U.P.S., TOULOUSE version Janvier 2001 par lafosse@cict.fr

1 - INTRODUCTION AUX SÉRIES CHRONOLOGIQUES

1.1 **Objectifs** de ce cours sur les chroniques

_ la description.
_ le filtrage pour extraire, présenter, ou analyser, une partie de l'information.
_ la modélisation stochastique élémentaire de l'explication du présent par le passé,

pour prévoir, gérer (un stock), contrôler (un processus chimique), expliquer, détecter (une rupture de comportement), ou vérifier qu'il est impossible de prévoir quoi que ce soit avec la technique employée…

Nombreux sont les domaines scientifiques concernés.

1.2 **Logiciels** (seuls quelques-uns sont cités selon un choix arbitraire)

a) **logiciels spécialisés**

Moujik, (1980), de la Société Processus, sur P.C., français. Résidence Simon, 8 D rue de la Ceinture, 7800 Versailles. Un des plus anciens du monde, avec interface graphique sur tektro, puis PC (mis à jour? Existe encore? venu trop tôt en France, où les enseignants sont souvent plus soucieux de théorie que de mise en œuvre).

Mandrake, (1990), de l'Université Paris Sud, sur P.C., français.
Laboratoire U.A. 743 de Statistiques Appliquées, Bât. 425. Facile d'emploi. Assez pédagogique et assez automatisé.

TSE (1991), Système expert pour la prévision économique, Pr. Mélard, Université libre de Bruxelles. Institut de Statistique. Campus de la Plaine. C.P.210, Bd du Triomphe, B-1050, Bruxelles. Fax +32-2-6505890. Prévision non automatisée et fine + livre: Méthodes de prévision à court terme, Mélard, ISBN2- 7298-9054-8 Ed. Ellipses. Sur PC.

Timeslab (1988). de Newton H.J., de Texas A&M University, Wadsworth & Brooks/ Cole Publishing Company, Advanced Books & Software. Pacific Grove, California, 93950. ISBN 0-534-09198-9. Assez convivial en uni - bi-varié. Dans le book, se trouve la disquette P.C.

RATS http://www.intesoft.com/ info@intesoft.com Des ARIMA aux GARCH.

3 SE, de la Société SSP Conseil, sur P.C. 72 rue du château de Rentiers, 75013, Paris.

SCA system (bien développé pour les chroniques multiples), **ITSM** (1990), **TSA** de la NAG (1981), **X11**,
TSP, etc. ...

b) **logiciels de statistiques**

Pour traiter des séries, ces logiciels sont le plus souvent moins conviviaux que les précédents, puisqu'ils n'ont pas été faits pour cela uniquement. Ils contiennent cependant des modules ou des fonctions plus ou moins développés de traitement de séries.

Logiciel **A-TEMP** de **SPAD**, CISIA, 1, av. Herbillon, 94160, St-Mandé, Fax (1) 43 74 17 29. graphiques, prévision uni et multivariée, français sur PC.

Logiciel **TIMElab** de **STATLAB**, slp STATISTIQUES 51/59 rue L. Rollin, 94853 Ivry sur Seine, Cedex. Fax (1) 45 15 26 77. Univarié. français sur Unix, Windows, OS/2, Mac.

Le logiciel **Splus**, que nous utiliserons en T.P., contient des fonctionnalités sophistiquées permettant des traitements multi-séries, une gestion de calendrier, des lissages spécifiques aux séries avec de bonnes sorties graphiques. Langage de développement de méthodes autant que logiciel de commandes, il est intéressant dans une formation ingénieur. Deux adresses, avec des services rendus différents:
Sigma-Plus 5 place St Pierre 31000 Toulouse. Fax 61 23 39 82.
Ritme 34 bd Haussmann 75009 Paris. Fax (1) 42 46 00 33
Introduction à Splus, notice en français par Marcel Baumgartner, baumgart@dma.epfl.ch ou helblin@dma.epfl.ch

SAS contient un module série plutôt complet sur gros système, parfois plus cher à l'achat qu'un logiciel spécialisé (pour compenser, disons que JMP est une version conviviale réduite intéressante de SAS sur Mac, faisant penser à Data Desk sur certains points).
SPSS est bien développé sur les chroniques (pc, mac, etc.). Normal pour ce logiciel orienté Sciences Sociales.
BMDP, un produit ancien qui s'est bien modernisé.
Systat-Sygraph ou encore **Minitab** permettent de faire une bonne partie ce qui est indiqué dans ce cours (sur PC ou Mac).
EDA, permet de jouer facilement avec tous les lissages à la Tukey (sur Mac ou Pc-windows, excellent rapport qualité-prix si on s'intéresse au point de vue exploratoire, Prof. Horber, Dpt. Science Politique, Uni. Genève, 102 Bd Carl Vogt, 1211 Genève 4). Mais l'interface graphique manque.
MODE Etc.…

c) **logiciels de calculs numériques**

Avant que les statisticiens ne s'intéressent aux séries chronologiques, les physiciens avaient élaboré de nombreux outils, notamment dans le cadre du traitement du signal. On trouve dans ces logiciels des commandes relatives aux fonctions de transfert, armax, fft (fast fourier transform), etc., qui peuvent servir au traitement des chroniques. Les aspects probabilistes et statistiques y sont cependant peu ou pas développés.
Citons **Matlab**, de MathWorks, avec ses boîtes à outils spécialisées dont: **Control System**, **System identification**, **Signal Processing**, **Spline, Statistics, neural networks**. S.C.G. 72-78 grand rue. 92310 Sèvres. Fax 45340109. La NAG propose un interfaçage poussé avec Matlab. Il existe une

boîte pour calcul formel dont le noyau est celui de Maple. Génération de code C depuis matlab. (Mac, PcWindows, Crays, etc.).
Mais aussi **Mathematica** (Mac, PcWindows, etc.)**, Matrix$_x$-Xmath** (Pc, Stations), etc.

La question du choix d'un logiciel est d'autant plus complexe que souvent nous sommes en face de "marchés captifs": En sciences sociales, on travaille souvent avec SPSS. Une grosse boîte "se doit" d'avoir le logiciel le plus cher (SAS) et ne peut imaginer que des logiciels gratuits puissent être plus performants de certains points de vue (tel Xlisp.stat, langage de développement très intéressant par exemple pour concevoir graphiques et fenêtres liés ou R, version réduite de Splus). On pourrait dire qu'une bonne formation en statistique suppose une connaissance de plusieurs logiciels, ne serait-ce parce que certaines façons de traiter les données rendent obsolètes certaines méthodes.

Une revue publie parfois une fiche synoptique de logiciels de statistique (80 fiches dans le n°2 de février 1993): **Computational Statistics and data analysis**. North Holland.

Journal of Statistical Software est un récent journal consacré à la publication et à la description de programmes utiles pour les statisticiens.

Forum sur réseau fr pour tous statisticiens francophones et tous thèmes:
Pour s'inscrire, envoyer un message à **servmast@cict.fr** avec le sujet: oui (pour annuler l'inscription, idem avec sujet: non); Envoyer ses messages à **lmastat@cict.fr**.

2 - ÉTUDE PRÉLIMINAIRE D'UNE SÉRIE

2.1 Les données

Elles sont constituées de n observations notées x_i, i = 1, …, n ou x_t, t = 1, …, n.
n est la **taille** de l'échantillon, de la série observée.

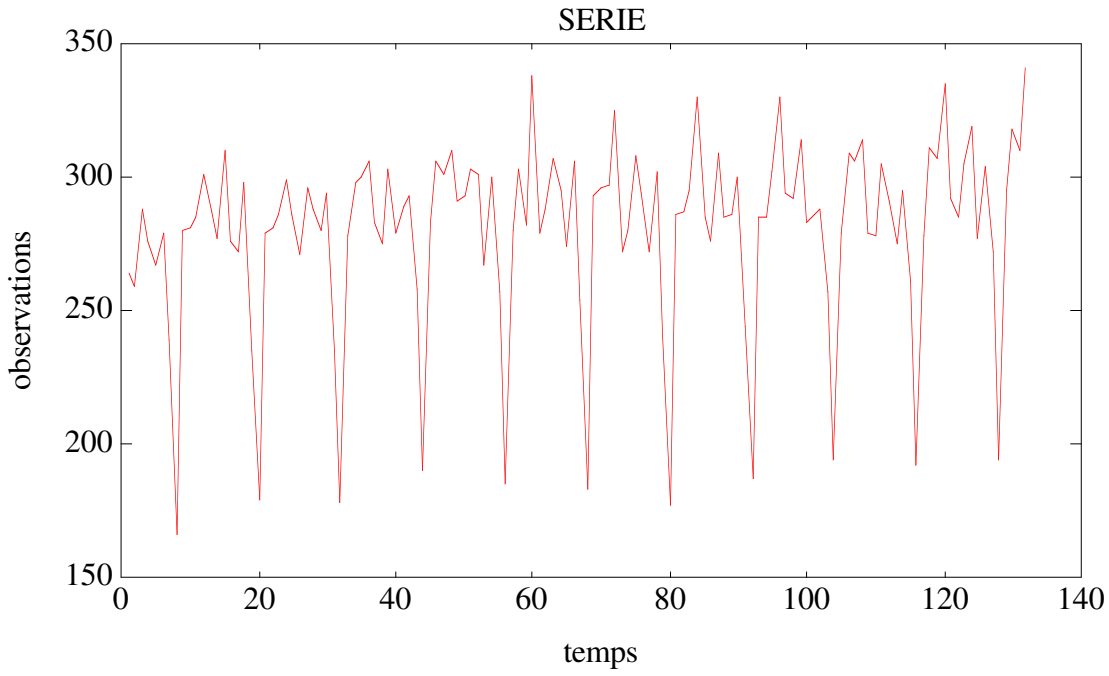
Les observations sont faites à intervalles de temps **régulièrement espacés**.
L'intervalle de temps égal à 1, entre t et t+1 ∀t, représente un jour, ou une seconde, ou autre.
Cet échantillon ne pourra pas être vu comme issu d'un n-uplet de variables aléatoires indépendantes, la succession dans le temps des valeurs induisant le plus souvent une dépendance.

