

## Chapitre 9, exercice 5

### Instructions pour étudier les séries du répertoire CH09EX05

*Le répertoire CH09EX05 comporte un exercice de base destiné à tous les apprenants et un exercice avancé réservé aux seuls apprenants de la version avancée du cours.*

#### Exercice de base (Pour tous les utilisateurs du cours)

**Préalable** Le chapitre 9 du cours doit avoir été suivi jusqu'à la page 134.



**Objectif** Le but est d'étudier les autocorrélations de séries temporelles générées au moyen de processus autorégressifs d'ordre 1 ou AR(1).



**Données** Les données sont artificielles. Elles ont été générées au moyen de plusieurs processus autorégressifs d'ordre 1 ou AR(1). La série BLANC, présentée au chapitre 8, générée au moyen d'un processus bruit blanc a été employée pour toutes les simulations. Plusieurs jeux de paramètres sont utilisés. Les séries sont de longueur 400.



#### Structure de l'exercice

L'exercice comporte une partie :

- Dans la partie 1, le but est d'étudier, de manière empirique, les autocorrélations de séries générées au moyen de plusieurs processus autorégressifs d'ordre 1 ou AR(1). Non seulement plusieurs valeurs du paramètre sont employées mais plusieurs longueurs de séries sont considérées.

**Partie 1** Un processus autorégressif d'ordre 1 ou AR(1) est défini à partir d'un processus bruit blanc. Notons  $\{e_t\}$  un processus bruit blanc de moyenne 0 et de variance  $\sigma^2$ , égale à 1, par exemple. Notons  $\{y_t\}$  le processus AR(1). Il est défini par la relation  $y_t - \phi y_{t-1} = e_t$ . Pour générer des séries artificielles selon un processus AR(1), on génère une série par le processus  $\{e_t\}$  puis on calcule les  $y_t$  par la formule ci-dessus. Ceci nécessite 1 valeur initiale. On peut par exemple, prendre  $y_1 = e_1$ .

**Remarque**



1. On recommande souvent d'abandonner un certain nombre des premières données, 30 ou 100, par exemple, pour réduire l'effet du démarrage.

2. Ici, nous avons employé le moteur de calcul de Time Series Expert pour générer les séries. L'algorithme employé permet de ne pas abandonner de données au début parce qu'il n'y a pas d'effet de démarrage. Voir Mélard (1984b).

L'utilisation de Time Series Expert for Windows (TSE) a été déjà présentée lors de la réalisation des exercices 1 à 3. En cas de problème avec les instructions sommaires qui suivent, prière de revoir les instructions détaillées fournies lors de la réalisation de l'exercice 1.

### 1.1 EXAMEN DES AUTOCORRELATIONS D'UNE SÉRIE ARTIFICIELLE

Nous commençons par une série générée par le processus AR(1), définie par  $y_t - \phi y_{t-1} = e_t$ , avec  $\phi = 0,5$ , c'est-à-dire  $y_t - 0,5 y_{t-1} = e_t$ . La série s'appelle AR105.

- ⇒ Choisissez le répertoire de données approprié sur votre disque (pas sur le CD-ROM): menu File ⇒ Open. Choisissez DATA puis CHAP09 puis CH09EX05.
- ⇒ Chargez le problème déjà préparé : ARTIF. Dans le bas de l'écran, il est indiqué que la variable dépendante est BLANC, que l'échantillon d'estimation est 1 – 400 et que les prévisions seront calculées jusqu'en 400.
- ⇒ Pour visualiser la série AR105: menu Graphics ⇒ Series. Choisissez AR105. Cliquez OK.
- ⇒ Pour visualiser les autocorrélations de la série: menu Graphics ⇒ Autocorrelations and partials. Choisissez AR105. Cliquez OK.



Vérifiez quels sont les retards pour lesquels les autocorrélations sont significatives au seuil de 5%.

