

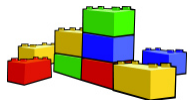
Chapitre 8, exercice 4

Instructions pour employer CH08EX04.XLS et la série BLANC du répertoire CH08EX04

Le fichier CH08EX04.XLS et le répertoire CH08EX04 comportent 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 8 du cours de base doit avoir été suivi jusqu'à la page 83 pour la partie 1 et jusqu'à la page 116 pour la partie 2.

Objectif



Le but de l'exercice est d'introduire les aspects statistiques relatifs à la corrélation avec retard et à l'autocorrélation.

Données



Les données sont des données artificielles produites par un générateur de nombres pseudo-aléatoires de manière à pouvoir apprécier la variabilité des autocorrélations.

Structure de l'exercice

L'exercice comporte deux parties :

- Dans la partie 1, le but de l'exercice est d'étudier la distribution empirique des autocorrélations dans le cas d'un processus bruit blanc et d'introduire le test de bruit blanc.
- Dans la partie 2, le but de l'exercice est d'effectuer dans Time Series Expert for Windows le test de bruit blanc sur une série générée par un processus bruit blanc et expliquer le paradoxe du rejet systématique.

Partie 1 Dans cette partie, on procède comme dans l'exercice 2, c'est-à-dire on calcule les corrélations avec retard et les autocorrélations, sauf qu'au lieu d'une série temporelle artificielle courte sur laquelle les calculs sont aisés à vérifier, on emploie des séries artificielles simulées à l'aide d'un générateur de nombres aléatoires. On en déduit de manière empirique la distribution des autocorrélations dans le cas le plus simple d'observations indépendantes, qu'on appelle processus bruit blanc.

1.1 LES DONNÉES

Les données sont des données artificielles simulées à l'aide d'un générateur de nombres aléatoires. Au lieu d'une seule série courte de 10 observations, on considère autant de séries que l'on veut. Elles sont relativement longues puisqu'elles comportent 400 observations.

⇒ Pour en voir le tableau dans la feuille Main, pressez F5 et sélectionnez DATA ou cliquez sur le lien prévu en haut de la feuille « données artificielles ».

Rappelons que la fonction `RAND()` d'Excel renvoie un nombre pseudo-aléatoire entre 0 et 1 et que la fonction `NORMSINV(x)` renvoie le quantile d'ordre p de la distribution normale centrée réduite.



Examinez les premières cellules des colonnes B à E et essayez d'expliquer leur contenu.



1.1.1 Votre réponse

Les données sont vivantes. Chaque fois qu'un calcul est effectué, les nombres de la colonne B changent de manière aléatoire.

⇒ Pressez la touche fonction F9 pour demander le calcul intégral du classeur.



Que voyez-vous ?



1.1.2 Votre réponse

1.2 L'ASPECT GRAPHIQUE DE LA CORRÉLATION AVEC RETARD

Notons y la variable et $\{y_t\}$ les observations de cette variable. Comme dans les exercices 1, 2 et 3, on peut apprécier de manière graphique l'autocorrélation dans la série. Pour apprécier l'autocorrélation de retard 1, on trace un diagramme de type x - y , en employant, par exemple, les y_t en ordonnées (y) et les y_{t-1} , en abscisses (x). Pour apprécier l'autocorrélation de retard 2, on porte les y_{t-2} en abscisses (x). Pour apprécier l'autocorrélation de retard 3, on porte les y_{t-3} en abscisses (x).

⇒ Cliquez sur l'onglet LAG1 pour visualiser la corrélation éventuelle entre la série et elle-même avec un décalage d'une unité de temps.



Que pensez-vous de cette corrélation de retard 1 ?



1.2.1 Votre réponse

⇒ Cliquez sur l'onglet LAG2 pour visualiser la corrélation éventuelle entre la série et elle-même avec un décalage de deux unités de temps.



Que pensez-vous de cette corrélation de retard 2 ?



1.2.2 Votre réponse

⇒ Cliquez sur l'onglet LAG3 pour visualiser la corrélation éventuelle entre la série et elle-même avec un décalage de trois unités de temps.

**?**

Que pensez-vous de cette corrélation de retard 3 ?

1.2.3 Votre réponse

Revenez par exemple sur le graphique LAG1.



Pressez la touche fonction F9 pour demander le calcul intégral du classeur.

**?**

Décrivez ce que vous voyez ?

1.2.4 Votre réponse

1.3 LES CORRÉLATIONS AVEC RETARD ET LES AUTOCORRÉLATIONS

L'écran est divisé en deux parties. Nous allons d'abord observer les corrélations avec retard. Pour leur définition, voir l'exercice 2, partie 2. Pour les formules, voir l'exercice 2, partie A, dans le cadre du cours avancé.



Cliquez dans la partie supérieure de l'écran de la feuille Main. Allez dans la cellule A1, par exemple au moyen de la combinaison de touches CTRL HOME. Descendez un peu.



Pour en voir le tableau des corrélations avec retard, pressez F5 et sélectionnez CORRELATION ou cliquez sur le lien prévu en haut de la feuille « Corrélations avec retard ».

?