Exemples: Relevés mensuels d'un compte bancaire. Longueurs journalières de la tige d'une plante. Températures relevées toutes les minutes d'un processus chimique en cours. Relevés pluviométriques trimestriels. Évaluations d'heure en heure du stock pour chaque article d'un hypermarché.

2.2 Observations de graphiques

« Ne pas ajouter à la démence du réel, la niaiserie d'une explication », dixit J. Rostand. Il avait passé sa vie à observer les grenouilles et, de fait, est devenu célèbre pour avoir su regarder et décrire ce qu'il voyait.
On n'hésitera pas, avant et pendant l'analyse, à user de graphiques.

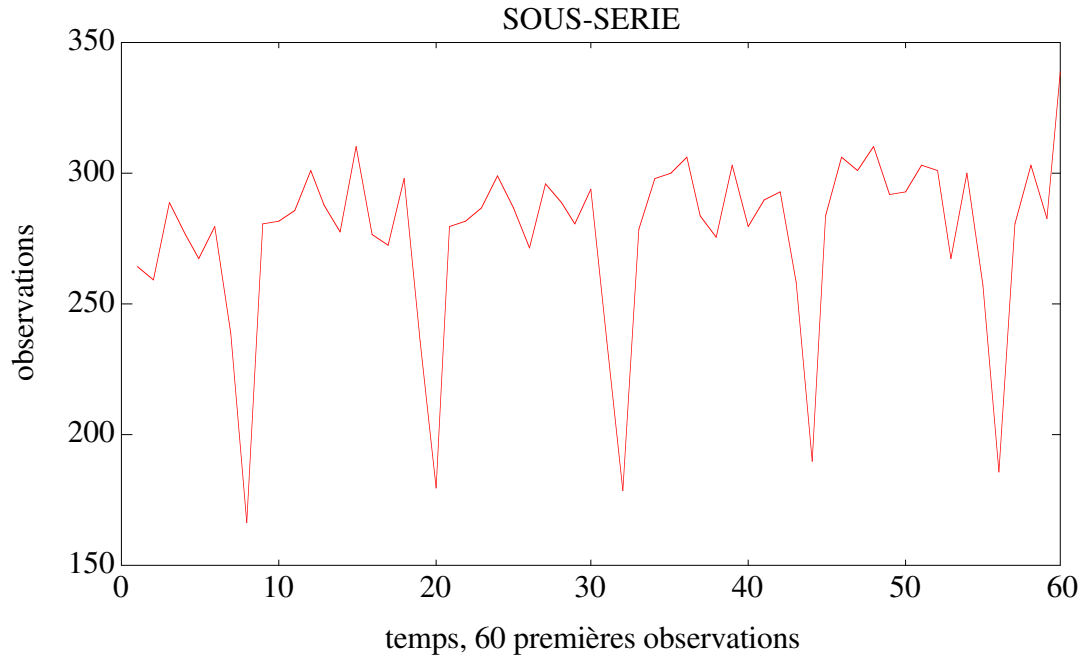
Par l'observation, on pourra parfois s'apercevoir de l'existence de valeurs manquantes, de valeurs accidentelles, d'un mouvement tendanciel, d'un mouvement périodique, de ruptures de comportement, d'une analogie de comportement de plusieurs séries représentées simultanément.

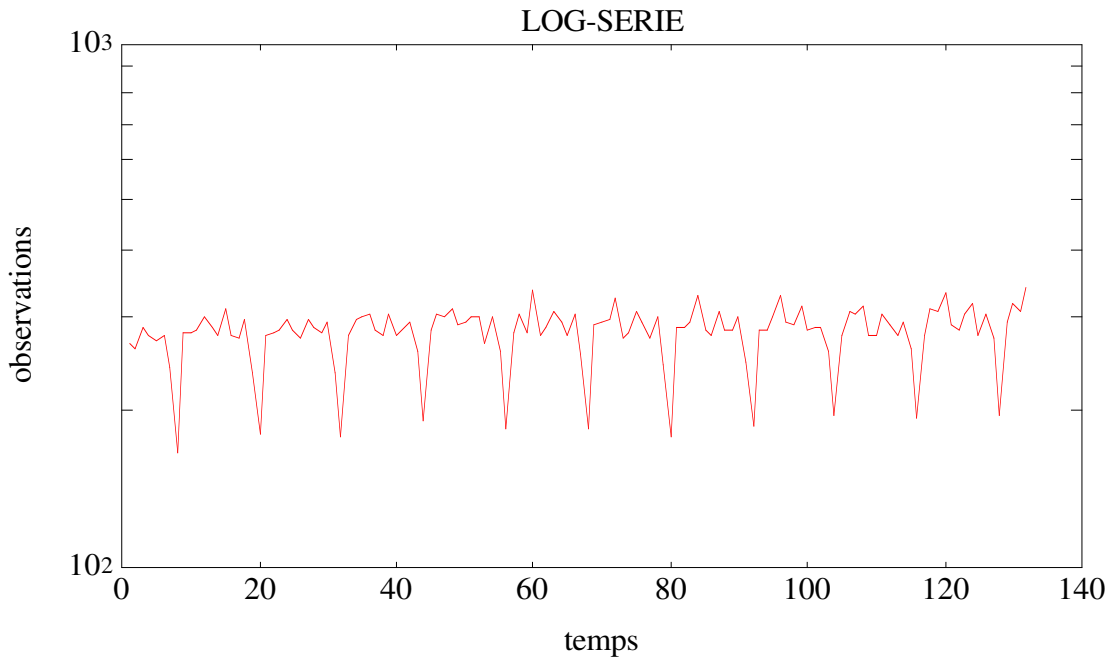


Quand l'outil est convivial, on observe aussi bien les données initiales que des données transformées: par fonction puissance

$$[(x_t)^{\lambda} + a] / b], a \in \mathbf{R}, b \in \mathbf{R}, \lambda \in \mathbf{R} - \{0\},$$

ou par logarithme. La classe de transformations, dite de **BOX-COX**, usitée en économie, correspond à cette famille, avec a=-1 et b=λ (on convient que le Log correspond à $\lambda = 0$, car un tracé de toutes les courbes lui donnerait cette position limite).

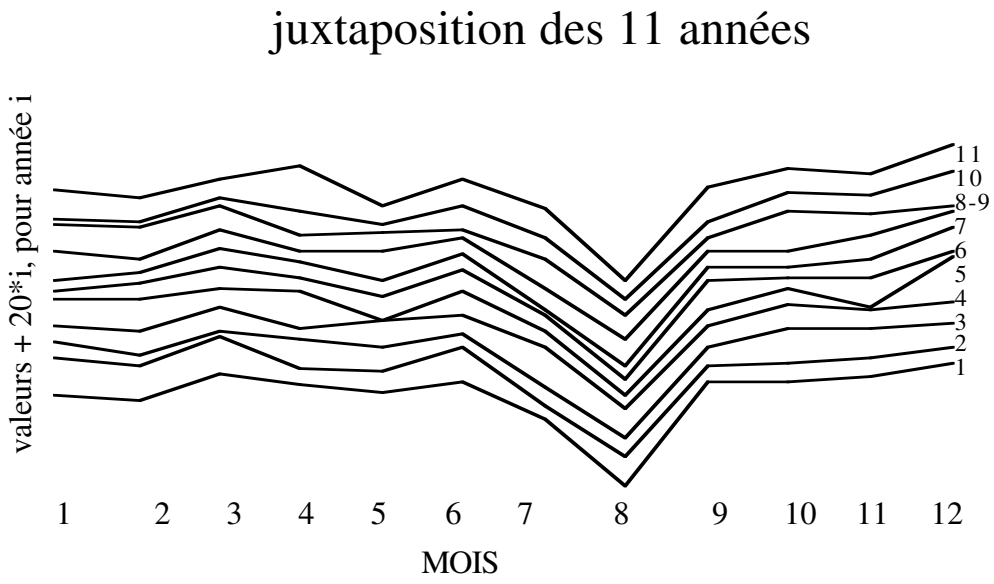




Tous ces graphiques ont un point commun: le repère est orthogonal et le temps est porté en abscisse.

On peut choisir le logarithme des valeurs pour atténuer une croissance exponentielle, ou bien remplacer les valeurs par les taux $[(x_{t+1} - x_t) / x_t]$ jugés a priori plus intéressants que les valeurs initiales.

L'usage de la couleur n'est pas inutile, car souvent nous ferons des représentations simultanées de différentes séries, même quand une seule série est à l'étude, notamment pour comparer différents lissages.



2.3 Corrections des données

Il s'agit de modifier les données brutes avant de se lancer dans une étude préliminaire, ou à la suite de cette étude.

Citons quelques situations fréquemment rencontrées:

* Évaluation préliminaire de données manquantes; remplacement de données accidentelles.

* Découpage en sous-séries, associé aux ruptures de comportement.

* Extraction de sous-séries de valeurs appartenant à un intervalle donné (pour une étude robuste par exemple), ou de valeurs extrêmes, ou des mois d'août qui ont un comportement original à dissocier de celui des autres mois, ou, pour un commerce, les mois d'été à dissocier des autres mois (et donc 2 séries à traiter)….

