Projet 01 : Estimation de l'évolution du cours d'un actif financier (LS moindres carres)

Introduction.

Soit un processus stochastique $X_n,\ n \in \mathbf{N}$ représentant l'évolution du cours d'un actif en fonction des jours. On suppose que : $X_n = m_n + Y_n$
où la composante m_n, que l'on appelle tendance, est déterministe et le processus stochastique Y_n est stationnaire, centré.
L'objectif de ce projet est d'élaborer, à partir de données historiques $X_n,\ n = 0, \cdots, N-1$, le modèle d'évolution de l'actif.

Etape 1 Analyse de tendance

Hypothèse. On suppose que la tendance du processus est linéaire, c'est à dire qu'il existe deux coefficients réels a et b tels que : $m_n = a + b \cdot n$

1.1 En utilisant la méthode des moindres carrés, trouver des estimations $\hat{a}$ et $\hat{b}$ des coefficients a et b telles que :

$$\sum_n \left(X_n - \hat{a} - \hat{b} \cdot n\right)^2 \quad \to \quad \min$$

1.2 Calculer le signal erreur d'estimation : $Y_n = X_n - \hat{a} - \hat{b} \cdot n$ et analyser ses caractéristiques statistiques : moyenne et fonction d'autocorrélation (fonction d'autocovariance).

Etape 2 Modélisation AR du processus Y_n.
Hypothèse. On travaillera dans cette partie avec des bruits blancs gaussiens.

2.1 A l'aide des équations de Yule-Walker déterminer un filtre formeur de type AR d'ordre p (que vous choisirez) qui représente le mieux le modèle du processus Y_n.

2.2 Simuler un bruit blanc gaussien de variance obtenue par les équations de Yule-Walker. Programmer le fonctionnement du filtre formeur. On notera $\hat{Y}_n$ le signal obtenu par simulation à partir du filtre formeur.

Etape 3 Analyse des résidus

Calculer et tracer la différence : $e_n = Y_n - \hat{Y}_n$
et analyser ses caractéristiques : moyenne et fonction d'autocorrélation. On considère que le modèle obtenu à l'étape précédente est conforme aux données si le signal résiduel e_n est un bruit blanc.

Application

Réaliser les calculs, comparer et analyser les résultats sur deux jeux de données :
– le premier (fichier données.xls) correspond aux données artificielles, obtenues par simulation avec un modèle AR (ordre 2) ;
– le second (fichier AL.xls) correspond aux données réelles.

ANNEXE

1. Fonction de covariance et fonction de corrélation d'un Processus Aléatoire

La Fonction de covariance d'un Processus Aléatoire devient fonction de corrélation lorsque le processus est *SSL* (stationnaire au second ordre au sens large) :

Fonction de Covariance $R_{XY}(t_1,t_2)$ *de 2 Processus Aléatoires* $X(t)$ *et* $Y(t)$ *SSL*

$$R_{XY}(t_1,t_2) = E\left[X^0(t-t_1)Y^0(t-t_2)\right]$$

Fonction de Corrélation $R_{XY}(t_1-t_2)$ *de 2 Processus Aléatoires* $X(t)$ *et* $Y(t)$ *SSL*

$$R_{XY}(t_1-t_2) = E\left[X^0(t-t_1)Y^0(t-t_2)\right] = R_{XY}(t_1,t_2)$$

FILTRES FORMEURS

IDENTIFICATION PAR FILTRES FORMEURS - MODELES AR, MA ARMA

1. Processus MA (*Moving Average* - Moyenne mobile) **d'ordre** M

$$W_n \longrightarrow \boxed{H(z)} \longrightarrow X_n$$

Bruit blanc centré **Filtre MA**

Un signal à TD x_n peut être identifié comme sortie d'un filtre MA excité par un bruit blanc (car il y a ***tout*** dans un *bruit blanc*, un bruit blanc excite tous les modes du processus, il y a toutes les fréquences dans un bruit blanc) :

$$x_n = \sum_{k=0}^{M} b_k w_{n-k} = b_0 w_n + b_1 w_{n-1} + \cdots + b_M w_{n-M}$$ avec $b_0 = 1$ (le fait de diviser tous les coefficients par b_0 conduit à $b_0 = 1$)

où :

w_n désigne un processus aléatoire réel blanc, centré, stationnaire, de variance σ^2

et : $\{b_m\}$ est une séquence de $M+1$ coefficients réels.