Notez les valeurs des coefficients de corrélation avec retard, de retard 1, 2 et 3 situées dans les cellules C426, D426 et E426.



1.3.1 Votre réponse

Si vous le souhaitez, vous pouvez examiner les formules utilisées pour les calculs.

- ⇒ Revenez à la feuille Main.
- ⇒ Cliquez dans la partie inférieure de l'écran. Allez dans la cellule A1, par exemple au moyen de la combinaison de touches CTRL HOME. Descendez un peu.
- ⇒ Pour en voir le tableau des autocorrélations, pressez F5 et sélectionnez AUTOCORRELATION ou cliquez sur le lien prévu en haut de la feuille « Autocorrélations ».



Notez les valeurs des autocorrélations de retard 1, 2 et 3 situées dans les cellules F840, H840 et J840.



1.3.2 Votre réponse

Si vous le souhaitez, vous pouvez examiner les formules utilisées pour les calculs.

- ⇒ Pressez la touche fonction F9 pour demander le calcul intégral du classeur.



Décrivez ce que vous voyez ?

*1.3.3 Votre réponse*

Synthétisez les constatations réalisées et tentez d'expliquer les résultats.

*1.3.4 Votre réponse*

Si la série temporelle était une vraie série temporelle, s'attendrait-on aux mêmes résultats ?

*1.3.5 Votre réponse***1.4 PROCESSUS ALÉATOIRE ET DE PROCESSUS BRUIT BLANC**

Revenons un instant sur la manière dont les données artificielles ont été générées. Elles l'ont été à l'aide d'un générateur de variables aléatoires normales mais c'est relativement secondaire. Ce qui est important ici c'est le fait d'employer un générateur de nombres aléatoires comme RAND d'Excel.

Il y a beaucoup à dire sur les générateurs de nombres aléatoires. On les appelle souvent générateurs de nombres pseudo-aléatoires. En réalité, les nombres pseudo-aléatoires sont obtenus par des procédés de calcul contrôlés connus qui simulent le hasard, et ceci de manière à éviter les problèmes dans la plupart des applications.

Supposons un instant que RAND soit parfait de telle sorte qu'on puisse considérer que le résultat produit par B25 soit indépendant, au sens des probabilités, du résultat produit par B26 et les autres. On dira alors que le résultat de B25 et de B26 sont des réalisations de deux variables aléatoires

indépendantes.

Si l'on considère toute la série, de longueur 400, on a ici des réalisations de 400 variables aléatoires indépendantes. Chaque observation de la série temporelle est donc considérée comme une réalisation d'une variable aléatoire. Une séquence de variables aléatoires est ce qu'on appelle un processus aléatoire ou processus stochastique. La série temporelle est vue comme une réalisation particulière du processus aléatoire. En particulier, une séquence de variables aléatoires indépendantes est ce qu'on appelle un processus aléatoire de type bruit blanc ou simplement processus bruit blanc. Pour plus de détails, voir la partie A de l'exercice avancé.

1.5 LA DISTRIBUTION EMPIRIQUE DES AUTOCORRÉLATIONS

Dans la suite, nous allons nous baser sur des expériences qui ont été réalisées dans le même contexte que ce qui a fait l'objet des paragraphes précédents. Nous avons généré successivement 100 séries temporelles de longueur 400 et avons gardé trace des autocorrélations de retard 1, 2 et 3. Nous avons constitué un tableau à cette fin.

⇒ Pour atteindre ce tableau, pressez F5 et sélectionnez SAMPLES ou cliquez sur le lien prévu en haut de la feuille « Échantillons d'autocorrélations ».

On voit le numéro de l'échantillon entre 1 et 100, puis les autocorrélations de retard 1, 2 et 3, respectivement dans les colonnes F, H et J. Les résultats sont synthétisés au moyen d'un histogramme.

⇒ Cliquez sur l'onglet DISTRIB pour visualiser la distribution et l'histogramme de la distribution empirique de l'autocorrélation de retard 1.



À quelle distribution ceci vous fait-il penser.

1.5.1 Votre réponse



Les cellules B21 et B22 contiennent respectivement la moyenne et la variance des 100 autocorrélations de retard 1. D'après la théorie statistique, la distribution est, en première approximation, de moyenne 0 et de variance $1/T$.

**?**

Ceci vous surprend-il au vu des résultats ?

1.5.2 Votre réponse

Notez que les 100 échantillons sont vivants dans la mesure où ils sont renouvelés à chaque fois que le classeur est recalculé.

Pressez plusieurs fois la touche fonction F9 en laissant à l'ordinateur le temps de refaire les calculs et regardez les différentes distributions obtenues, ainsi que la moyenne et la variance.

1.6 LE TEST DE BRUIT BLANC OU DE COMPORTEMENT ALÉATOIRE

Le problème n'a pas beaucoup de sens ici. À moins que vous soupçonniez le générateur de nombres aléatoires d'Excel, il n'y a pas de raison à tester que les séries que nous avons observées sont bien produites par un processus bruit blanc. Nous savons que les séries ont un comportement aléatoire. Néanmoins, la question se posera quand nous traiterons des séries réelles ou mieux des séries résiduelles provenant de l'application d'une méthode de prévision. À ce moment, la question prendra tout son intérêt. Il faut donc être prêt à y répondre. Nous pouvons nous entraîner sur les séries simulées, d'autant qu'elles vont nous permettre de préciser certains points intéressants.