### 1.1.1 Votre réponse

## 1.2 EXAMEN DES AUTOCORRELATIONS D'AUTRES SÉRIES ARTIFICIELLES

Nous traitons maintenant des séries générées à partir de trois autres processus AR(1), définis par  $y_t - \phi y_{t-1} = e_t$ , mais avec respectivement  $\phi = 0,5$ ,  $\phi = 0,2$  et  $\phi = 0,9$ . Les séries s'appellent respectivement AR1\_5, AR102 et AR109. Pour chacune de ces séries :

- ⇒ Visualisez la série: menu Graphics ⇒ Series. Choisissez son nom. Cliquez OK.
- ⇒ Visualisez les autocorrélations de la série : menu Graphics ⇒ Autocorrelations and partials. Choisissez son nom. Cliquez OK.
- ⇒ Visualisez les autocorrélations d'extraits de la série, en procédant comme indiqué au paragraphe 1.2.

Dans les conclusions relatives aux autocorrélations d'un processus AR(1), on dit dans le cours :

- Les séries générées par un processus AR(1) présentent de l'autocorrélation pour tous les retards, décroissant de manière exponentielle (comme indiqué dans l'exercice) sans ou avec oscillations (comme indiqué dans l'exercice).
- La longueur  $T$  de la série et la valeur du coefficient  $\phi$  ont un effet.
- On dit qu'il n'y a *pas* troncation des autocorrélations.
- Il est intéressant de voir dans les graphiques que ce sont souvent pour les mêmes retards que les autocorrélations sont légèrement significatives.
- C'est parce que toutes les séries ont été générées à partir de la même réalisation d'un bruit blanc.



Essayez de justifier ces conclusions.



### 1.2.1 Votre réponse

On peut en effet montrer (ce sera fait plus loin dans la cadre du *cours avancé*) que les autocorrélations de retard  $k \geq 1$  d'un processus AR(1) de paramètre  $\phi$  sont égales à  $\phi^k$ .

## 1.3 EXAMEN DES AUTOCORRELATIONS PARTIELLES DE SÉRIES ARTIFICIELLES

Dans le cours, nous avons regardé des autocorrélations de différentes séries générées par des processus autorégressifs. Nous les reverrons dans l'exercice 6. Il est rapidement apparu que les autocorrélations ne sont ici d'aucun secours.

### Remarque



Dans un certain nombre de livres, à commencer par Box et Jenkins (1976) et Box et al. (1994), on montre que les autocorrélations de processus autorégressifs sont une combinaison linéaire d'exponentielles et de cosinusoïdes amorties. C'est vrai mais cela n'aide pas beaucoup à trouver l'ordre  $p$  du processus.

On introduit alors les *autocorrélations partielles*. Nous les introduirons brièvement dans la partie A de l'exercice avancé. Il y a plusieurs définitions équivalentes. Malheureusement, la définition la plus simple est encore trop complexe pour ce cours. Disons simplement qu'elles se déduisent des autocorrélations de telle manière que,

- pour un processus bruit blanc, elles sont nulles pour tout  $k > 0$ .
- Pour un processus autorégressif d'ordre 1, elles sont nulles pour tout  $k > 1$ .
- Pour un processus autorégressif d'ordre  $p$ , elles sont nulles pour tout

$$k > p.$$

Les autocorrélations partielles d'une série, que nous notons  $p_k$ , se calculent à partir des autocorrélations de la série, notées  $r_k$ , de la même manière que les autocorrélations partielles d'un processus se calculent à partir des autocorrélations du processus. Nous n'expliquons pas comment les autocorrélations partielles doivent être calculées parce que genre de calcul est relativement complexe et sujet à des erreurs d'arrondi, de sorte que seul un programme d'ordinateur bien conçu peut être recommandé. Un tableur n'est pas conseillé à cette fin.

On peut développer un test de comportement aléatoire ou test de bruit blanc à partir des autocorrélations partielles. Les limites du test de signification, au niveau de 5%, sont les mêmes que pour les autocorrélations, à savoir  $\pm 1,96/\sqrt{T}$ .

Nous allons maintenant regarder systématiquement les autocorrélations partielles et les autocorrélations des différentes séries.

Nous traitons d'abord des séries générées à partir de quelques processus, autorégressifs sauf le premier. Voici les noms des séries et les équations qui ont servi à les générer.