Le processus ainsi construit a pour RI la séquence $\{b_m\}$: $\{h_n\} = \{b_n\}$

Ce filtre a donc pour FT :

$$H(z) = TZ[b_n] = \sum_{n=0}^{M} b_n z^{-n} = b_0 + b_1 z^{-1} + \cdots + b_M z^{-M}$$

Cette FT ne possède pas de pôles. Elle conduit donc à un filtre causal stable, facilitant ainsi l'identification du signal x_n.

Relations entre coefficients du modèle et covariances

Malheureusement, les relations entre coefficients du modèle et covariances ne sont pas linéaires.

Il suffit pour s'en convaincre de traiter en exemple le processus MA d'ordre $M = 2$ à TD engendré par :

$$x_n = \sum_{k=0}^{2} b_k w_{n-k} = w_n + b_1 w_{n-1} + b_2 w_{n-2}$$

Calculons l'autocovariance $R_{XX_k} = E[X_{n+k} X_n]$ où : X_n est la VA dont x_n est une réalisation

W_n est la VA dont w_n est une réalisation

(on rappelle qu'un bruit blanc de variance σ^2 est tel que : $E[W_{n+k} W_n] = \sigma^2 \delta_{n,k}$ et que R_{XX_k} est paire)

$$\begin{cases} R_{XX_0} = E\left[X_n^2\right] = E\left[W_n^2\right] + b_1^2 E\left[W_{n-1}^2\right] + b_2^2 E\left[W_{n-2}^2\right] = \left(1 + b_1^2 + b_2^2\right)\sigma^2 \\ R_{XX_1} = E\left[X_{n+1} X_n\right] = E\left[(W_{n+1} + b_1 W_n + b_2 W_{n-1})(W_n + b_1 W_{n-1} + b_2 W_{n-2})\right] = (b_1 + b_1 b_2)\sigma^2 \\ R_{XX_2} = E\left[X_{n+2} X_n\right] = E\left[(W_{n+2} + b_1 W_{n+1} + b_2 W_n)(W_n + b_1 W_{n-1} + b_2 W_{n-2})\right] = b_2 \sigma^2 \\ R_{XX_k} = E\left[X_{n+k} X_n\right] = 0 \quad \text{pour} \quad k \geq M + 1 \end{cases}$$

soit :

$$\begin{cases} R_{XX_0} = \left(1 + b_1^2 + b_2^2\right)\sigma^2 \\ R_{XX_1} = (b_1 + b_1 b_2)\sigma^2 \\ R_{XX_2} = b_2 \sigma^2 \\ R_{XX_k} = 0 \quad \text{pour} \quad k \geq M + 1 \end{cases}$$

Lorsque l'on veut identifier (≡ modéliser) un signal x_n comme un processus MA d'ordre donné (ici 2) d'entrée w_n : $x_n = w_n + b_1 w_{n-1} + b_2 w_{n-2}$ alors la covariance R_{XX_k} est connue (déduite de la séquence $\{x_k\}$ du signal par une méthode d'estimation de la *fonction d'autocovariance*, par exemple, la méthode par *ergodicité*) et b_1, b_2 et σ^2 sont les inconnues.

Le système d'équations est non linéaire et pour le résoudre on fait appel à des algorithmes de Programmation Non Linéaire.

2. Processus AR (*Auto-Regressive* - Auto-Régressif) d'ordre N

$$W_n \longrightarrow \boxed{H(z)} \longrightarrow X_n$$

Bruit blanc centré **Filtre AR**

Un signal à TD x_n peut être vu comme sortie d'un filtre modèle AR excité par un bruit blanc (car il y a ***tout*** dans un *bruit blanc*, un bruit blanc excite tous les modes du processus, il y a toutes les fréquences dans un bruit blanc) :

Modèle AR : $x_n = w_n - \sum_{k=1}^{N} a_k x_{n-k} \quad \rightarrow \quad \sum_{k=0}^{N} a_k x_{n-k} = w_n$ avec $a_0 = 1$

soit : $a_0 x_n + a_1 x_{n-1} + \cdots + a_N x_{n-N} = w_n$

où : w_n désigne un processus aléatoire réel blanc, centré, stationnaire, de variance σ^2

et : $\{a_n\}$ est une séquence de $N+1$ coefficients réels.

