ANALYSE NUMÉRIQUE – TP N° 2

MÉTHODE DES MOINDRES CARRÉS

semaine du 27 mai 2013

1 Contexte et problématique

L'objectif de ce TP est l'application de la méthode des moindres carrés à la prévision des valeurs des séries chronologiques.

Une *série chronologique* (on dit aussi *série temporelle*) est une suite des valeurs numériques (dans notre cas réelles), indicée par le temps t

$$y(0), y(1), \ldots, y(t), \ldots$$

Ces valeurs représentent l'évolution au cours du temps d'un phénomène dont l'étude nous intéresse soit parce que nous voulons expliquer son comportement, soit parce que nous voulons prévoir son comportement. Les domaines d'applications sont extrêmement variés. Citons, à titre d'exemple, médecine, biologie, épidémiologie, traitement de données, métrologie, assurance, économétrie, finance, traitement du signal, sciences de la Terre, etc.

En règle générale, la prévision $\widehat{y}(t)$ de la valeur $y(t)$ à l'instant t se fait en se référant aux valeurs antérieures de y suivant un modèle. Si le modèle est linéaire, alors on peut exprimer la prévision selon la relation

$$\widehat{y}(t) = a_1 y(t-k_1) + \cdots + a_n y(t-k_n) \ ; \ a_i \in \mathbb{R}, \ k_i \in \mathbb{N}$$

où k_i représente un décalage temporel.

Il est aussi possible d'envisager que les valeurs de la variable y dépendent non seulement des valeurs antérieures de y mais aussi des valeurs d'autres variables x_j, $j = 1, \ldots, q$. Ainsi, dans ce cas le modèle linéaire s'écrit

$$\begin{aligned}
\widehat{y}(t) = \ & a_{01} y(t-k_{01}) + \cdots + a_{0n_0} y(t-k_{0n_0}) \\
& + a_{10} x_1(t) + a_{11} x_1(t-k_{11}) + \cdots + a_{1n_1} x_1(t-k_{1n_1}) \\
& + a_{20} x_2(t) + a_{21} x_2(t-k_{21}) + \cdots + a_{2n_2} x_2(t-k_{2n_2}) \\
& \vdots \\
& + a_{q0} x_q(t) + a_{q1} x_q(t-k_{q1}) + \cdots + a_{qn_q} x_q\left(t-k_{qn_q}\right)
\end{aligned}$$

Les valeurs des coefficients a_{ij} du modèle ainsi que les valeurs des différents décalages k_{ij} du temps ne sont pas connues et il faut donc les estimer.

En ce qui concerne les valeurs a_{ij} on pourrait envisager d'utiliser pour leur estimation la méthode des *moindres carrés*.

Par contre pour l'estimation des différents décalages k_{ij}, il y a plusieurs techniques. Ici on effectuera une *analyse spectrale* de chaque signal x_i et aussi de y à l'aide de la fonction scilab `fft`. L'analyse spectrale d'un signal x consiste à appliquer la transformation de Fourier sur la partie $[x(1), x(2), \ldots, x(N)]$ du signal, avec $N = 2^r$. On considère que la période d'échantillonnage du signal y (et aussi des signaux x_i) est égale à $T = 1$ unité du temps. L'analyse spectrale fournit comme résultat le *spectre du signal*, c'est-à-dire l'amplitude du signal $X(l)$ pour chaque fréquence $f_l = \frac{l}{NT} = \frac{l}{N}$ du signal. En examinant le graphique du spectre, on évalue les fréquences qui ont une grande amplitude par rapport aux autres fréquences. Ce sont essentiellement ces fréquences qui contribuent à la formation du signal. Si donc on on a pour une fréquence particulière f_{l_0} une amplitude $X(l)$, alors le signal présente une périodicité de fréquence l/N et qui influe pour beaucoup à la formation du signal. Il est donc possible d'estimer la valeur du signal à l'instant t en utilisant sa valeur à l'instant $t-l$. Il va de soi que ce raisonnement il faut le faire pour chaque fréquence f_l d'amplitude $X(l)$ importante. Cette démarche nous permet de calculer le décalages k_{ij}.

2 Travail à faire

Les données sont présentées sous la forme d'un tableau des séries chronologiques représentant des données financières issues du marché américain. Il s'agit d'estimer le prix d'une option sur le marché. Pour cela, on dispose de 1530 observations de 5 variables qui sont dans l'ordre

- la date ;
- la *volatilité* de l'option calculée d'après la formule de Black-Sholes. (La volatilité mesure l'ampleur des variations d'une option dans le temps. Elle est liée à la notion en statistique de la variance) ;
- la *date de maturité* qui est la date d'échéance de l'option ;
- le *prix de levée* (ou *prix d'exercice*) de l'option en dollars. C'est le prix auquel une option peut être achetée ou vendue ;
- le *prix d'option* (ou *prime d'option*) en dollars. C'est le prix que l´acheteur d´un contrat d´option verse au vendeur initial pour avoir le droit d´acheter ou de vendre cette option à la date de la maturité.