Les conclusions du paragraphe précédent impliquent que la probabilité d'avoir une autocorrélation de retard 1 qui soit inférieure à $-1,96/\sqrt{T}$ ou qui soit supérieure à $1,96/\sqrt{T}$ est approximativement égale à 0,05. Par conséquent, si la série temporelle est produite par des variables aléatoires indépendantes, ce qu'on appelle un processus bruit blanc, on s'attend avec une probabilité de 0,95 à ce que l'autocorrélation de retard 1 soit comprise entre $-1,96/\sqrt{T}$ et $1,96/\sqrt{T}$. Et il en est de même pour ce qui concerne les autocorrélations pour les autres retards.

?

Que valent ces limites ici ?



1.6.1 Votre réponse

On dira qu'on rejette l'hypothèse de bruit blanc, au niveau de probabilité de 5 %, si l'autocorrélation est hors de l'intervalle défini par les limites $-1,96/\sqrt{T}$ et $1,96/\sqrt{T}$.

- ⇒ Revenez à la feuille Main.
- ⇒ Allez dans la cellule A1, par exemple au moyen de la combinaison de touches CTRL HOME. Descendez un peu.
- ⇒ Pour voir le tableau des autocorrélations, pressez F5 et sélectionnez AUTOCORRELATION ou cliquez sur le lien prévu en haut de la feuille « Autocorrélations ».
- ⇒ Pressez la touche fonction F9 de manière répétée pour demander plusieurs fois le calcul intégral du classeur.

? Observez une des 3 autocorrélations et notez 1 chaque fois qu'on rejette l'hypothèse et 0 chaque fois qu'on l'accepte.



1.6.2 Votre réponse

? Si vous avez la patience de faire ceci de manière répétée, pensez-vous que vous aurez un ou plusieurs rejets.



1.6.3 Votre réponse

? Si vous avez assez de patience pour le faire un grand nombre

de fois, pourquoi devrait-on s'attendre à 5 % de rejets ?

1.6.4 Votre réponse



SYNTHÈSE

Nous avons généré des séries temporelles artificielles de longueur 400. Chaque observation est produite par le générateur de variables aléatoires, selon une loi normale. Nous avons eu l'occasion de découvrir de manière empirique, au travers de 100 échantillons, la distributions des autocorrélations de retard 1. Dans le contexte utilisé, c'est-à-dire celui d'une série générée par T variables aléatoires indépendantes de même distribution, la théorie statistique montre que la distribution de l'autocorrélation de retard k est normale de moyenne 0 et de variance $1/T$, donc d'écart-type $1/\sqrt{T}$, ceci pour des séries suffisamment longues. On écrit en bref $r_k \sim N(0 ; 1/T)$. C'est ce que nous avons pu vérifier de manière expérimentale.

Ceci implique que la probabilité d'avoir une autocorrélation de retard 1 qui soit inférieure à $-1,96/\sqrt{T}$ ou qui soit supérieure à $1,96/\sqrt{T}$ est approximativement égale à 0,05. Par conséquent, si la série temporelle est produite par des variables aléatoires indépendantes, ce qu'on appelle un processus bruit blanc, on s'attend avec une probabilité de 0,95 à ce que l'autocorrélation de retard 1 soit comprise entre $-1,96/\sqrt{T}$ et $1,96/\sqrt{T}$. Et il en est de même pour ce qui concerne les autocorrélations pour les autres retards.

Nous avons pu étudier de manière expérimentale aussi bien le test que l'interprétation de la probabilité de 5 % qui est appelée le niveau du test.

Partie 2

Nous allons effectuer une étude similaire à celle de la partie 1, mais cette fois en employant Time Series Expert for Windows au lieu d'Excel. Au lieu de nous limiter à des retards 1 à 3, nous allons considérer des retards plus élevés, comme ceux que nous utiliserons pour des séries temporelles mensuelles, c'est-à-dire au moins 2 ans ou 24 mois. Chacune des 400 observations de la série BLANC a été obtenue à l'aide d'un générateur de variables pseudo-aléatoires normales de moyenne 0 et d'écart-type 1. On peut donc considérer que la série BLANC est une réalisation d'un processus bruit blanc.

2.1 AUTOCORRÉLATIONS DE LA SÉRIE BLANC

L'utilisation de Time Series Expert pour la détermination des autocorrélations a déjà été présentée lors de la réalisation de l'exercice 3 du chapitre 8. En cas de problème avec les instructions sommaires qui suivent, prière de revoir les instructions détaillées fournies lors de la réalisation de l'exercice 3 du chapitre 8.

- ⇒ Pour lancer le logiciel, suivez les instructions données en annexe de l'introduction du cours.
- ⇒ Choisissez le répertoire de données approprié sur votre disque (pas sur le CD-ROM) : menu File ⇒ Changing Data Directory. Choisissez DATA puis CHAP08 puis CH08EX04.
- ⇒ Chargez le problème déjà préparé : BLANC. Vous voyez dans le bas de l'écran 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 graphiquement la série BLANC : menu Graphics ⇒ Series. Sélectionnez BLANC et cliquez Open puis OK.