* Discrétisation de données recueillies continûment (relevés de température), ou de données lissées continûment (parce que les données initiales n'auraient pas été régulièrement espacées dans le temps).

* Standardisation, calcul de moyennes, pour se ramener à un intervalle de temps d'amplitude fixe entre les relevés successifs (mois standard de [365,25/12] jours en consommation, mois standard prenant en compte les jours fériés en production, hauteurs de pluie observée chaque mois / nombre de jours du mois).

* Transformations des données, pour des raisons diverses (passage des coordonnées cartésiennes à des coordonnées polaires ou réciproquement, modélisation plus aisée, etc.).

3 UN MODÈLE DÉTERMINISTE: LE MODÈLE ADDITIF

3.1 Présentation du modèle

Le modèle est celui d'une décomposition additive :

$$x_t = \mathbf{z_t} + \mathbf{s_t} + e_t \ , \quad \forall\ t$$

c'est à dire le modèle proprement dit: **Tendance + Saisonnier,** les résidus e_t correspondant à la notion d'écart au modèle, de fluctuations autour du modèle, ou encore de "bruit" quand les valeurs du modèle sont assimilées à celles d'un "signal" (langage de physiciens).
En anglais les résidus de modèles pour chroniques sont traduits par irregular, error, residuals, remainder or noise.

Le saisonnier est un phénomème qui se reproduit analogue à lui-même sur chaque intervalle de temps successif, l'étendue de cet intervalle qui est constante étant appelée **période**. Pour des données mensuelles un saisonnier de période 12 peut exister, pour des données trimestrielles un saisonnier de période 4.

Le modèle le plus simple suppose que le saisonnier, de période T, est constant (rigoureusement égal à lui même d'année en année, ce qui ne signifie pas que la variance du saisonnier soit nulle):

$$s_{t+T} = s_t \ , \quad \forall\ t.$$

Il est en général défini par T coefficients saisonniers $c_1, c_2, \ldots, c_T$, avec la contrainte suivante:

$$\Sigma_{i=1\ldots T}\ \ c_i = 0.$$

Le saisonnier se caractérise alors par des valeurs positives et négatives, pour créer une moyenne nulle. Cette contrainte est aussi voulue (et à peu près vérifiée) quand le saisonnier n'est pas constant, c'est à dire quand les coefficients saisonniers se modifient un peu au fur et à mesure que sont considérées les périodes d'amplitude T successives.
De toutes façons, il y a une infinité de solutions au problème de la décomposition, correspondant d'abord à l'infinité des choix possibles de la tendance.

$x_t - z_t$ est la série **détendancialisée**. Elle est égale à saisonnier + résidu.
$x_t - s_t$ est la série **désaisonnalisée**. Elle est égale à tendance + résidu. On la nomme aussi **série corrigée des variations saisonnières**.

On estime le saisonnier sur une série détendancialisée. Par exemple on fait alors un calcul des coefficients par moyennes: par exemple pour des données mensuelles, si la période 12 existe, le coefficient du mois de janvier sera la moyenne de tous les mois de janvier de la série $x_t - z_t$.
Alors, si n=kT, et si le centre de gravité du nuage des points (t, x_t) coïncide avec celui des points du mouvement tendanciel (t, z_t), et quand le phénomène saisonnier est nul en moyenne, alors on a aussi:

$$\Sigma_t\ \ e_t = 0.$$

Mais rien n'empêche de préférer la médiane plutôt que la moyenne pour le calcul des coefficients saisonniers, de penser à une modélisation de chaque mois plus fine que par une simple constante (par exemple modélisation par lissage). L'égalité précédente n'est ainsi en général qu'à peu près vérifiée.

Exercice:

On suppose qu'une série x possède un mouvement saisonnier trimestriel constant, les 4 coefficients saisonniers valant: $c_1 = -1$, $c_2 = 2$, $c_3 = 0$, $c_4 = -1$.

On suppose que le mouvement tendanciel est une droite de pente 1, passant par l'origine du repère des points (t, x_t).

On suppose en outre que le mouvement résiduel e_t est le suivant:

année \ trimestre:	1	2	3	4
1		1	- 1	1
2	1	- 1	1	- 1
3	- 1	- 1	1	

Représentez dans un même repère orthonormé z_t, puis $(z_t + s_t)$ à partir du tracé de z_t, et enfin x_t, pour les dix valeurs de t = 1, 2, …, 10.

Remarques:

On pourrait préférer le **modèle multiplicatif**

$$x_t = z_t \, s_t \, e_t$$

mais alors, la relation

$$\log(x_t) = \log(z_t) + \log(s_t) + \log(e_t),$$

montre que cela revient à considérer un modèle additif (s'il existait des valeurs négatives, on prendrait par exemple les valeurs: $x_t - \min_t(x_t) + 1$, au lieu des valeurs x_t).

C'est dire que le modèle additif contient le multiplicatif, et bien d'autres, quand on s'autorise la possibilité de transformer les données.

Dans cet ordre d'idées, citons le modèle dit "**logistique**":

$$\mathrm{Log}[(1 - x_t) / x_t] = z_t + s_t + e_t \,, \quad \forall\, t.$$

Cependant, le modèle **mixte** (mi-multiplicatif, mi-additif):

$$x_t = z_t \, s_t \; + \; e_t$$

ne peut être associé au modèle additif via une transformation.

Par ailleurs, une transformation préalable des données est parfois introduite pour rendre plus acceptable l'hypothèse d'un mouvement saisonnier constant. Ainsi, si on remarque sur un graphique une croissance

de la variabilité du phénomène saisonnier d'années en années, au lieu de la constance souhaitée, cette croissance peut être fortement amoindrie par une transformation, telle que le logarithme.
Un objectif parfois recherché par une transformation est de linéariser le mouvement tendanciel. Plus lisse que la droite, tu meurs (en géométrie euclidienne). Comme nous le verrons, la droite de Buys Ballot, ou le filtrage par moyenne mobile, peuvent être les véritables motivations qui sous-tendent l'intérêt que l'on peut porter à une tendance linéaire (plutôt qu'un lissage super lisse).
Un autre objectif visé par une transformation est parfois de rendre par là les deux mouvements, tendanciel et saisonnier, les plus indépendants possibles, ce qui se traduit en pratique par la recherche d'estimateurs des paramètres correspondants les moins corrélés possibles.
Un quatrième serait de se débarrasser du mouvement tendanciel, en le supprimant à l'aide d'une ou plusieurs différenciations (confer § 6.2, pour se ramener au cas stationnaire).

Le nombre de paramètres à estimer pour un modèle additif simple est T + nombre de paramètres éventuellement nécessaires à l'expression de la tendance. Quand la période est 12, cela fait déjà beaucoup de paramètres, et les 12 estimateurs du saisonnier sont rarement indépendants entre eux. La technique du lissage pour modéliser la tendance et le saisonnier, permet d'éviter cette complexité.

3.2 Mouvement tendanciel et lissages

Il y a une infinité de lissages possibles pouvant servir chacun à exprimer le mouvement tendanciel: une succession de points (t, z_t) situés au milieu du nuage de points (t, x_t).

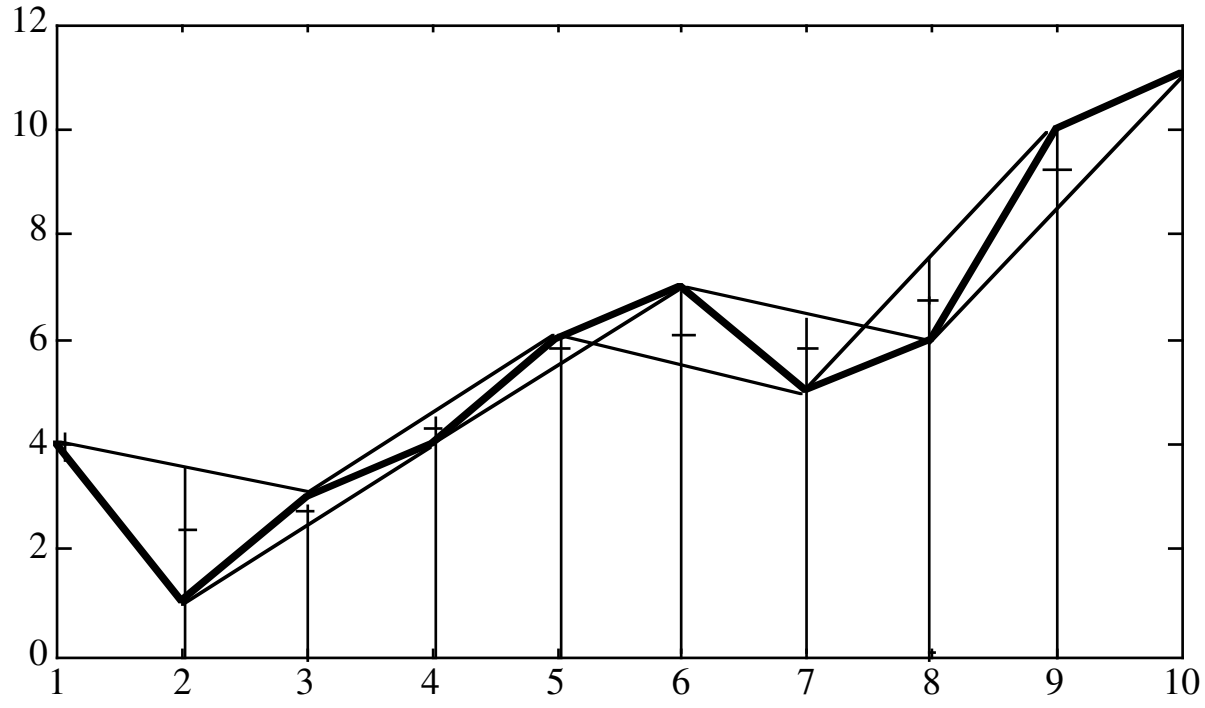
Remarquons ici que le problème ainsi posé peut être lié au type de représentation. Pour nous ce sera dans un repère orthonormé, où le temps est porté en abscisse. Pourtant il existe d'autres représentations intéressantes pour certaines chroniques. Par exemple, les représentations en spirale, de séries ayant une tendance constamment croissante, et de période T, T axes partageant le plan en T secteurs égaux et servant à la représentation de chacun des mois. On pourrait alors travailler en coordonnées polaires plutôt que cartésiennes.
Tous les logiciels sont encore pauvres quant à la variété des graphiques offerts, mêmes ceux spécialisés pour les séries.
Si on désire exprimer la tendance avec une courbe très lisse (c'est doux, c'est joli, est-ce bête aussi?), une méthode consiste à proposer une expression analytique (pour la prolonger aisément?). Notamment pour des raisons de résistance, on calcule souvent cette expression à partir de données lissées jugées convenables (visuellement ou d'après un critère à optimiser), et non à partir des données initiales.
Souvent il s'agira d'estimer par les moindres carrés les paramètres d'une expression polynomiale (par exemple, 2 paramètres pour une droite). Ces moindres carrés sont ceux des distances calculées entre points de la courbe lissée et points de la courbe analytique, pour chaque valeurs du temps, parallèlement à l'axe des ordonnées (ce qui est tout à fait justifié, le temps ne devant pas intervenir).