Le processus ainsi construit a pour RI la séquence $\{h_n\} = TZ^{-1}[H(z)]$ où $H(z)$ est la FT du filtre :

$$H(z) = \frac{1}{\sum_{n=0}^{N} a_n z^{-n}} = \frac{1}{a_0 + a_1 z^{-1} + \cdots + a_N z^{-N}} \qquad \text{(filtre } \textit{tout-pôle}\text{)}$$

Cette FT possède des pôles. Ceux-ci doivent être de module <1 pour assurer la stabilité du filtre causal.

Relations entre coefficients du modèle et covariances (cas $N=1$)

Prenons pour exemple le processus AR d'ordre $N=1$ à TD engendré par :

$x_n + a_1 x_{n-1} = w_n$ avec $|a_1| < 1$ pour assurer la stabilité du filtre.

En prenant l'espérance des 2 membres, et en utilisant l'hypothèse que w_n est centré, on en déduit que le processus x_n est lui-même centré.

Fonction d'autocovariance : $R_{XX_k} = E[X_{n+k} X_n] = E[X_n X_{n-k}]$ où : X_n est la VA dont x_n est une réalisation

W_n est la VA dont w_n est une réalisation

(on rappelle qu'un bruit blanc de variance σ^2 est tel que : $R_{WW_k} = [W_{n+k} W_n] = \sigma^2 \delta_{n,k}$ et que R_{XX_k} est paire)

Comme : $x_n = w_n - a_1 x_{n-1}$

on a : $x_n^2 = x_n w_n - a_1 x_n x_{n-1} = (w_n - a_1 x_{n-1})w_n - a_1 x_n x_{n-1} = w_n^2 - a_1 x_{n-1} w_n - a_1 x_n x_{n-1}$

d'où, comme il n'y a pas de dépendance causale entre x_{n-1} et w_n

(x_n s'exprime causalement en fonction de w_n , ce qui signifie que x_{n-1} est fonction de w_{n-1} , w_{n-2} , ...

Comme w_{n-1} , w_{n-2} , ... sont indépendants de w_n (bruit blanc), alors x_{n-1} et w_n sont indépendants donc non corrélés, ce qui s'écrit : $E[X_{n-1}W_n] = E[X_{n-1}]E[W_n] = 0$ car le bruit blanc est centré)

on a : $E[X_n^2] = E[W_n^2] - a_1 E[X_n X_{n-1}]$ soit :

$R_{XX_0} = \sigma^2 - a_1 R_{XX_1}$

On a aussi : $x_n x_{n-1} = x_{n-1} w_n - a_1 x_{n-1}^2$

d'où : $E[X_n X_{n-1}] = -a_1 E[X_{n-1}^2]$ (car $E[X_{n-1}W_n] = E[X_{n-1}]E[W_n] = 0$) soit :

$R_{XX_1} = -a_1 R_{XX_0}$

On a donc le système d'équations : $\begin{cases} R_{XX_0} + a_1 R_{XX_1} = \sigma^2 \\ R_{XX_1} + a_1 R_{XX_0} = 0 \end{cases}$

Ce système linéaire d'équations peut s'écrire sous forme matricielle : $\mathbf{R}\,\underline{\mathbf{a}} = \underline{\mathbf{v}}$

avec : $\mathbf{R} = \begin{bmatrix} R_{XX_0} & R_{XX_1} \\ R_{XX_1} & R_{XX_0} \end{bmatrix}$ $\underline{\mathbf{a}} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$ $\underline{\mathbf{v}} = \begin{bmatrix} \sigma^2 \\ 0 \end{bmatrix}$

$a_0 = 1$

La solution de ce système linéaire d'équations est immédiate : $\begin{cases} a_1 = -\dfrac{R_{XX_1}}{R_{XX_0}} \\ \sigma^2 = R_{XX_0} - \dfrac{R_{XX_1}^2}{R_{XX_0}} \end{cases}$

Relations entre coefficients du modèle et covariances (cas N quelconque)