BLANC  $y_t = e_t$

AR109  $y_t - 0,9y_{t-1} = e_t$

AR105  $y_t - 0,5y_{t-1} = e_t$

AR102  $y_t - 0,2y_{t-1} = e_t$

AR1\_5  $y_t + 0,5y_{t-1} = e_t$

Pour chacune de ces cinq séries :

⇒ Visualisez les autocorrélations de la série: menu Graphics ⇒ Autocorrelations and partials. Choisissez le nom. Cliquez OK.

⇒ Pour visualiser les autocorrélations partielles de la série, pressez Enter de plus.

⇒ Visualisez les autocorrélations d'extraits de la série, en

procédant comme indiqué au paragraphe 1.2 de l'exercice 3.

Nous allons maintenant réduire la longueur de la série. La manière la plus simple est de changer, dans la boîte de dialogue Autocorrelation parameters, soit la date de début sur la ligne First date, soit la date de fin sur la ligne Last date, de manière à avoir le nombre d'observations demandé. Par exemple :

- ⇒ Pour visualiser les autocorrélations des 100 dernières données de la série AR105 : menu Graphics ⇒ Autocorrelations and partials. Choisissez AR105. Ensuite sur la ligne First date, tapez 301. Cliquez OK. Pressez Enter pour visualiser les autocorrélations partielles.
- ⇒ Expérimentez avec différentes longueurs de séries extraites de la série AR105 et des autres séries, aussi bien pour les autocorrélations que pour les autocorrélations partielles.



Comment pourrait-on résumer les résultats obtenus?



1.3.1 Votre réponse

Ceci est étudié dans le cadre du cours avancé, à titre facultatif, dans la partie A du présent exercice.

## SYNTHESE

Nous avons introduit les autocorrélations *partielles* et nous avons montré leur intérêt pour caractériser les processus autorégressifs d'ordre 1. Nous avons montré que si l'autocorrélation partielle de retard 1 est souvent statistiquement significative, il n'en est pas de même pour les retards suivants 2, 3, ..., qui sont souvent dans la bande de signification. On en déduit une caractérisation des processus autorégressifs d'ordre 1: les autocorrélations partielles sont tronquées au delà du retard 1. Cette caractérisation est plus simple que l'inspection des autocorrélations, dont on a vu qu'elles

décroissent de façon exponentielle, avec ou sans oscillations. Dans l'exercice 4, partie 2, nous montrerons comment se comportent les autocorrélations partielles de série générées par des processus moyenne mobile.

En conséquence, chaque fois que les autocorrélations présenteront une allure tronquée au delà d'un certain retard  $q$ , il faudra penser à un processus moyenne mobile d'ordre  $q$ . En revanche, quand ce seront les autocorrélations partielles qui présenteront une allure tronquée au delà d'un certain retard  $p$ , il faudra penser à un processus autorégressif d'ordre  $q$ .



## Exercice avancé

(Pour les utilisateurs de la version avancée du cours)

### Préalable



Le chapitre 9 du cours de base et avancé doit avoir été suivi jusqu'à la page 167.

### Objectif



Le but est de développer de manière légèrement plus ample que dans le cours, *à titre facultatif*, la théorie relative aux autocorrélations d'un processus  $AR(1)$ .

### Données

Néant.

### Structure de l'exercice

L'exercice comporte une partie :

- Dans la partie A, le but est de procéder, *à titre facultatif*, à une démonstration mathématique qui explique le résultat des expériences vues dans la partie 1. La démonstration repose sur des concepts relativement simples. Il est conseillé d'attendre la fin de la partie du cours consacrée aux processus autorégressifs avant d'examiner cette partie de l'exercice.



**Partie A** Dans la partie 1, nous avons considéré les autocorrélations de processus AR(1) de manière purement empirique, en nous basant sur des graphiques obtenus pour quelques séries artificielles. Les expériences ont permis de faire varier la valeur du paramètre et la longueur de la série. Ici, nous procédons à une démonstration mathématique qui explique le résultat des expériences. La démonstration repose sur des concepts relativement simples : évaluation d'une espérance mathématique et d'une variance d'une variable aléatoire qui est la somme, ou plus généralement une combinaison linéaire, de deux variables aléatoires, ainsi que l'évaluation d'une covariance entre deux sommes, ou plus généralement deux combinaisons linéaires, de variables aléatoires. Ces concepts ont déjà été rappelés au début de l'exercice 3, partie A, auquel nous renvoyons le lecteur.