Rappelons que les logiciels (sauf ceux de calcul formel) fournissent les valeurs calculées de la fonction analytique pour les abscisses temps considérées, et non l'expression de la fonction elle-même, et cela pour des raisons de rapidité de mis en œuvre.

Cela implique qu'une représentation n'apparaîtra avec des courbures arrondies que si les points successifs sont suffisamment rapprochés.
Cela explique pourquoi il est rarement prévu dans ce calcul de pouvoir prolonger une courbe dans le futur, alors que cette visualisation du futur est souvent désirée.
Dans tous les cas on recherche donc une "bonne" série lissée. Il est difficile de dire exactement ce que l'on entend par là, les motivations pouvant être très variées. Cependant il ne s'agit pas le plus souvent d'atteindre un cas limite: la droite, ou de s'ajuster au point de passer par chaque point observé.

Il s'agit souvent de lisser le moins fort possible tout en atteignant un objectif précis: par exemple se débarrasser des résidus, ou des résidus et du saisonnier, de sorte que ces résidus, ou ces résidus plus saisonnier, semblent fluctuer autour de la courbe lisse. Cette courbe lisse doit alors apparaître en position centrale par rapport à cette composante résiduelle, ou aux résidus plus saisonnier.
On contrôle un lissage en superposant courbes lisses et données à lisser.
On souhaite en général que la courbe lisse passe au milieu du nuage de points, qu'il y ait à peu près autant de points au dessus qu'en dessous de la courbe lisse quelle que soit la zone où on l'observe. Il est alors utile pour un bon contrôle visuel d'effectuer la représentation par points des données sans les relier entre eux, avec la représentation des valeurs lissées reliées entre elles.
Il n'existe pas de technique de lissage automatisée qui donnerait satisfaction quelques soient les données.
On est donc le plus souvent amené à faire usage de différentes techniques, et de différents paramétrages d'une même technique, en procédant plus ou moins à tâtons. Pour certaines séries bien typées, on saura d'avance si l'emploi de certaines techniques sera avantageux ou non, connaissant leurs propriétés.

3.2.1 "High-low-mid points" ou Hanning



C'est avec le "3R" le lissage le plus facile à réaliser à la main sur un graphique. Convient tout particulièrement aux séries de très faible taille, tout comme le 3R.
La série précédente est représentée ici en gras. Les points marqués au milieu des segments verticaux sont les points lissés, non reliés successivement entre eux ici. On note l'absence de valeurs lissées aux temps extrêmes 1 et 10.

3.2.2 Moyennes mobiles d'ordre T

Pour t = (T+1)/2, (T+1)/2 + 1, …, n + 1 - (T+1)/2, on calcule les valeurs lissées:

$$l_t = (\Sigma\, x_i) / T,$$

l'indice i allant de t - (T-1)/2 à t + (T-1)/2. Il s'agit donc d'un calcul de moyenne fait dans une **fenêtre mobile** de largeur T, contenant toujours T points car on la déplace chaque fois parallèlement à l'axe du temps d'une unité, de sorte que le point le plus à gauche sort de la fenêtre et qu'un nouveau point entre à droite dans la fenêtre. La technique précédente travaillait avec une fenêtre mobile contenant 3 points.

$a_1 = \Sigma_{1\ldots T}\ x_i,\ l_1 = a_1 / T,\ a_2 = a_1 - x_1 + x_{T+1},\ l_2 = a_2 / T,\ a_3 = a_2 - x_2 + x_{T+2}, \ldots$

Si T est impair, les valeurs lissées sont fournies pour les valeurs de t entières en nombre n-(T-1), T-1 temps extrêmes ne possédant pas de valeurs lissées.

Si T est pair, les valeurs lissées correspondraient à des temps moyens non entiers. C'est pourquoi ces valeurs sont elles-mêmes relissées à l'ordre 2. Ce lissage final donne alors des valeurs associables à (n-T) valeurs entières du temps. On dit qu'il s'agit d'un lissage à l'ordre T, **centré**.
Le lissage par moyenne mobile est intéressant dans le contexte du modèle additif dans la mesure où il est compatible avec ce modèle: pour un ordre donné, la série lissée est la somme des lissés de chacune des composantes du modèle:

$$l_t(x) = l_t(z) + l_t(s) + l_t(e)\ ,\quad \forall\ t$$

Quand l'ordre est choisi égal à la période T, comme la somme des coefficients saisonniers d'un modèle additif simple est nulle, on a: $l_t(s) = 0$. Le lissage d'ordre T est alors un **filtre linéaire** qui élimine le saisonnier.
Les résidus sont en moyenne nuls. Les moyennes de T valeurs successives de ces résidus sont en général assez proches de zéro. Quoiqu'il en soit, on aurait:

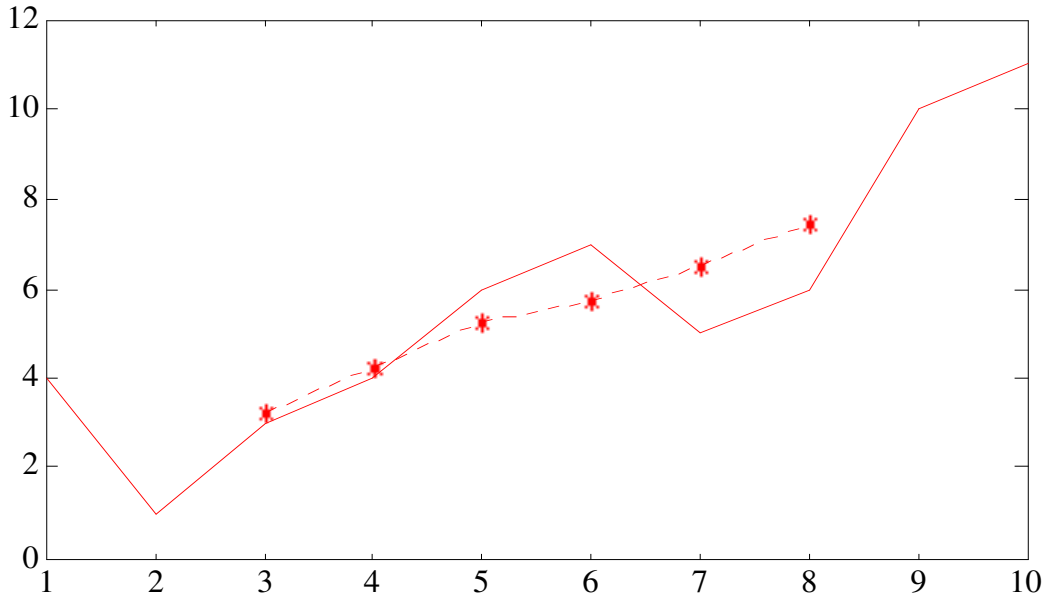
$$\mathrm{var}[l_t(e)] = \mathrm{var}[e_t] / T,$$

si les e_t étaient indépendants entre eux . C'est faux le plus souvent, mais quand même la variabilité des résidus (souvent déjà faible comparée à celle du saisonnier et de la tendance) diminue le plus souvent assez fortement par lissage. C'est ainsi qu'alors:

$$l_t(x) \cong l_t(z),$$

Pour obtenir, par ce lissage, le mouvement tendanciel, il faudrait: $l_t(z) = z_t$. Cela est vérifié *seulement si la tendance est linéaire*. C'est donc, un lissage qui peut être préconisé dans le contexte du modèle additif et avec l'ordre égal à la période, quand la tendance est sensiblement linéaire.

La série ci-après est celle de l'exercice précédent, de période 4; le lissage d'ordre 4 a été centré.