L'expression matricielle permet la généralisation à un modèle AR d'ordre N dont la fonction d'autocovariance vérifie le système linéaire d'équations : (ce sont les équations de **Yule-Walker**)

$$\begin{bmatrix} R_{XX\,0} & R_{XX\,1} & \cdots & R_{XX\,N} \\ R_{XX\,1} & R_{XX\,0} & \cdots & R_{XX\,N-1} \\ \vdots & \vdots & \vdots & \vdots \\ R_{XX\,N} & R_{XX\,N-1} & \cdots & R_{XX\,0} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

ou encore : $\boxed{\mathbf{R}\,\underline{\mathbf{a}} = \underline{\mathbf{v}}}$

$a_0 = 1$

La matrice d'autocovariance $\mathbf{R}$ de terme général $r_{i,j}$ ne dépendant que de $i-j$ ($R_{XX\,k}$ paire) est dite matrice de Toeplitz.

Pour réaliser ce calcul, une inversion matricielle est possible mais on utilise généralement l'algorithme efficace itératif de **Levinson** ou encore de **Burg**, qui évitent une inversion matricielle en mettant à profit le caractère Toeplitz de la matrice $\mathbf{R}$.

La séquence $\{a_n\}$, solution avec σ^2, du système linéaire de **Yule-Walker** à partir de la connaissance de la matrice d'autocovariance $\mathbf{R}$, conduit à la relation de récurrence du filtre.

La détermination de la séquence $\{a'_n\}$ des pôles de

$$H(z)=\frac{1}{1+a_1z^{-1}+\cdots+a_Nz^{-N}}=\frac{1}{(1-a'_1z^{-1})(1-a'_2z^{-1})\cdots(1-a'_Nz^{-1})}$$

conduit, de façon équivalente à celle $\{a_n\}$, à l'obtention du filtre d'identification de la séquence x_n.

L'avantage de la modélisation AR par rapport à l'identification par un modèle MA est qu'elle conduit à un système linéaire d'équations pour la détermination des coefficients du filtre.

Applications :

. synthèse de signal

. compression de signal (de parole par exemple) :

un signal x_k de K échantillons d'un signal de parole peut être codé non pas par mémorisation de la séquence brute $\{x_k\}$ mais compressée pour être ramenée aux coefficients a_i du filtre modèle (AR) d'ordre N et au coefficient σ^2, variance du bruit blanc placé à l'entrée du filtre pour produire, synthétiser le signal x_k. La décompression consiste à générer, synthétiser x_k à partir du bruit blanc de variance σ^2 par application de l'algorithme AR du filtre formeur d'ordre N au bruit blanc.

3. Processus ARMA d'ordre $N-M$

$$W_n \longrightarrow \boxed{H(z)} \longrightarrow X_n$$

Bruit blanc centré **Filtre ARMA**

Un processus ARMA s'obtient en mettant en série une structure AR et une structure MA.

Soit l'équation récurrente :

$$\sum_{k=0}^{N} a_k x_{n-k} = \sum_{i=0}^{M} b_i w_{n-i} \qquad \text{avec} \qquad a_0 = 1$$

(le fait de diviser tous les coefficients par a_0 ne change rien et conduit à $a_0 = 1$)

soit : $a_0 x_n + a_1 x_{n-1} + \cdots + a_N x_{n-N} = b_0 w_n + b_1 w_{n-1} + \cdots + b_M w_{n-M}$

où :

w_n désigne un processus aléatoire réel blanc, centré, stationnaire, de variance σ^2

et : $\{a_n\}$ et $\{b_m\}$ sont des séquences de coefficients réels.

Le processus ainsi construit a pour RI la séquence $\{h_n\} = TZ^{-1}[H(z)]$ où $H(z)$ est la FT du filtre :

$$H(z) = \frac{\sum_{m=0}^{M} b_m z^{-m}}{\sum_{n=0}^{N} a_n z^{-n}} = \frac{b_0 + b_1 z^{-1} + \cdots + b_M z^{-M}}{a_0 + a_1 z^{-1} + \cdots + a_N z^{-N}}$$

Cette FT possède des pôles. Ceux-ci doivent être de module <1 pour assurer la stabilité du filtre causal.

L'équation récurrente admet une solution unique x_n, stationnaire au 2nd ordre, unique, qui s'exprime causalement en fonction de w_n, si et seulement si les racines du dénominateur (≡ pôles de $H(z)$) sont de module <1.