**Remarque**

Des instructions plus détaillées ont été données sous forme de remarques lors de la partie 1 de l'exercice 3. Prière de s'y référer.



Que pensez-vous de la série ?



2.1.1 Votre réponse

⇒ Pour visualiser graphiquement les autocorrélations de la série BLANC : menu Graphics ⇒ Autocorrelations and partials. Sélectionnez BLANC (cliquez sur ce nom puis sur Open). Cliquez OK pour obtenir le graphique.

Comme pour tous les graphiques de TSE, il est possible de focaliser sur un point. Ceci permet de lire la valeur des autocorrélations correspondant à un retard donné.



Que valent les autocorrélations de retard 1 ? De retard 2 ? De retard 12 ? De retard 24 ?



2.1.2 Votre réponse

⇒ Quittez le graphique.

2.2 TEST DE BRUIT BLANC SUR LA SERIE BLANC

Rappelons (voir l'exercice 3) que les lignes discontinues correspondent aux valeurs critiques pour des tests au niveau de probabilité de 5 %. Elles sont données par la formule $\pm 1,96/\sqrt{T}$. Considérons un retard k spécifié. Le fait que la barre représentant l'autocorrélation pour ce retard passe *au-dessus de la ligne discontinue du haut* indique que cette autocorrélation est statistiquement significative. De même, le fait que la barre représentant l'autocorrélation pour ce retard passe *au-dessous de la ligne discontinue du bas* indique que cette autocorrélation est statistiquement significative.



Vérifiez la valeur critique donnée par la formule ci-dessus par rapport au graphique. Est-ce correct ?



2.2.1 Votre réponse



Parcourez les retards. Pour chacun de ces retards, examinez si l'autocorrélation correspondante est statistiquement significative. Dans ce cas, notez le retard et la valeur de l'autocorrélation.



2.2.2 Votre réponse



En conclusion, les tests de bruit blanc associés aux autocorrélations de la série des résidus BLANC sont-ils en accord avec la manière dont la série a été construite ?



2.2.3 Votre réponse



Comment peut-on interpréter ce résultat et quels en sont les implications pratiques ?



2.2.4 Votre réponse

2.3 EFFET DE LA LONGUEUR DES SÉRIES SUR LES AUTOCORRÉLATIONS

Nous avons expliqué le calcul des valeurs critiques par la formule $\pm 1,96/\sqrt{T}$, mais il serait utile de confirmer ceci de manière expérimentale. À cette fin, nous allons 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 souhaité. Par

exemple :

⇒ Pour visualiser les autocorrélations des 100 dernières données de la série BLANC : menu Data ⇒ Autocorrelations and partials. Choisissez BLANC. Ensuite sur la ligne First date, tapez 301 au lieu de 1, puis cliquez OK.



Remarques

1. Pour changer la date de fin, déplacez le curseur sur la ligne Last date, tapez la date de fin, par exemple 350. Ce n'est pas demandé ici.
2. La date de début doit être antérieure à la date de fin.
3. Les séries artificielles traitées ici sont non datées. Quand on traite des données annuelles, mensuelles ou trimestrielles, les dates introduites doivent être valides, c'est-à-dire notamment
 - des nombres sans décimales, pour des données annuelles,
 - des nombres avec une décimale, la partie décimale variant de 1 à 4, pour des données trimestrielles,
 - des nombres avec deux décimales, la partie décimale variant de 1 à 12, pour des données annuelles.

⇒ Expérimentez avec différentes longueurs de séries extraites de la série BLANC.



Comment pourrait-on résumer les résultats obtenus ?



2.3.1 Votre réponse

2.4 ÉTUDE DE RÉSISTANCE DES AUTOCORRÉLATIONS

Nous allons utiliser la série modifiée en 2 points afin d'examiner les effets de ces changements sur les autocorrélations. Nous allons simuler une grève au temps 200 qui réduit la production de 10 unités et un rattrapage au temps 201 qui consiste à ajouter 10 unités supplémentaires. Cette série a déjà été créée et s'appelle MODIF. Nous allons simplement vérifier qu'elle est comme souhaitée. Nous employons pour cela le tableur incorporé à TSE.

- ⇒ Pour employer le tableur : menu Data ⇒ Spreadsheet.
- ⇒ Chargez la série BLANC : pressez la touche fonction F3; sélectionnez la série BLANC. La série se charge dans la colonne A du tableau.
- ⇒ Cliquez dans la colonne B et chargez MODIF.
- ⇒ Cliquez dans la colonne C. Utilisez le menu Operation du tableur et cliquez Enter Oper pour effectuer une opération entre les colonnes du tableau. Cliquez B dans la colonne Operand 1 puis – comme opérateur puis A dans la colonne Operand 2 et cliquez OK. Pressez la touche fonction F8. Tapez successivement « B » puis « – » puis « A ». La colonne C contient maintenant la différence des colonnes B et A.
- ⇒ Avec la flèche vers le bas ou l'ascenseur, descendez sur les temps 200 et 201.