L'objectif est, en utilisant ce tableau, de prévoir le prix d'option.

La méthode à utiliser est la méthode des moindres carrés. Pour ce faire, on peut utiliser les trois variables – volatilité, maturité et prix de levée – ou certaines parmi elles, avec éventuellement des décalages que l'on calculera à l'aide des analyses spectrales des variables correspondantes. Ainsi on créera un tableau des données X avec autant de colonnes que de variables utilisées. Remarquons que chaque décalage d'une variable constitue une nouvelle variable du tableau avec instant de début et instant de fin différents des autres. Par conséquent le nombre d'observations dans le tableau X dépend de la valeur des décalages. Par exemple considérons la variable volatilité que l'on note x_1, on aura les signaux $x_1(1), \ldots, x_1(1530)$. Si on utilise dans le tableau cette variable sans décalage et aussi avec un décalage de $k_{11} = 10$, alors dans le tableau X on doit trouver une colonne ayant les signaux $[x_1(11), \ldots, x_1(1530)]^\top$ et une autre colonne avec le signaux $[x_1(1), \ldots, x_1(1520)]^\top$. Ainsi ce tableau représente le modèle

$$y(t) = a_{10}x_1(t) + a_{11}x_1(t-10);\ 11 \leq t \leq 1530$$

et il a 1520 lignes. Si, de plus on doit utiliser la variable maturité, notée x_2, avec un décalage $k_{21} = 20$, alors dans le tableau on aura les trois colonnes $[x_1(21), \ldots, x_1(1530)]^\top$, $[x_1(11), \ldots, x_1(1520)]^\top$ et $[x_2(1), \ldots, x_2(1510)]^\top$ et le modèle devient

$$y(t) = a_{10}x_1(t) + a_{11}x_1(t-10) + a_{21}x_2(t-20);\ 21 \leq t \leq 1530$$

Ce tableau a 1510 lignes. En fait on peut établir que le nombre de lignes du tableau X est égal à $1530 - \max_{i,j} k_{ij}$.

Le travail à faire est décomposé comme suit :

1. Utiliser un modèle linéaire simple, c'est-à-dire en utilisant comme entrées les trois variables sans décalages et aussi le prix d'option avec un décalage de 1, à savoir le modèle

$$\widehat{y}(t) = a_{01}y(t-1) + a_{10}x_1(t) + a_{20}x_1(t) + a_{30}y(t) \; ; \; a_{ij} \in \mathbb{R},$$

   Calculer sa qualité en comparant les sorties réelles aux sorties calculées du modèle selon un critère que vous établirez.
2. En utilisant des analyses spectrales, trouver quelles sont les variables et les décalages à utiliser afin d'améliorer la qualité de l'estimation. Établir ainsi un modèle optimale.

   Notons qu'un tel modèle doit faire une erreur d'estimation nettement inférieure à 10 dollars pour chaque prix d'option.
3. Conclusion générale.

3 Livrable

Un rapport écrit impérativement en LaTeX, dans lequel on présente :
- les expériences effectuées avec leurs résultats ;
- les conclusions que vous avez pu en tirer des différentes expériences.

Le livrable doit contenir
- le rapport en format pdf et le source en LaTeX, et
- les programmes Scilab associés à une notice succincte d'utilisation de ces programmes.

Ce livrable sera stocké dan un fichier compacté (format `zip` ou `rar`). Pour l'envoi vous utiliserez à Cergy le site `sifoci`, rubrique `Analyse numérique` et vous cliquerez sur le bouton correspondant à votre groupe de TD. N'oubliez pas de noter sur le nom du fichier votre numéro de binôme.

NB.- Un livrable qui contient seulement une de deux parties – rapport ou programme – aura la note 0 (zéro). De même un livrable dont le programme ne permet pas d'obtenir les résultats indiqués au rapport sera aussi noté 0.

Date de livraison : 14.06.2013

4 Recommandations

Sur la forme, il est vivement conseillé d'utiliser des tableaux et des graphiques, afin d'améliorer la lisibilité des résultats. En particulier les graphiques permettent d'apprécier l'adéquation de l'estimation avec la réalité.

Le rapport doit être écrit selon les règles de l'art, dont un bref rappel est donné par le document "Écriture d'un rapport en Analyse Numérique" qui se trouve sur le site `http //sifoci eisti fr`, rubrique *Analyse Numérique*.