### A.a LE PROCESSUS AR(1)

Partons d'un processus autorégressif d'ordre 1. Cela signifie que la variable au temps  $t$  est solution d'une équation de la forme

$$y_t = \phi y_{t-1} + e_t.$$

#### 1) La moyenne

La moyenne du processus est donnée par

$$E(y_t) = E[\phi y_{t-1} + e_t] = \phi E(y_{t-1}) + E(e_t) = \phi E(y_{t-1}).$$

La supposition de stationnarité entraîne que  $E(y_t) = E(y_{t-1})$  et donc  $E(y_t) = \phi E(y_t)$  d'où  $E(y_t) = 0$  sauf si  $\phi = 1$ . La moyenne est donc nulle.

#### 2) La variance

Pour la variance, on a

$$\text{var}(y_t) = \text{var}(\phi y_{t-1} + e_t) = \phi^2 \text{var}(y_{t-1}) + \sigma^2 + 2\phi \text{cov}(y_{t-1}, e_t).$$

Le dernier terme est nul parce que  $e_t$  est indépendant de  $e_{t-1}, \dots$  et donc du passé constitué par les variables aléatoires  $y_{t-1}, y_{t-2}, \dots$

La stationnarité entraîne que  $\text{var}(y_t) = \text{var}(y_{t-1}) = \gamma_0$  et, par conséquent que  $\gamma_0 = \phi^2 \gamma_0 + \sigma^2$ , d'où

$$\gamma_0 = \frac{\sigma^2}{1 - \phi^2}.$$

Puisque  $\gamma_0 \geq 0$ , en tant que variance, il faut que  $1 - \phi^2 > 0$  et donc que

$-1 < \phi < 1$ . C'est la forme que prend la *condition de stationnarité* pour un processus autorégressif d'ordre 1.

### 3) Autocovariances du processus

Multiplions les deux membres de l'équation du processus AR(1) par  $y_{t-k}$  :  $y_t y_{t-k} - \phi y_{t-1} y_{t-k} = e_t y_{t-k}$ , d'où  $E(y_t y_{t-k}) - \phi E(y_{t-1} y_{t-k}) = E(e_t y_{t-k})$  et donc la relation

$$\gamma_k - \phi \gamma_{k-1} = 0.$$

### 4) Autocorrélations du processus

Il s'ensuit que

$\gamma_1 = \phi \gamma_0$	donc	$\rho_1 = \gamma_1 / \gamma_0 = \phi$
$\gamma_2 = \phi \gamma_1$	donc	$\rho_2 = \gamma_2 / \gamma_0 = \phi^2$
...		...
$\gamma_k - \phi \gamma_{k-1} = 0$	donc	$\rho_3 = \gamma_k / \gamma_0 = \phi^k$ , etc.

Si  $0 < \phi < 1$ , la fonction d'autocorrélation se présente donc comme une exponentielle décroissante.

Si  $-1 < \phi < 0$ , il y a simultanément décroissance exponentielle de  $|\rho_k|$  et alternance en signe des  $\rho_k$ . Par exemple, si  $\phi = -0,9$ , on a  $\rho_1 = -0,9$ ,  $\rho_2 = 0,81$ ,  $\rho_3 = -0,729$ , etc.

Nous avons généré des séries temporelles artificielles de longueur 400 à l'aide de ce processus AR(1), en employant toujours la série générée selon un processus bruit blanc de l'exercice 4 du chapitre 8. Les  $y_t$  ont été déterminés à l'aide de l'équation  $y_t = e_t - \phi y_{t-1}$ . Les figures obtenues dans la partie 1 montrent les séries obtenues pour  $\phi = -0,5$ ,  $\phi = 0,2$ ,  $\phi = 0,5$  et  $\phi = 0,9$ .