La série est-elle bien la série désirée ?



2.4.1 Votre réponse

- ⇒ Pour visualiser graphiquement la série : menu Graphics ⇒ Series. Sélectionnez MODIF.

Inspectez les autocorrélations de cette série modifiée en procédant comme suit.

- ⇒ Pour visualiser les autocorrélations de la série MODIF : menu Data ⇒ Autocorrelations and partials. Sélectionnez MODIF. La suite est comme ci-dessus.



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



2.4.2 Votre réponse

Nous allons maintenant apprécier la corrélation avec retard de manière graphique.

- ⇒ Pour visualiser graphiquement les séries : menu Data ⇒ Xy graphic/Scatter diagram
- ⇒ Cliquez sur Select Y. Sélectionnez MODIF (cliquez sur ce nom puis sur Open).
- ⇒ Cliquez sur Select X. Sélectionnez de même MODIF.
- ⇒ Dans le cadre Regression line cliquez sur le bouton Yes. Cliquez OK pour obtenir un graphique — sans intérêt parce que tous les points sont sur une droite.
- ⇒ Pressez la touche Page Down pour retarder la variable sur l'axe horizontal par rapport à l'axe vertical. Le retard de 1 est indiqué. Chaque fois que vous pressez Page Down, le retard augmente.

?

Que constatez-vous ? Que pouvez-vous en déduire ?



2.4.3 Votre réponse

SYNTHÈSE

Nous avons d'abord examiné les autocorrélations d'une série produite par un processus bruit blanc. Nous avons constaté que le test de bruit blanc conduit à un rejet pour plusieurs retards mais que cela peut s'expliquer assez simplement. Il n'y a donc pas de raison,

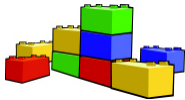
dans ce cas, de mettre en cause la manière dont la série artificielle a été générée. Nous avons confirmé l'étude en faisant varier la longueur de la série. Enfin, nous avons examiné le manque de résistance ou de robustesse des autocorrélations



Exercice avancé

(Pour les utilisateurs de la version avancée du cours)

Préalable



Le chapitre 8 du cours de base et du cours avancé doit avoir été suivi jusqu'à la page 148 pour la partie A et jusqu'à la page 178 pour la partie B.

Objectif



Le but est ici de présenter les formules des corrélations et des autocorrélations et de justifier en partie l'avantage des ces dernières.

Données



Pour la partie A, les mêmes données que celles de l'exercice de base, partie 2. Pour la partie B, les mêmes données que celles de l'exercice de base, partie 1.

Structure de l'exercice

L'exercice avancé comporte deux parties :

- Dans la partie A, le but de l'exercice est d'introduire, de manière plus précise que dans le cours de base, les concepts de processus aléatoire, de processus stationnaire, de processus bruit blanc et le test de bruit blanc.
- Dans la partie B, le but de l'exercice est de généraliser le test de bruit blanc pour un retard déterminé, ou test individuel, qui a été vu dans la partie 1 de l'exercice de base, et de présenter le test global portant sur plusieurs autocorrélations, plus précisément sur la somme des carrés de ces autocorrélations.

Partie A Dans cette partie, on introduit les processus aléatoires, les processus stationnaires, les processus bruit blanc et les tests de bruit blanc de manière plus précise que dans le cours de base.

A.a LES PROCESSUS ALÉATOIRES

Dans l'introduction du chapitre, l'accent a été mis sur la nécessité de disposer d'un concept de population afin d'appliquer les méthodes statistiques aux échantillons que sont les séries chronologiques. Chaque observation est considérée comme la réalisation d'une variable aléatoire. La collection de ces variables aléatoires s'appelle un *processus aléatoire*. Nous sommes amenés à imposer des contraintes à ces variables aléatoires, qui ne pourront pas être distribuées n'importe comment. Provisoirement, nous désignons par Y_t la variable aléatoire correspondant à l'observation y_t .

A.b LES PROCESSUS STATIONNAIRES

Il faut maintenant définir pour la population, c'est-à-dire pour le processus aléatoire $\{Y_t\}$, les différentes notions introduites au niveau de la série chronologique. Commençons par la moyenne \bar{y} . Pour justifier son emploi, il convient que chaque variable aléatoire Y_t , quel que soit t , ait la même espérance mathématique ou moyenne $E(Y_t)$, soit m . La moyenne est alors une estimation de m . Un raisonnement identique pour la variance conduit à supposer que les variables aléatoires Y_t ont des variances identiques σ^2 . Ceci conduit à considérer

$$\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2$$

comme une estimation de $\sigma^2 = \text{var}(Y_t) = E[(Y_t - m)^2]$.