Les lissages suivis de relissage, à un ordre différent ou non, reviennent à des lissages pondérés, pour lesquels les poids des points au centre de la fenêtre mobile sont plus importants que ceux aux bords. L'ordre dans lequel s'effectuent les lissages successifs est indifférent.
Ainsi, le lissage à l'ordre 12 suivi d'un lissage à l'ordre 2 est un lissage par moyenne mobile pondéré à l'ordre 13 désigné par les 13 poids utilisés, et on peut noter:
[12 fois 1/12][2 fois 1/2] = [2 fois 1/2][12 fois 1/12] = [1/24, 11 fois 1/12, 1/24].

Hanning: [2 fois 1/2][2 fois 1/2] = [1/4, 1/2, 1/4]. Comparer avec le High-low-mid points.

Très utilisé, paraît-il, le **I**anning: [1/8, 1/4, 1/4, 1/4, 1/8].
Le **J**anning étant: [1/8,1/8,1/8,1/4,1/8,1/8,1/8], le **K**anning....

Spencer: [4 fois 1/4][4 fois 1/4][-3/4, 3/4, 1, 3/4, -3/4], qui préserve les polynômes de degré inférieur ou égal à 3 (la moyenne mobile équipondérée préserve seulement la droite), ou bien qui préserve les polynômes locaux, estimés dans la fenêtre (confer Gouriéroux-Monfort).

Le calcul de valeurs lissées aux **temps extrêmes** 1, n, …, peut s'effectuer par moyenne pondérée unilatéralement. Par exemple, après un lissage à l'ordre 3, on peut proposer: $l_1 = (3x_1 + 2x_2 + x_3) / 6$ et $l_n = (3x_n + 2x_{n-1} + x_{n-2}) / 6$.
Une autre approche revient à user de techniques de prévision élémentaires sur les valeurs lissées déjà calculées, en prévoyant (ou "postvoyant") suffisamment loin pour proposer un lissage final complet du temps 1 au temps n.
On peut évidemment user des techniques de reconstitution introduites en section 3.3.3.

3.3.3 Médianes mobiles à l'ordre k

Ce n'est pas un filtre linéaire pouvant éliminer le saisonnier du modèle additif en prenant un ordre égal à la période.

C'est une **technique résistante** (peu sensible aux valeurs accidentelles, alors que la technique précédente y est sensible, une seule valeur influant sur le calcul de k valeurs lissées à l'ordre k; soulignons qu'il peut être *bien gênant* avec les moyennes mobiles d'avoir un outil de calcul de milieu de nuage de points plus influencé par les points les plus éloignés de ce milieu que par les autres).

S'il est nécessaire de recentrer dans le temps un lissage par **médiane** d'**ordre pair**, on use de la *moyenne* mobile d'ordre pair.

Les différentes combinaisons de lissages définis par les codes ci-après permettent une grande variété de tentatives. Le logiciel EDA possède une commande spécifique très souple pour ce faire.

code k
Notons L et L' les lissages par médiane mobile d'ordres respectifs k et k'.
Un lissage L suivi d'un relissage L' est noté L'oL et est codé kk'.

code twice
Il signifie qu'à une série lissée par médiane mobile, on rajoute les résidus relissés au même ordre (le plus souvent), ce qui redonne un peu de rugosité au lissage initial jugé trop lisse, ne collant pas avec assez de souplesse aux données.
Notons (I - L)(x) la série des résidus d'un premier lissage.
$L_o(I - L)(x) = (L - L_oL)(x)$ est la série des résidus, relissée par L.

Par exemple, en notant L l'opérateur de lissage par médiane mobile à l'ordre 5, le code "5,twice" désigne le lissage:

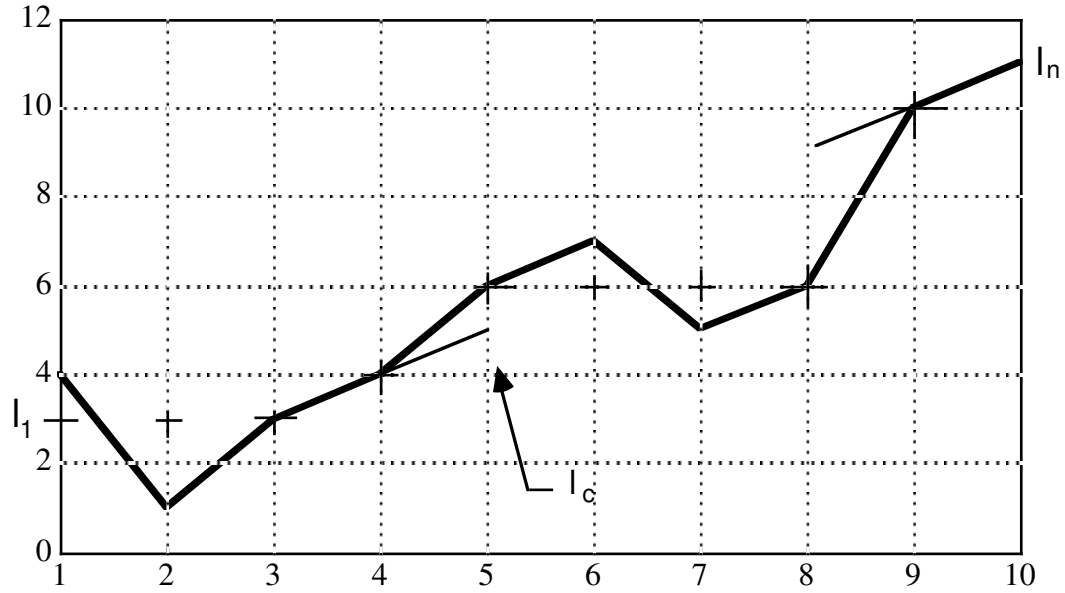
$$x \rightarrow L(x) + (L - L_oL)(x) = 2L(x) - L_oL(x)$$

code R (reroughing)
Il signifie qu'on relisse sans cesse le lissage à l'ordre indiqué par le code qui précède la lettre R. La procédure d'arrêt se fait à l'étape précédant celle pour laquelle aucun changement ne s'est produit.

Le lissage **3R** est très utilisé pour les séries de taille très faible.

Ci-après, un lissage 3R (points codés par "+", pour t = 2 à 9). Que donnerait le relissage de ces points par médiane mobile d'ordre 3 ?

code S (split)
Il donne une forme arrondie aux plateaux (remarquez le plat des valeurs lissées entre les temps 5 et 8), notamment produits lors d'un lissage 3R. Ainsi, pour un plateau ayant un bord gauche au temps 5, on substitue à la valeur lissée l_5 la valeur: $l_c = \text{med}(l_5, l_4, 2l_4 - l_3)$.



code E (extremum)

Il permet de proposer des valeurs lissées pour tous les temps (de la même façon que pour les moyennes mobiles, plus l'ordre est grand, plus est élevé le nombre de valeurs lissées manquantes aux temps extrêmes).

Ainsi, si l_1, l_2 et l_3 sont manquantes, ce code produit les valeurs:
$l_3 = \text{med}(x_1, x_2, x_3, x_4, x_5)$, $l_2 = \text{med}(x_1, x_2, x_3)$ et $l_1 = \text{med}(2l_2 - l_3, x_1, l_2)$, (ou $l_1 = x_1$). On procède de façon symétrique pour évaluer l_{n-2}, l_{n-1}, l_n.
Sur le graphe précédent, on a pris $l_n = x_n$ et $l_1 = \text{med}(2l_2 - l_3, x_1, l_2)$.

code H

Si les données ne sont pas centrées par les lissages successifs précédents, ce code rajoute un lissage par moyenne mobile d'ordre 2. Sinon, 2 lissages successifs par moyennes mobiles d'ordre 2 sont effectués, ce qui revient à un seul lissage à l'ordre 3 par moyennes mobiles pondérées avec les poids 1/4, 2/4, 1/4. C'est le Hanning, qui arrondit les angles.

Dans EDA, d'autres lissages finaux par moyennes mobiles pondérées sont relatifs à des fenêtres temporelles d'étendues de plus en plus grandes, et sont codés **I, J et K**.
I comme Ianning défini par [3 fois 1/3][3 fois 1/3], J comme Janning avec 4, K comme Kanning...
Faire des lissages par moyennes, une fois éliminées par médianes mobiles l'essentiel des valeurs exceptionnelles, est cette fois concevable.

Pour conclure sur ces codages, dus à Tukey à l'origine, les praticiens de cette technique modulable au gré de l'utilisateur, préconisent cependant les lissages codés suivants, **avec twice**:
4253HE ou **3RSSHE** (Tukey, ou Velleman et Hoaglin)
5RIE ou **7RJE** (Ladiray et Roth)
et **3RSE** surtout pour les petites tailles.
Ces codages sont de bons filtres passe-bas, laissant à peu près intact ce qui est du domaine des basses fréquences et éliminant totalement ce qui est de celui des hautes fréquences. Des simulations suivies

d'étude spectrale ont permis de vérifier que l'efficacité de **5RIE** est comparable à celle de **4253HE**, alors que **7RJE** filtre plus sévèrement, donnant un lissage plus intense, alors que **3RSSHE** agit dans l'autre sens, laissant passer quelques basses fréquences de plus.
Ces codages de référence sont ceux à partir desquels d'autres éventualités de codages sont à envisager.
Une chronique **monotone** ($\forall$ t, $x_t \geq x_{t-1}$ ou $\forall$ t, $x_t \leq x_{t-1}$) est invariante par lissage par médiane mobile. Un mouvement saisonnier pur de période 2 l'est aussi avec les ordres 4k+1 ($k \geq 1$). Certains autres saisonniers, de périodes différentes, le sont pour un ordre donné…

Parce qu'un logiciel d'analyse exploratoire s'attache d'abord aux techniques résistantes, ces techniques font souvent appel à la notion de rang que possède chacune des valeurs par rapport aux autres. Le calcul de la médiane s'y effectue de façon précise (et rapide), ce qui n'est pas toujours le cas de tous les logiciels de statistique, même très réputés.
Il semble bien que cette technique de lissage (sans l'option S) soit supérieure à beaucoup d'autres dans le cas d'une série possédant des chutes et remontées verticales que l'on veut préserver, où seulement les petites oscillations relatives à des plateaux doivent être lissées.