### Remarques



1. Un processus AR(1) ne pose pas de problème d'inversibilité puisque l'équation qui le définit s'écrit déjà sous la forme traditionnelle  $y_t = f(X_{1t}, X_{2t}, \dots) + e_t$ , avec ici une valeur retardée de  $y$  comme variable explicative.

2. Le problème de la détermination de  $\phi$  en fonction des  $\rho_k$  se présente sous de multiples aspects puisque :

$$\phi = \rho_1 = \frac{\rho_2}{\rho_1} = \dots = \frac{\rho_k}{\rho_{k-1}} = \dots$$

### A.b SPÉCIFICATION ET ESTIMATION D'UN MODÈLE AR(1)

Comme nous venons de le remarquer, les allures exponentielles (avec ou sans alternance de signe) sont évidemment déformées quand on examine les autocorrélations d'une série temporelle générée par un processus AR(1) au lieu du processus lui-même. À l'exception des premières, les autocorrélations sont rapidement non significatives et ceci d'autant plus vite que  $|\phi|$  est proche de zéro et que la longueur  $T$  de la série est petite. Néanmoins, il est généralement assez facile de spécifier un modèle AR(1).

L'estimation de  $\phi$  à partir d'une série temporelle peut se résoudre de plusieurs façons. D'abord, on peut procéder par régression (en se rappelant toutefois que les suppositions du modèle linéaire général n'étant plus remplies, les résultats statistiques deviennent approchés au lieu d'être exacts). Ensuite, on peut utiliser les autocorrélations  $r_k$  de l'échantillon comme estimateurs des  $\rho_k$ . Or, nous avons vu une multitude de façons de lier  $\phi$  aux coefficients d'autocorrélations. Cependant, il est possible de montrer que, dans un certain sens, le meilleur estimateur de  $\phi$  est  $r_1$ .

### A.c CORRÉLATION PARTIELLE ET CORRÉLATION MULTIPLE

Nous nous contentons d'introduire ces notions pour les besoins de la suite. La notion de corrélation multiple a déjà été introduite lors de la définition du coefficient de détermination. Le *coefficient de corrélation multiple* entre une variable  $y$ , d'une part, et  $k - 1$  variables  $x_1, \dots, x_{k-1}$ , d'autre part, est égal au coefficient de corrélation entre  $y$  et la variable des résidus de la régression linéaire de  $y$  en fonction de  $x_1, \dots, x_{k-1}$  (avec constante et au sens des moindres carrés). On peut montrer que ce coefficient, qu'on peut noter  $R(y; x_1 \dots x_{k-1})$ , est la racine carrée (positive) du coefficient de détermination de  $y$  en fonction de  $x_1, \dots, x_{k-1}$ .

Dans le cas de la *distribution de  $k$  variables*  $x_1, \dots, x_{k-1}, x_k$  (ou  $y$ ), le *coefficient de corrélation partielle entre  $x_1$  et  $x_2$ , en éliminant l'influence de  $x_3, \dots, x_k$* , est égal au coefficient de corrélation entre les variables  $x_1^*$  et  $x_2^*$ , définies comme les résidus de la régression linéaire de  $x_1$  et de  $x_2$ , respectivement, en fonction des variables  $x_3, \dots, x_k$ . Notons ce coefficient  $\text{corr}(x_1, x_2; x_3, \dots, x_k)$ . En quelque sorte, c'est la corrélation qui subsiste entre  $x_1$  et  $x_2$  quand on élimine l'effet des autres variables  $x_3, \dots, x_k$ . On peut montrer que  $\text{corr}(x_1, x_2; x_3, \dots, x_k)$  est également le coefficient de corrélation entre  $x_1$  et

$x_2^*$ , c'est-à-dire entre  $x_1$  et les résidus de la régression linéaire de  $x_2$  en fonction des variables  $x_3, \dots, x_k$ , ou entre  $x_1^*$  et  $x_2$ , c'est-à-dire entre  $x_2$  et les résidus de la régression linéaire de  $x_1$  en fonction des variables  $x_3, \dots, x_k$ .