Envisageons enfin l'autocorrélation. Au numérateur on trouve

$$\frac{1}{T} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})$$

qui devrait estimer $\text{cov}(Y_t, Y_{t-k}) = E[(Y_t - m)(Y_{t-k} - m)]$. Or, cette expression dépend non seulement du retard k mais aussi du temps t . Il faut donc imposer qu'elle ne dépende pas de t . On est donc conduit à formuler des contraintes sur les processus aléatoires à prendre en considération. Un processus aléatoire $\{Y_t\}$ est *stationnaire* s'il remplit les conditions suivantes :

- 1) $E(Y_t)$ ne dépend pas de t et vaut m ;
- 2) $\text{var}(Y_t) = E[(Y_t - m)^2]$ ne dépend pas de t et vaut γ_0 ;

3) $\text{cov}(Y_t, Y_{t-k}) = E[(Y_t - m)(Y_{t-k} - m)]$ ne dépend pas de t et vaut γ_k .

Remarque



Les suppositions (connues sous le nom de stationnarité faible ou de stationnarité du second ordre) ne sont pas suffisantes mais nous nous en contenterons ici. Il faut parfois supposer la stationnarité dite forte exprimant qu'il y a invariance dans le temps de toutes les caractéristiques du processus.

A.C LES AUTOCORRÉLATIONS DE RETARD k

Les définitions suivantes sont très naturelles dans le cas de processus stationnaires :

l'autocovariance de retard k : $\gamma_k = \text{cov}(Y_t, Y_{t-k})$

l'autocorrélation de retard k :

$$\rho_k = \frac{\text{cov}(Y_t, Y_{t-k})}{\text{var}(Y_t)} = \frac{\gamma_k}{\gamma_0}.$$

Le retard k varie de 0 à l'infini. On peut aussi considérer k négatif, en notant que $\rho_{-k} = \rho_k$. On parle aussi de la *fonction d'autocorrélation* du processus comme constituée de toutes les autocorrélations de retard k :

$\{\rho_0 = 1, \rho_1, \rho_2, \dots\}$ ou $\{\rho_k\}$.

Remarque



Une partie de la théorie statistique suppose, en outre, que le processus $\{Y_t\}$ est *gaussien* ou *normal*, c'est-à-dire que la distribution de n'importe quel sous-ensemble des Y_t est une loi *multinormale* (ou normale multivariée). En particulier, chacun des Y_t a une loi normale. Si le processus $\{Y_t\}$ est gaussien, toute combinaison linéaire des variables aléatoires Y_t est normale. Une caractéristique des distributions multinormales est que l'indépendance et la non-corrélation s'y confondent. Deux variables non corrélées sont indépendantes. En général, l'indépendance entraîne la non-corrélation, mais pas l'inverse.

A.d LES PROCESSUS DE TYPE BRUIT BLANC

Avant d'introduire les processus stationnaires plus généraux, considérons le plus simple d'entre eux : le *processus bruit blanc*. Il s'agit d'une suite de variables aléatoires e_t de même distribution (même moyenne et variance) et *mutuellement indépendantes* :

1) on peut supposer que ces variables aléatoires sont de moyenne nulle,

$E(e_t) = 0$, sans restreindre la généralité;

2) elles sont de même variance $E(e_t^2) = \sigma^2$;

3) puisque e_t est indépendante de e_s quand $t \neq s$, ces deux variables aléatoires sont non corrélées, c'est-à-dire que $E(e_t e_s) = 0$.

Dire que les $\{e_t\}$ constituent un processus bruit blanc implique que (pour tout t)

$$\rho_k = \text{corr}(e_t, e_{t-k}) = 0, \quad k = 1, 2, \dots$$

A.e LES TESTS DE BRUIT BLANC OU TEST DE COMPORTEMENT ALÉATOIRE

Revenons au problème de départ. Il s'agit de tester l'hypothèse qu'un processus aléatoire (par exemple celui qui génère les erreurs de prévision d'horizon 1) est un bruit blanc. Cela peut se faire en testant que les autocorrélations (du processus) sont nulles. Deux types de tests sont envisagés dans la suite. Il s'agit, d'une part, de tests individuels relatifs à une autocorrélation particulière, de retard k , utiles quand on craint précisément qu'il y ait de l'autocorrélation de retard k dans le processus. D'autre part, un test global sur plusieurs autocorrélations peut être effectué quand aucun retard particulier n'est suspecté *a priori*. Il sera évidemment impossible de bâtir un test portant sur tous les retards puisque les séries dont on dispose sont de longueur finie.

Remarque

Les processus aléatoires sont généralement considérés pour t allant de $-\infty$ à ∞ et non pour t allant de 1 à T , pour plusieurs raisons. D'abord, on s'intéresse aux valeurs futures, donc à la distribution des variables aléatoires Y_{T+1} , Y_{T+2} , etc. Ensuite, l'idée même de stationnarité implique la pérennité. Dans le cas contraire, il faudrait préciser les valeurs de t et de k pour lesquelles les propriétés sont vraies.