3.3.4 Buys-Ballot

On veut estimer **simultanément** les T+2 coefficients, du saisonnier et de la droite exprimant le mouvement tendanciel, de sorte que le modèle additif corresponde aux résidus les plus faibles (au sens des moindres carrés).

Soit T le nombre de "mois" et k le nombre d'"années", avec n=kT. Si n≠kT, on supprime les premières valeurs de la série, pour avoir un multiple de la période.
Soient g_i, i=1…k, les centres de gravité des k sous-séries relatives aux k années, d'abscisses les moyennes des abscisses, d'ordonnées les moyennes des ordonnées.
La solution, appelée droite de Buys-Ballot, est la droite qui s'ajuste au sens des moindres carrés à ces k points. Elle passe par le centre de gravité du nuage de tous les points.

3.3.5 Lowess

(locally weighted regression scatterplot smoothing)
Prévu à l'origine pour les nuages de points dans $\Re^2$ de deux variables statistiques x et y mesurées sur de mêmes individus (scatterplot). Donc intéressant aussi pour lisser des données non régulièrement espacées dans le temps et pour lisser le saisonnier. La bonne gestion du calcul des valeurs lissées aux temps extrêmes pourra donc souvent constituer une base honnête pour extrapoler vers le futur un mouvement tendanciel ou saisonnier.

On veut décrire par un lissage la liaison entre x et y (du type l_t=f(t) comme lors des lissages précédents) à partir d'un nuage de points observés (x_i, y_i).
Dans une fenêtre mobile d'ordre k, on calcule un point lissé, en le prenant sur la droite de régression linéaire (les moindres carrés sont pondérés) calculée avec les k points de la fenêtre, son abscisse étant au milieu de la fenêtre en général. Parfois il existe des options à la commande où le polynôme de régression

de degré fixé à 1 peut être remplacé par le degré fixé à 2 ou le degré 0 (alors moyenne pondérée). Le lissage est moins lisse avec un degré moins élevé.
Le calcul s'effectue dans chaque fenêtre en affectant des poids différents aux points dans la fenêtre: élevés pour les points ayant une abscisse proche de celle du point à lisser, donc proche du milieu de la fenêtre, faibles pour ceux éloignés.
On considère ainsi la fenêtre dans une position donnée et on considère le sous-ensemble de points (x_i, y_i) de cette fenêtre.
Soit (x_j, y_j) le point à lisser par le point (x_j, l_j) où l_j est l'ordonnée à trouver (pour les temps extrêmes, l'indice j peut désigner un point autre que le point au milieu de la fenêtre).
Les poids w_i alors affectés aux valeurs y_i des points (x_i, y_i) de la fenêtre sont évalués par:

$$[1 - (|x_i - x_j| / \max_i |x_i - x_j|)^3]^3.$$

En fait la méthode ne serait pas résistante en s'arrêtant là. Les résidus calculés à partir de cette première série lissée définissent les écarts entre valeurs observées et valeurs lissées. Ces écarts servent à définir un nouveau système de poids pour l'ensemble des valeurs observées. Des poids faibles sont donnés aux points associés aux écarts élevés. Des poids forts (proches de 1) sont donnés aux points présentant des écarts faibles. Le calcul de ces poids s'effectue souvent en se référant à une densité gaussienne de moyenne zéro, avec un mode ramené à 1. Mais une densité triangulaire donne autant de satisfaction le plus souvent. On réitère alors le lissage initial, cette fois sur la série des valeurs multipliées par ces poids respectifs.

On peut répéter plusieurs fois cette procédure pour affiner le calcul de ces poids affectés à chacune des valeurs, car il dépend toujours du calcul de la série lissée précédente (le plus usuellement, 2 ou 3 fois seulement).
Si la taille de la série est élevée, un lissage lowess est consomateur de temps calcul.

Loess: c'est le lowess sans la pondération des valeurs visant la résistance. Est une commande très générique dans Splus.
Running line: c'est le loess sans pondération lors des régressions mobiles.
Gaussian kernel: les poids des régressions sont évalués selon une densité gaussienne centrée sur le point à lisser, au lieu d'être évalués par la formule précédente. On rejoint alors l'approche des lisseurs à noyau.
Lisseurs à noyau. Les développements actuels répondant à des préoccupations diverses sont très importants, tout comme pour la technique suivante.
Smoothing spline: des polynômes sont estimés dans la fenêtre mobile d'une façon particulière, la concavité (la dérivée seconde du polynôme) intervenant dans le calcul. Les méthodes splines sont très variées et ici ce sont les splines de régression qui sont les plus adaptés.
Des méthodes splines peuvent tenir compte de contraintes dans l'évaluation de la courbe de lissage, telles de valeur minimale indépassable, saisonnier sinusoïdal, etc. Pour les séries multivariées, les splines peuvent inclure des techniques d'analyse des données.
Etc.

3.3.6 Lissage du saisonnier

L'objectif du lissage du mouvement tendanciel est de tracer une courbe autour de laquelle oscillent saisonnier et résidus, assez lisse pour que saisonnier et résidus n'apparaissent plus.

Souvent on ne se contente pas de résumer par une constante chaque mois de la série détendancialisée. Par exemple un lissage de chacunes des sous-séries associées à chaque mois permet de les débarrasser des résidus.

Le lissage du saisonnier se conduit avec l'une ou l'autre des techniques décrites pour la tendance, à partir de la série détendancialisée.

Les courbes lissées obtenues peuvent alors servir de modèle, d'évaluation de la composante s_t du modèle additif.

D'autres préoccupations que celles d'obtenir des courbes lissées autour desquelles les résidus oscillent peuvent intervenir dans la recherche d'un "bon" lissage.

Ainsi par exemple:

* on peut rechercher une certaine continuité sur la façon dont l'allure des courbes lisses se déforment quand on passe d'un mois à un autre mois voisin (et un mois de Janvier doit être considéré comme un voisin du mois de Décembre), de façon notamment à s'abstraire de l'arbitraire du découpage de mois en mois (au lieu du 15 d'un mois au 15 du mois suivant par exemple.

* on peut rechercher dans le même temps un lissage des niveaux moyens, de sorte qu'apparaisse une continuité dans la succession des T moyennes relatives aux T courbes lissées précédentes. Cependant certaines discontinuités observées peuvent être tout à fait naturelles et ne doivent pas forcément être atténuées par le lissage (un mois d'Août peut introduire une discontinuité évidente en production, à préserver et non à diminuer par lissage).

* on peut rechercher une classification des mois en 2 ou 3 catégories distinctes (formées de mois voisins) basées sur une homogénéité d'allure de courbes lisses, de façon à ramener le problème d'une série hétérogène (de construction de modèle additif, … ou de prévision) à un problème de 2 ou 3 sous-séries plus homogènes.

On regarde la série saisonnière lissée: elle constitue notre modélisation du mouvement saisonnier. On analyse alors la croissance en **amplitude** (en variabilité) éventuelle du saisonnier d'année en année, ou la décroissance, et autres phénomènes particuliers.

3.4 Calcul du modèle additif

3.4.1 Calcul

Une fois choisi le mouvement tendanciel qui ne doit pas contenir la moindre trace de mouvement saisonnier, on fait le calcul de la série détendancialisée (nommons-la d_t).

A partir de cette dernière série, on fait l'estimation du saisonnier, par moyennes, par lissages, etc., si un mouvement périodique existe. Certaines méthodes conduisent à déterminer le saisonnier avant de s'occuper de trouver une expression définitive pour la tendance.

Le mouvement saisonnier est parfois détectable visuellement sur le graphique de d_t. *Mais la lecture devient précise à l'aide du corrélogramme ou du périodogramme* (nommé aussi parfois spectrogramme) *calculé sur cette série d_t.*
Ensuite on calcule les résidus, en soustrayant à d_t le mouvement saisonnier évalué.

On dira que le modèle additif construit est **correct** notamment s'il n'existe plus dans e_t de phénomène saisonnier pour la période indiquée: *cela se vérifie en regardant le corrélogramme ou le périodogramme des résidus.* Notamment, la valeur ρ_T ne doit pas être significativement plus élevée que les valeurs dans le voisinage. Si ces graphes indiquent le contraire, l'hypothèse adoptée (par exemple) d'un saisonnier constant de période fixe peut être remise en question. Le modèle est alors à abandonner et on pourra envisager un modèle sur données transformées, ou une modélisation plus complexe du saisonnier.

Les séries économiques sont plus souvent décomposées en 4 ou 5 composantes au lieu de 3.
Ainsi, ce que nous avons appelé tendance peut constituer ce qui est nommé tendance à long terme + mouvement cyclique. Le **mouvement cyclique** est un mouvement apériodique dont les pseudo-cycles s'étalent sur de longues périodes. C'est un phénomène très basses fréquences.
De mauvais choix pour les composantes du modèle additif conduisent souvent à des résidus répartis autour de zéro par paquets de valeurs plutôt que distribués plus aléatoirement.
Quand le modèle suppose le saisonnier d'amplitude variable, par exemple avec les variations d'années en années de chacun des mois traduites par des courbes lisses au lieu d'être traduites par une constante, le saisonnier peut alors être extrapolé plus finement que par moyennes, en vue d'une estimation pour l'année suivante.
Parfois des **variations calendaires** sont à prendre en compte (Dimanche, jours fériés, heures de pointe, etc.).
Toutes ces variantes et bien d'autres proviennent de solutions apportées à des problèmes particuliers posés. Une bonne connaissance du contexte, et une pratique fréquente de ce type de modélisation parfois finalement assez complexe, peut mener à une description objective des données étudiées (et par suite conduire à un meilleur choix de technique de prévision).