Par exemple, la corrélation entre le nombre de souris capturées dans une région en été et les précipitations tombées au printemps dans cette région peut être de 0,5, non pas parce que les souris femelles sont fécondées par l'eau de pluie, mais simplement parce que de fortes précipitations au printemps impliquent une moisson abondante l'été (coefficient de corrélation de 0,8) et que l'abondance de nourriture favorise la reproduction des souris (coefficient de corrélation de 0,6). Le coefficient de corrélation partielle entre les précipitations et le nombre de souris, en éliminant l'influence de la récolte, est alors de 0,04. Ceci montre que la relation apparente entre les deux variables est due à l'influence d'une troisième variable, la récolte.

Il est possible de démontrer que

$$\text{corr}(x_1, x_2; x_3, \dots, x_k) = b_2(x_1; x_2, x_3, \dots, x_k) \sqrt{\frac{\text{var}(x_2; x_3, \dots, x_k)}{\text{var}(x_1; x_3, \dots, x_k)}} \quad (*)$$

où

- $b_2(x_1; x_2, x_3, \dots, x_k)$  est le coefficient de régression de  $x_2$  dans la régression de  $x_1$  en fonction de  $x_2, x_3, \dots, x_k$ ,
- $\text{var}(x_2; x_3, \dots, x_k)$  est la variance résiduelle de la régression de  $x_2$  en fonction de  $x_3, \dots, x_k$ ,
- $\text{var}(x_1; x_3, \dots, x_k)$  est la variance résiduelle de la régression de  $x_1$  en fonction de  $x_3, \dots, x_k$ .

Il est aussi possible de montrer que les coefficients de corrélation partielle peuvent se calculer uniquement en fonction des coefficients de corrélation entre les  $k$  variables (parfois appelés *coefficients de corrélation totale*), en effectuant un rapport entre deux déterminants.

On peut aussi considérer le coefficient de corrélation partielle entre  $x_1$  et  $x_2$ , en éliminant l'influence de  $x_3$ , seulement, soit  $\text{corr}(x_1, x_2; x_3)$ . On peut montrer que ce coefficient de corrélation partielle peut s'obtenir à partir des coefficients de corrélation par la formule

$$\text{corr}(x_1, x_2; x_3) = \frac{\text{corr}(x_1, x_2) - \text{corr}(x_1, x_3)\text{corr}(x_2, x_3)}{\sqrt{[1 - \text{corr}^2(x_1, x_3)][1 - \text{corr}^2(x_2, x_3)]}}.$$

C'est ainsi que le calcul a été effectué dans l'exemple des souris :

$$(0,5 - 0,6 \times 0,8) / [(1 - 0,6^2)(1 - 0,8^2)]^{1/2} = 0,02/0,48 = 0,04.$$

De même,

$$\text{corr}(x_1, x_2; x_3, x_4) = \frac{\text{corr}(x_1, x_2; x_4) - \text{corr}(x_1, x_3; x_4)\text{corr}(x_2, x_3; x_4)}{\sqrt{[1 - \text{corr}^2(x_1, x_3; x_4)][1 - \text{corr}^2(x_2, x_3; x_4)]}},$$

ou, en permutant les rôles de  $x_3$  et  $x_4$ ,

$$\text{corr}(x_1, x_2; x_3, x_4) = \frac{\text{corr}(x_1, x_2; x_3) - \text{corr}(x_1, x_4; x_3)\text{corr}(x_2, x_4; x_3)}{\sqrt{[1 - \text{corr}^2(x_1, x_4; x_3)][1 - \text{corr}^2(x_2, x_4; x_3)]}}, \quad (*)$$

au choix.

Il y a également des relations entre les coefficients de corrélation partielle et les coefficients de corrélation multiple, par exemple

$$1 - R^2(x_k; x_1, x_2, \dots, x_{k-1}) = [1 - \text{corr}^2(x_k, x_1)][1 - \text{corr}^2(x_k, x_2; x_1)] \dots [1 - \text{corr}^2(x_k, x_{k-1}; x_1, x_2, \dots, x_{k-2})].$$

Cette relation peut être aisément interprétée. Le premier membre représente la proportion de la variance de  $x_k$  qui n'est pas expliquée par  $x_1, \dots, x_{k-1}$ . Le second membre est la proportion de la variance de  $x_k$  qui n'est pas expliquée par  $x_1$  multipliée par la proportion de la variance des résidus de la régression  $x_k$  en fonction de  $x_1$  qui n'est pas expliquée par  $x_2$ , multipliée par la proportion de la variance des résidus de la régression de  $x_k$  en fonction de  $x_1$  et de  $x_2$  qui n'est pas expliquée par  $x_3$ , etc.