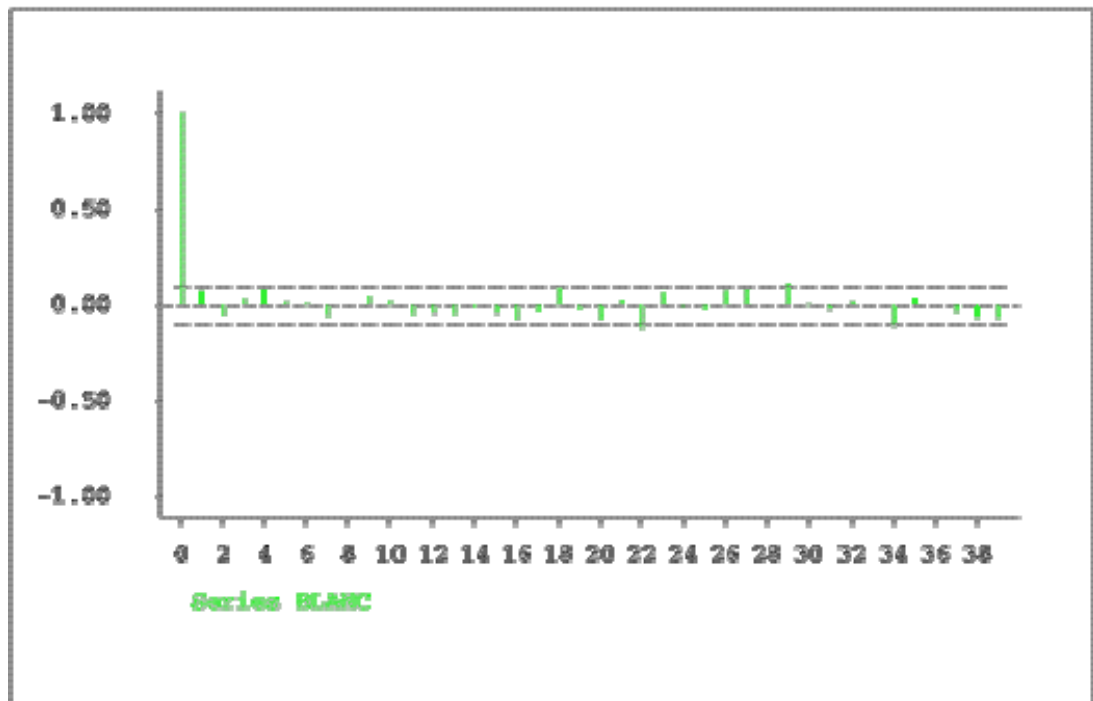
Que le processus des $\{e_t\}$ constitue un processus bruit blanc implique que $\rho_1 = \rho_2 = \dots = 0$. Avant d'examiner un test global portant sur plusieurs retards (voir la partie B), considérons le cas où on s'intéresse à un retard k particulier. C'est encore un test de bruit blanc mais avec comme contre-hypothèse le fait qu'il y a de l'autocorrélation de retard k dans le processus. On met ceci en évidence en écrivant l'hypothèse à tester sous la forme :

$$H_k : \rho_k = 0.$$

Si le processus est un bruit blanc et pourvu que T soit grand, la statistique de test $\sqrt{T} r_k$ a une loi normale centrée réduite, $N(0 ; 1)$. On rejette donc

l'hypothèse, au niveau de probabilité de 5 %, si la statistique a une valeur inférieure à $-1,96$ ou supérieure à $1,96$. Il est toutefois préférable d'exprimer la décision en terme de coefficient d'autocorrélation r_k . Celui-ci a approximativement une distribution $N(0 ; 1/T)$. On rejette donc l'hypothèse de bruit blanc si $\sqrt{T} r_k < -1,96/\sqrt{T}$ ou si $\sqrt{T} r_k > 1,96/\sqrt{T}$. Cette autocorrélation est dite *significative*. Une autocorrélation située entre $-1,96/\sqrt{T}$ et $1,96/\sqrt{T}$ est dite *non significative* (au niveau de 5%). Pour fixer les idées, ces limites de signification sont de $\pm 0,5$ si $T = 16$, de $\pm 0,2$ si $T = 100$ et de $\pm 0,1$ si $T = 400$. Il faut noter que $T = 16$ est insuffisant pour que l'approximation normale soit valable.

Dans la partie 2, nous avons examiné la série BLANC, pour laquelle nous avons obtenu les autocorrélations suivantes.



Il est maintenant possible de commenter les lignes horizontales parallèles à l'axe des abscisses de cette figure qui montre les autocorrélations d'une série artificielle de longueur $T = 400$, engendrée à partir d'un processus bruit blanc. Il s'agit des limites de signification au niveau de 5 %, égales à $\pm 1,96 \times (1/20) = \pm 0,1$. On voit donc qu'il y a quelques autocorrélations significatives, la plus importante pour le retard 22. Nous commentons ceci davantage plus loin mais nous pouvons déjà remarquer qu'on doit s'attendre à $5 \% \times 40 = 2$ rejets en moyenne sur les 40 tests statistiques effectués au niveau de 5 %, en supposant que ceux-ci soient indépendants

SYNTHÈSE

Nous avons présenté les concepts de processus aléatoire, de pro-

cessus stationnaire, de processus bruit blanc et le test de bruit blanc, de manière plus complète que dans le cours de base.
--

Partie B Dans la partie 1 de l'exercice de base, nous avons introduit le test de bruit blanc pour un retard déterminé, ou test individuel, qui porte sur une seule autocorrélation. L'idée ici est de généraliser ce test en un test global portant sur plusieurs autocorrélations.