3.4.2 Corrélogramme estimé

Il est estimé à partir d'une série échantillon $\{x_t\}$. C'est un outil descriptif des liaisons linéaires simples entre la série initiale et la (même) série décalée dans le temps.
Soit X_t le vecteur de composantes (x_2, x_3, …, x_{n-2}, x_{n-1}, x_n) et soit X_{t-1} celui de composantes (x_1, x_2, x_3, …, x_{n-2}, x_{n-1}). X_{t-1} est la représentation du vecteur X_t, décalé dans le passé d'une unité, comme le montre la lecture colonnes par colonnes des deux lignes suivantes:

$\mathbf{X_t}$	x_1 |	$\mathbf{x_2}$	$\mathbf{x_3}$	…	$\mathbf{x_{n-2}}$	$\mathbf{x_{n-1}}$	$\mathbf{x_n}$
$\mathbf{X_{t-1}}$		$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$	…	$\mathbf{x_{n-2}}$	$\mathbf{x_{n-1}}$ | x_n

On note ρ_1 le coefficient de corrélation linéaire entre les deux vecteurs X_t et X_{t-1} de taille n-1.

De manière analogue, quand on considère un **décalage** égal à k, on fabrique les deux vecteurs X_{t-k} et X_t correspondants, qui ont n-k composantes, et l'estimation ρ_k.

Bien que ces vecteurs aient n-k composantes, la recherche de bonnes propriétés pour les estimateurs donne lieu dans ce contexte à des estimations particulières.
On centre au préalable la variable X_t.
On obtient l'échantillon noté

$$y_1\ y_2\ y_3\ \cdots\ y_{n-2}\ y_{n-1}\ y_n.$$

Puis on calcule la variance estimée σ_X^2 de x_t, et les covariances estimées c_k:

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 \quad \text{et} \quad c_k = \frac{1}{n}\sum_{i=1}^{n-k} y_i y_{i+k}, \ \text{pour } k = 1, 2, \ldots$$

Remarquez que l'on divise par n, et non par n-k-1. Alors:

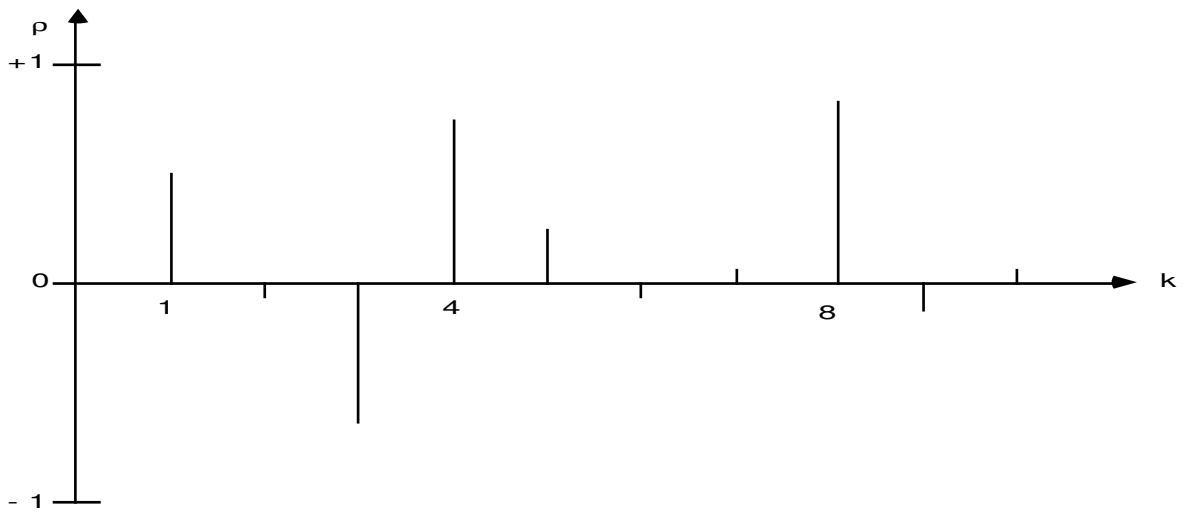
$$\rho_k = c_k \ / \ c_o,$$

est une estimation indépendante de n.

La fonction: $k \rightarrow \rho_k$, pour k = 1, 2, …, est nommée **fonction d'autocorrélation** (acf). On la note ρ.

En théorie, elle n'existe que si la série est stationnaire (§ 5.1). En particulier, cela implique que la série ne doit pas avoir de mouvement tendanciel. Et un utilisateur peut décider qu'un mouvement cyclique n'est pas à inclure dans l'expression du mouvement tendanciel... Quoiqu'il en soit, en pratique, le *corrélogramme n'est estimé que sur des séries détendancialisées.*

La plus grande valeur de k est souvent de l'ordre de n/4. Cependant, quand (n - k) n'est pas trop petit, des valeurs de k plus grandes peuvent être considérées.
Le corrélogramme suivant, qui possède des valeurs élevées (**positives**) pour ρ_4 et ρ_8, indiquerait l'existence d'un mouvement périodique de période 4.
Prolongé à des valeurs de k plus élevées, l'hypothèse de périodicité est alors confirmée s'il présente aussi de fortes valeurs positives pour ρ_{12}, ρ_{16}, … .



3.4.3 Corrélogramme résistant

Le corrélogramme venant d'être décrit n'est pas résistant: par exemple deux valeurs exceptionnelles (exceptionnellement élevées par rapport aux autres valeurs, en valeur absolue) séparées dans le temps par un décalage k, induisent un ρ_k élevé (en valeur absolue) quand la série est de taille assez faible.

Cela étant, on peut cependant comparer un corrélogramme classique ρ à celui recalculé sur données corrigées (les une ou deux valeurs les plus extrêmes étant remplacées par des valeurs moyennes; cela revient alors à se proposer un calcul de corrélogramme résistant). La comparaison avec le corrélogramme des données non corrigées permet de vérifier si ce corrélogramme est très dépendant de ces 2 valeurs exceptionnelles.

3.4.4 Périodogramme

Comme pour un corrélogramme, ce graphique ne peut être réalisé que si l'hypothèse de stationnarité est vraie. Comme elle est fausse si un mouvement tendanciel existe, la série sera souvent détendancialisée avant de calculer le périodogramme (procédure parfois automatiquement incluse dans une commande logiciel).

De cette façon la stationnarité ne sera pas forcément obtenue, mais en pratique l'outil devient alors utilisable pour la détection de périodicités.

(cependant, le calcul du périodogramme sur données initiales non détendancialisées capte le mouvement tendanciel quand celui-ci s'apparente assez bien à une sinusoïde; on pourrait parler alors de **mouvement cyclique** plutôt que de mouvement tendanciel).

Le périodogramme peut permettre de s'apercevoir de l'existence de plusieurs mouvements périodiques superposés, chose trop difficile avec un corrélogramme.

De plus, il permet d'avoir une estimation de la période plus fine, ce nombre n'étant plus forcément un multiple de l'intervalle de temps séparant deux valeurs successives. L'estimation de la période fournie par le périodogramme est à remplacer par l'entier T, le plus proche ou le plus vraisemblable, si on admet a priori l'existence du phénomène périodique de période T.

Quand le phénomène périodique n'existe pas vraiment, c'est que "la période" fluctue au cours du temps, les "années" n'étant pas toutes de même longueur. Quand n est assez grand, cela se constate en estimant les périodes de sous-séries obtenues par découpage de la série initiale, en 2 ou 3 morceaux, avec des évaluations nettement différentes de "la" période sur chaque morceau. On parle alors de **pseudo-période** (celle d'un mouvement cyclique par exemple). Cette pseudo-période n'est pas entière le plus souvent et le périodogramme en donne une évaluation d'autant plus fine que la taille de la série est importante. Le pic n'est pas bien marqué dans la mesure où les valeurs voisines les plus proches sont également élevées de sorte que la plus grande valeur (lecture locale) n'est pas nettement isolée (les points voisins sont estimations voisines de la pseudo période).

Si on présuppose de l'existence d'une période entière unique, le corrélogramme est aussi facile à lire que le périodogramme. Il existe d'ailleurs des relations (formules de convolution) entre les formules de calcul d'un corrélogramme et celles d'un périodogramme.

Supposons que n = 2q+1 soit la taille d'une série détendancialisée et centrée $\{y_t\}$. On pose $\omega_i = 2\pi i/n$.

1/n est la fréquence **fondamentale**, et i/n est celle de la i-ème **harmonique**.

On recherche une décomposition de y_t en ondes sinusoïdales, ajustée par les moindres carrés, en calculant les valeurs α_i et β_i du modèle:

$$y_t = \sum_{i=1}^{q} [\alpha_i \cos(\omega_i t) + \beta_i \sin(\omega_i t)] + e_t, \quad \forall t$$

La solution est obtenue pour

$$n\,\alpha_i = \sum_{t=1}^{n} y_t \cos(\omega_i t) \quad \text{et} \quad n\,\beta_i = \sum_{t=1}^{n} y_t \sin(\omega_i t), \forall\, i = 1, \ldots, q.$$

Le **périodogramme** $(\omega_i, I(\omega_i))$ indique en ordonnée l'intensité de la ième fréquence

$$I(\omega_i) = n\,(\alpha_i^2 + \beta_i^2)/2, \quad i = 1, \ldots, q$$

pour une abscisse proportionnelle à la fréquence, égale à ω_i ($0 < \omega_i < \pi$, en pratique).
Le périodogramme ainsi défini est nommé aussi **spectre** de y_t, étant associable à une transformée de Fourier, à sa décomposition spectrale.