Ces notions de corrélation partielle et de corrélation multiple peuvent être introduites aussi bien au niveau de l'échantillon qu'à celui de la population. Il faut néanmoins remarquer que les suppositions sous-jacentes aux inférences statistiques sont plus exigeantes puisqu'elles requièrent la normalité de la population. Cela signifie que la loi de probabilité dans la population doit être *multinormale* (ou *normale multivariée*). En particulier, chaque variable a une loi normale, chaque couple de variable a une loi normale, etc. De même la loi conditionnelle de chaque variable, conditionnellement à certaines des autres variables, est normale, et ainsi de suite. Cette supposition exclut toute possibilité d'asymétrie. De plus les variables ne peuvent être discrètes ni *a fortiori* binaires.

Supposons donc la normalité. Si, en outre, les variables sont mutuellement indépendantes (deux à deux, trois à trois, etc.), donc si tous les coefficients de corrélation totale, partielle et multiple sont nuls dans la population, on peut montrer que :

1) la loi asymptotique (c'est-à-dire quand la taille  $n$  de l'échantillon tend vers l'infini) des coefficients de corrélation totale de l'échantillon est nor-



male de moyenne 0 et de variance  $1/n$  (ce qui justifie le raisonnement relatif à la matrice de corrélation dans le paragraphe sur la sélection des variables);

2) la loi asymptotique des coefficients de corrélation partielle est également normale de moyenne zéro et de variance  $1/n$ ;

3) la loi asymptotique de  $n$  fois le carré du coefficient de corrélation multiple, avec  $k - 1$  variables explicatives, suit une loi  $\chi^2$  à  $k - 1$  degrés de liberté.

### Remarque

Une loi khi carré à  $K$  degrés de liberté est celle de la somme des carrés de  $K$  variables normales centrées réduites et mutuellement indépendantes; voir par exemple Dreesbeke [1988].

## A.d AUTOCORRÉLATIONS PARTIELLES

L'autocorrélation partielle de retard  $k$  est définie à partir de la notion de *coefficient de corrélation partielle* entrevue dans le paragraphe précédent et qui relève du chapitre sur la régression multiple. Nous y avons mentionné que le coefficient de corrélation partielle entre  $X_1$  et  $X_2$ , en éliminant l'influence de  $X_3, \dots, X_r$ , peut être calculé comme un rapport entre deux déterminants.

On appelle autocorrélation partielle de retard  $k$  le coefficient de corrélation partielle entre  $y_t$  et  $y_{t-k}$  en éliminant l'influence de  $y_{t-1}, \dots, y_{t-k+1}$ . On le note ici  $\pi_k$ . On a toujours  $\pi_1 = \rho_1$ . On peut montrer que les  $\pi_k$  peuvent s'obtenir en fonction des  $\rho_k$  à l'aide des formules de rapport entre déterminants suivantes :

$$\pi_k = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \vdots & & & & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & & & & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{vmatrix}}$$

Pour un processus AR(1),



$$\pi_2 = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = 0.$$

et, plus généralement, le déterminant au numérateur vaut 0 pour  $k > 1$  parce que la dernière colonne est égale à la première multipliée par  $\phi$ , compte tenu de ce que  $\rho_j = \phi \rho_{j-1}$ .

### Remarques

1. On peut vérifier cette formule en employant que

$$\text{corr}(x_1, x_2; x_3) = \frac{\text{corr}(x_1, x_2) - \text{corr}(x_1, x_3)\text{corr}(x_2, x_3)}{\sqrt{[1 - \text{corr}^2(x_1, x_3)][1 - \text{corr}^2(x_2, x_3)]}}.$$

et en remplaçant  $x_1$  par  $y_t$ ,  $x_2$  par  $y_{t-1}$ , et  $x_3$  par  $y_{t-2}$ , tout en tenant compte de ce que la corrélation entre  $y_t$  et  $y_{t-1}$ , ou entre  $y_{t-1}$  et  $y_{t-2}$  vaut  $\rho_1$ , et que celle entre  $y_t$  et  $y_{t-2}$  vaut  $\rho_2$ .