B.a INTRODUCTION

Plus précisément la statistique de test sera la longueur de la série multipliée par la somme des carrés des autocorrélations de retard 1, 2 jusque K , où K est un nombre entier choisi à l'avance :

$$Q = T \sum_{k=1}^K r_k^2.$$

La raison de ce choix est le suivant. Nous avons vu que, dans le cas d'un processus bruit blanc et pour des séries suffisamment longues, l'autocorrélation r_k a approximativement une distribution $N(0 ; 1/T)$ et $\sqrt{T} r_k$ a une distribution $N(0 ; 1)$. Par conséquent, $(\sqrt{T} r_k)^2$ a une distribution χ^2 (khi deux) à 1 degré de liberté. Il est possible de montrer, toujours dans le cas d'un processus bruit blanc et pour des séries suffisamment longues, que les autocorrélations r_1, r_2, \dots, r_K , sont K variables aléatoires indépendantes. Comme une somme de K variables indépendantes ayant des distributions χ^2 à 1 degré de liberté a une distribution χ^2 à K degrés de liberté, nous avons ici une possibilité de réaliser un test statistique qui porte le nom de test de Box et Pierce.

Remarques



1. Une loi chi carré ou chi deux à K degrés de liberté est celle de la somme des carrés de K variables normales centrées réduites et mutuellement indépendantes; il s'agit ici des $\sqrt{T} r_k$ pour $k = 1, \dots, K$.

2. Certains logiciels fournissent la statistique de Ljung et Box Q' , qu'il est préférable d'utiliser au lieu de Q , sachant que

$$Q' = T(T+2) \sum_{k=1}^K \frac{r_k^2}{T-k}.$$

Les valeurs critiques sont identiques, c'est-à-dire obtenues à partir d'une loi khi deux.

3. Quand ces tests sont appliqués à des résidus, par exemple d'une méthode de lissage exponentiel, il convient de retirer un degré de liberté par paramètre estimé.

B.b LA STATISTIQUE

Nous avons complété les expériences de la partie 1 où nous avons généré successivement 100 séries temporelles de longueur 400 et avons gardé trace des autocorrélations de retard 1, 2 et 3. Un tableau a été constitué à cette fin.

⇒ Pour en voir le tableau dans la feuille Main du classeur CH08EX04.XLS, pressez F5 et sélectionnez SAMPLES ou cliquez sur le lien prévu en haut de la feuille “Echantillons d'autocorrélations”.

On voit le numéro de l'échantillon entre 1 et 100, puis les autocorrélations de retard 1, 2 et 3, respectivement dans les colonnes F, H et J. Un peu plus à droite, dans la colonne L, on a ajouté le calcul de la statistique.

? Par exemple, sur la ligne 845, on peut voir la formule suivante $=400*(F845^2+H845^2+J845^2)$. Interprétez-la.

B.b.1 Votre réponse



Les résultats sont synthétisés au moyen d'un histogramme.

⇒ Cliquez sur l'onglet DISTRIB pour visualiser la distribution et descendez en-dessous de l'histogramme de la distribution empirique de l'autocorrélation de retard 1.

? Ceci vous fait-il penser à une distribution normale ?

B.b.2 Votre réponse



B.C LE TEST GLOBAL DE BRUIT BLANC OU DE COMPORTEMENT ALÉATOIRE

D'après la théorie statistique, esquissée dans le paragraphe précédent, sous l'hypothèse d'un processus bruit blanc et pourvu que la série soit suffisamment longue, la distribution est, en première approximation, de type χ^2 à 3 degrés de liberté.

Dans le cas où l'hypothèse de bruit blanc n'est pas vraie — ce qui n'est pas le cas ici, en principe — on s'attend à ce que les autocorrélations soient assez éloignées de 0 donc que la somme de leur carrés, qui est nécessairement positive, soit assez grande. On dira qu'on rejette l'hypothèse de bruit blanc, au niveau de probabilité de 5 %, si la statistique de Box et Pierce dépasse la valeur critique obtenue comme suit : la valeur de la variable χ^2 à 3 degrés de liberté qui laisse à sa droite une probabilité de 0,05.

Nous avons placé dans la cellule D44 la valeur critique appropriée, pour le test au niveau de probabilité de 5 %.



Comment est obtenue cette valeur critique ? Que vaut-elle ?

B.b.3 Votre réponse



Combien parmi les 100 échantillons fournissent une valeur supérieure à cette valeur critique. Est-ce surprenant ?

B.b.4 Votre réponse



Remarques

1. Il peut arriver que le test global rejette l'hypothèse de bruit blanc alors qu'un des tests individuels ne rejette pas cette hypothèse.
2. Il peut arriver que le test global ne rejette pas l'hypothèse de bruit blanc alors qu'un des tests individuels la rejette.



SYNTHÈSE

Après le test individuel qui porte sur une autocorrélation de retard

déterminé, nous avons considéré ici le cas du test global qui porte simultanément sur plusieurs retards. C'est le test de Box et Pierce, dont il existe des variantes considérées comme plus précises, notamment le test de Ljung et Box.

[Retour au chapitre 8](#)