_N(q), \dots, \check{X}_N(q-p+1)$ qui permettent de déterminer P_1, \dots, P_n , sont appelées **valeurs pivotales**.

Erreur de prévision Notons $\check{d}_{N+h} = X_{N+h} - \check{X}_N(h)$. On a la formule de récurrence

$$\check{d}_{N+h} + \phi_1^* \check{d}_{N+h-1} + \dots + \phi_{p^*}^* \check{d}_{N+h-p^*} = \sum_{j=0}^{q \wedge h-1} \theta_{N+h,j} \tilde{\varepsilon}_{N+h-j}$$

Exemples :

$$\begin{aligned} \check{d}_{N+1} &= \tilde{\varepsilon}_{N+1} \\ \check{d}_{N+2} &= \tilde{\varepsilon}_{N+2} + (\phi_1^* + \theta_{N+2,1}) \tilde{\varepsilon}_{N+1} \end{aligned}$$

Par récurrence, on montre ainsi qu'il existe des coefficients $\zeta_{h,i}$ tels que

$$\check{d}_{N+h} = \sum_{i=1}^h \zeta_{h,i} \tilde{\varepsilon}_{N+i}$$

L'erreur de prévision est donc égale à

$$\|X_{N+h} - \check{X}_N(h)\|_2^2 = \mathbb{E}[d_{N+h}^2] = \sigma_\varepsilon^2 \sum_{i=1}^h \zeta_{h,i}^2 \delta_{N+i}$$

Le calcul des coefficients $\zeta_{h,i}$ s'effectue grâce à la formule de récurrence

$$\check{\zeta}_{h,i} = -\phi_1^* \check{\zeta}_{h-1,i} - \dots - \phi_{p^*}^* \check{\zeta}_{h-p^*,i} + \theta_{N+h,h-i}$$

avec $\zeta_{k,i} = 0$ si $k < i$.

Exercice 8. Soit X le processus ARIMA(2,1,0) vérifiant

$$\forall t \geq 0, (1-B) \circ (I-0,4B-0,05B^2)(X)_t = \varepsilon_t$$

avec ε bruit blanc. On suppose connus (X_1, \dots, X_N) . Déterminer la fonction de prévision \check{X}_{N+} , et calculer l'erreur de prédiction.

Exercice 9. Soit $X = (X_t, t \in \mathbb{Z})$ un processus ARIMA(1,1,1) vérifiant

$$(I-B)(I-0,5B)(X)_t = (I-0,8B)(\varepsilon)_t$$

avec ε bruit blanc de variance $\sigma_\varepsilon^2 = 0,25$. On suppose que $X_N = 12$ et que $\check{X}_{N+1} = P_{V^2(X_1, \dots, X_N)}^\perp(X_{N+1}) = 10$. On suppose aussi N grand. Donner un intervalle de prévision à 95% de X_{N+2} , en supposant X_1, \dots, X_N connus.

Exercice 10. Soit X un processus ARIMA(0,2,1). Déterminer la fonction de prévision \check{X}_{N+} , en précisant les valeurs pivôtales, où l'on suppose connus X_1, \dots, X_N . Donner l'expression formelle de l'erreur de prévision, puis son approximation explicite en supposant N grand. Mêmes questions avec un processus ARIMA(1,1,0) ou SARIMA(0,1,0)(0,1,0)₄.

4.III.D Prévision après une transformation logarithmique

Soit X un processus tel que $Y = \ln X$ soit un (S)AR(I)MA. Connaissant X_1, \dots, X_N , on cherche à déterminer le meilleur prédicteur. On fait l'hypothèse que Y est gaussien. Alors

$$\begin{aligned} \check{X}_N(h) &= \mathbb{E}[X_{N+h} | X_1, \dots, X_N] \\ &= \mathbb{E}[\exp(Y_{N+h}) | Y_1, \dots, Y_N] \\ &= \mathbb{E}[\exp(\check{Y}_N(h) + (Y_{N+h} - \check{Y}_N(h))) | Y_1, \dots, Y_N] \\ &= \exp(\check{Y}_N(h)) \mathbb{E}[\exp(Y_{N+h} - \check{Y}_N(h))] \\ &= \exp(\check{Y}_N(h)) \exp\left(\frac{1}{2} \mathbb{E}[(Y_{N+h} - \check{Y}_N(h))^2]\right) \end{aligned}$$

Un calcul similaire donne pour l'erreur de prévision

$$\mathbb{E}[(X_{N+h} - \check{X}_N(h))^2] = \mathbb{E}[X_{N+h}^2] (1 - \exp(-\mathbb{E}[(Y_{N+h} - \check{Y}_N(h))^2]))$$

4.IV Compléments

4.IV.A Test de Dickey-Fuller.

Supposons qu'on cherche à modéliser $x = (x_1, \dots, x_N)$ par un AR(1) décentré $(1 + \phi B)(X - \mu) = \varepsilon$. Si x est effectivement une réalisation d'un processus AR(1) avec $|\phi| < 1$, alors les estimateurs précédemment vus sont asymptotiquement des gaussiennes de moyenne ϕ et de variance proportionnelle à $1/N$. S'il s'agit en réalité d'un processus intégré, c'est-à-dire si $\phi = -1$, alors ce résultat asymptotique n'est plus vrai. Il convient donc de tester au préalable cette hypothèse.

On pose $\nabla X_t = X_t - X_{t-1} = \mu^* + \phi^* X_{t-1} + \varepsilon_t$, avec $\phi^* = -1 - \phi$. Tester $H_0 : \phi = -1$ contre $H_1 : \phi > -1$ revient à tester $\phi^* = 0$ contre $\phi^* < 0$.

On suppose connu le processus entre les instants 1 et N , et on estime les coefficients μ^* et ϕ^* par une régression linéaire ordinaire. L'estimateur $\hat{\phi}^*$ ainsi obtenu a pour variance

$$\hat{\sigma}_{\hat{\phi}^*}^2 = \frac{\sum_{t=2}^N (\nabla X_t - \hat{\mu}^* - \hat{\phi}^* X_{t-1})^2}{(N-3) \left(\sum_{t=1}^{N-1} X_t^2 - \frac{1}{N-1} \left(\sum_{t=1}^{N-1} X_t \right)^2 \right)}$$

Sous l'hypothèse H_0 , on a le résultat asymptotique suivant :

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\frac{\hat{\phi}^*}{\hat{\sigma}_{\hat{\phi}^*}} \in [-t_\alpha, 0] \right) = 1 - \alpha$$

avec $t_{1\%} = 3,43$, $t_{5\%} = 2,86$ et $t_{10\%} = 2,57$.

Remarque : Il existe une version très similaire de ce test dans le cas d'une modélisation $AR(p)$ avec $p > 1$.