2. Du point de vue des calculs, trop complexes pour être effectués manuellement, il ne faut surtout pas employer le calcul des rapports de déterminants. Il existe différents algorithmes, voir par exemple Box *et al.* (1994).

La suite des  $\pi_k$  définit la *fonction d'autocorrélation partielle*. La propriété essentielle des autocorrélations partielles est leur comportement pour un processus autorégressif. La fonction d'autocorrélation partielle d'un processus AR( $p$ ) est *tronquée au-delà de l'ordre  $p$* , c'est-à-dire que  $\pi_k = 0$  pour  $k > p$ .



Les  $\pi_k$  d'un processus sont estimés par les  $p_k$ , qui s'obtiennent à partir des  $r_k$  de la même façon que les  $\pi_k$  sont définis à partir des  $\rho_k$ . Une autre façon de calculer les  $p_k$  consiste à estimer les paramètres d'une *autorégression*, par exemple en considérant le modèle de régression multiple

$$y_t = \phi_1^{(k)} y_{t-1} + \dots + \phi_k^{(k)} y_{t-k} + e_t.$$

### Remarque

On peut vérifier ceci de la façon suivante. En effet, reprenons la formule (\*) du paragraphe A.c, adaptée pour  $k+1$  variables au lieu de  $k$ , et en posant  $x_1 = y_t$ ,  $x_2 = y_{t-k}$ ,  $x_3 = y_{t-1}$ , ...,  $x_{k+1} = y_{t-k+1}$ . Les deux variances résiduelles qui interviennent dans cette formule sont égales car, par stationnarité, la distribution de  $(y_t, y_{t-1}, \dots, y_{t-k+1})$  est la même que celle de  $(y_{t-1}, \dots, y_{t-k+1}, y_{t-k})$ , ou, par renversement du sens du temps, que celle de  $(y_{t-k}, y_{t-k+1}, \dots, y_{t-1})$ . On peut donc calculer l'autocorrélation partielle de retard  $k$  en estimant le coefficient de régression  $\phi_k^{(k)}$ , c'est-à-dire le coefficient du

terme en  $B^k$  dans le modèle autorégressif d'ordre  $k$  :  $p_k = \hat{\phi}_k^{(k)}$ .

### A.e TEST DE BRUIT BLANC BASÉ SUR LES AUTOCORRÉLATIONS PARTIELLES

Un test de bruit blanc peut être réalisé à partir des  $p_k$ . Il faut supposer à cette fin que le processus est gaussien (afin de vérifier la supposition de normalité multivariée de la population qui garantit que la distribution asymptotique des coefficients de corrélation partielle est normale de moyenne zéro et de variance  $1/n$ ). Pour le test de l'hypothèse que  $\pi_k = 0$ , les limites d'acceptation à 5% sont les mêmes que pour les autocorrélations, à savoir :  $\pm 1,96/\sqrt{T}$ .

Nous avons vu, dans la partie 1, l'application de ce test sur la série BLANC de longueur 400 déjà considérée dans le chapitre 8, exercice 4, ainsi que sur la série des 50 dernières données. On observe bien que presque toutes les autocorrélations partielles sont non significatives. Comme la fonction d'autocorrélation, la fonction d'autocorrélation partielle d'un bruit blanc est tronquée à partir du retard 1. Nous avons pu constater que les résultats sont conformes aux espérances.

### SYNTHESE

Nous avons montré, de manière plus générale que dans les exemples de la partie 1, une caractérisation des séries produites par un processus AR(1): bien que les autocorrélations ne soient pas tronquées au delà du retard 1, les autocorrélations partielles sont tronquées au delà du retard 1, ce qui veut dire que celle de retard 1 est non nulle, mais les suivantes ne s'écartent pas de 0 de façon significative.

[Retour au chapitre 9](#)