Si on nomme σ_X^2 la variance de la série, on a

$$\sigma_X^2 = \sum_{i=1}^{q} (\alpha_i^2 + \beta_i^2)/2,$$

de sorte que le périodogramme est une décomposition selon les différentes fréquences de la variabilité de série détendancialisée.

Les points du périodogramme sont parfois lissés dans une fenêtre mobile, dont l'utilisateur peut modifier l'étendue, de façon à obtenir une courbe plus ou moins lisse, visant à atténuer des fluctuations non significatives d'un phénomène réel. Mais un lissage fort s'accompagne d'une **perte en résolution**, un pic bien pointu étant remplacé par un arrondi.

*C'est dire que le **spectre brut** ("raw"), non lissé, est celui le plus adapté à la lecture de périodes de mouvements saisonniers.*

La période est lue en abscisse, correspondant à un pic (une ordonnée vers le haut) bien marqué.

Par exemple, si le pic vers le haut correspond à une valeur de ω_i sensiblement égale à 0.52 (radians par mois), alors la période T vaut 12, puisque

$$0.52 \cong 2\pi i/n \Rightarrow i/n \cong 1/12 = 1/T.$$

Un axe des abscisses qui est gradué:
- de 0 à π, permet de lire les radians,
- de 0 à .5, permet de lire directement les fréquences i/n,
- de 0 à T/2, permet de repérer par l'abscisse 1 la fréquence 1/T.

Notons que le mot "frequency" est parfois utilisé dans les logiciels à la place du mot "period".
La lecture de chacune des ordonnées d'un périodogramme doit être relativisée par la notion de zone de confiance, souvent fournie au coefficient de sécurité 95%.

Souvent le pic associable à la période cherchée est accompagné sur sa droite de pics secondaires ("sidelobes", un peu moins hauts le plus souvent), correspondants à des multiples de cette fréquence issus des **sous-multiples de la période** (tels 6, 4, 3 et 2 pour la période 12 suscitant les harmoniques 1/6, 1/4, 1/3 et 1/2 de la fréquence "fondamentale" 1/12). C'est dire que ces **pics secondaires** ne sont pas à prendre en compte, n'indiquant pas d'autres mouvements périodiques. Seul le pic le plus à gauche de ces pics correspond à la période, quand la période est unique.
La lecture du périodogramme sur série ayant plusieurs mouvements périodiques n'est ainsi pas toujours très évidente: le pic d'un autre mouvement périodique pourrait être vu à tord comme un pic secondaire.
Une période 2 est associée à quelle abscisse sur un périodogramme? Peut-il y avoir des pics secondaires à droite quand la période est 4 (mouvement trimestriel)?

A part cette lecture associée aux pics bien marqués, on peut regarder si, dans leur ensemble, les grandes valeurs des ordonnées sont observées plutôt pour les basses fréquences (pour les abscisses proches de l'origine), ou plutôt pour les hautes fréquences (pour les abscisses les plus éloignées de l'origine); dans la première éventualité, cela signifie que la variabilité de la série est principalement constituée de mouvements très espacés dans le temps (et d'amplitudes assez fortes par rapport aux amplitudes des autres mouvements), et dans la seconde, que cette variabilité provient surtout de mouvements très rapprochés dans le temps (et d'assez fortes amplitudes). La notion d'espacement est à relativiser par la taille de la série.
Il ne s'agit plus ici de repérer quelques mouvements saisonniers, mais d'une évaluation de mouvements non périodiques, à l'aide d'une décomposition par composantes périodiques. La lecture ne s'attache plus à une valeur précise du spectre, mais devient globale, par exemple relative à 3 zones de fréquences, les basses, les moyennes et les hautes (diviser les valeurs de l'axe des abscisses en trois parts égales). On préfère alors parfois un spectre lissé à un spectre brut.

Si aucune zone de fréquences n'est prépondérante, comme pour certains bruits, le périodogramme est dit **plat**. Par exemple, si la largeur de la zone de confiance associée aux ordonnées remplit pratiquement toute la hauteur de la page, aucune valeur du périodogramme n'est alors vraiment significative et on peut le considérer comme plat.

Le périodogramme est souvent fourni sous la forme standard suivante:
En ordonnées, sont portées les valeurs:

$$10 \log[(\alpha_i^2 + \beta_i^2) / 2\sigma_X^2].$$

L'unité est alors le **décibel**. La zone de confiance autour des valeurs est alors dissymétrique.

La procédure de calcul programmée dans les logiciels est la procédure dite **fft** (fast fourier transform). Cette procédure ajoute notamment des zéros au deux extrêmes de la série pour constituer une série de taille convenable (une puissance de 2) nécessaire à la mise en œuvre de l'algorithme. De là peut venir une estimation peu précise de la première valeur du périodogramme (associée à la fréquence la plus basse).

Certains logiciels permettent de **filtrer** certaines fréquences. En usant de la transformée de Fourier inverse sur fréquences filtrées, on revient alors à une série qui a les mêmes caractéristiques que la série initiale à part celles provenant des harmoniques disparues dans le filtrage. Cette approche peut être mise en œuvre pour obtenir l'expression d'un mouvement particulier comme le mouvement tendanciel.

Citons aussi la possibilité de représenter une densité spectrale*évaluée à partir d'un modèle*, fonction de transfert ou AR par exemple, et qui peut être comparée au périodogramme calculé directement sur une série observée. Cette densité peut être une façon de caractériser visuellement le modèle lui-même. Si le modèle provient de la série elle-même, on a par ce moyen une bonne façon de lisser un périodogramme.
Mais ce cours n'aurait plus de fin en abordant toutes les possibilités de l'**étude spectrale** (parfois dite **étude harmonique**).

3.5 Analyse des résidus, zone de confiance

Si les résidus contiennent encore des traces du mouvement saisonnier de période T, c'est que la modélisation a échoué. On use du périodogramme (pic significatif vers le haut pour la fréquence 1/T ?) ou du corrélogramme sur les résidus (ρ_T nul ?) pour le vérifier.

On regarde aussi le graphique des résidus eux-mêmes pour savoir si la question de l'existence de valeur exceptionnelle se pose ou non, quand une valeur apparaît trop extrême par rapport à l'ensemble des autres. Cela veut alors dire que la valeur initiale correspondante est très éloignée par rapport au comportement moyen traduit par le modèle additif lui-même.
Si on sait que la valeur extrême provient d'une cause connue accidentelle, alors on peut envisager de traiter cette valeur comme une donnée manquante, et on peut se servir du modèle additif pour en faire l'estimation (par la tendance + le saisonnier au temps considéré, si le modèle est correct).
Cette correction faite, on peut refaire l'estimation d'un modèle additif à partir de la série corrigée, cette fois peu influencée par la présence de la valeur exceptionnelle (si on veut "presque pas du tout" au lieu de "peu", on itère la procédure 1 ou 2 fois de plus). Cette démarche est inutile si la méthode employée pour l'évaluation de la tendance et du saisonnier est résistante.

Pour savoir si la valeur exceptionnelle est sans influence sur les valeurs voisines, on observe attentivement les valeurs des résidus voisins: par exemple une grande campagne publicitaire sur un produit peut provoquer une très forte vente un mois donné, et une vente assez forte les quelques mois qui suivront, jusqu'à un retour à la normale. Ou bien la très forte vente du mois est suivie de méventes les mois suivant, car les acheteurs ont stocké pour profiter de la promotion, sans vouloir consommer plus

pour autant, de sorte que la campagne tombe à l'eau (à moins que le problème consistait justement pour le vendeur à faire supporter par les acheteurs un problème de stockage…).
C'est dire que plusieurs valeurs, situées au voisinage de la valeur exceptionnelle, pourraient être alors corrigées aussi, en fonction des informations que l'on a par ailleurs (un statisticien "pur" ne peut travailler seul que sur des données qu'il aurait lui-même simulées. Cela fait penser à race pure, math. pure, …, de pures aberrations).
La difficulté existe lorsqu'une valeur paraît exceptionnelle, alors qu'aucune explication n'est connue: faut-il corriger et recommencer le calcul ou pas? Le résultat peut en effet en être sensiblement modifié si la taille de la série est relativement faible et l'exception vraiment bien marquée.
Ou bien faut-il se réjouir d'avoir choisi dès le départ un mode de calcul des différents termes du modèle, insensible à ces valeurs extrêmes, c'est à dire d'avoir choisi une méthode résistante?

Quand les résidus ont une moyenne sensiblement éloignée de 0, on peut penser à intégrer cette moyenne dans le mouvement tendanciel, en contrôlant (graphe superposé de la série et des deux expressions de la tendance) si l'effet de cette translation de la tendance vers le haut ou vers le bas est vraiment souhaité.

Dans le cas où les résidus sont non autocorrélés (corrélogramme nul, ce qui arrive plutôt rarement car le modele additif n'a pas cet objectif), on peut proposer autour du modèle additif une **zone de confiance** à un coefficient de sécurité donné:
Par exemple, si l'hypothèse de distribution gaussienne des résidus est envisageable [qq-plots, boîte à pattes (boxplot), branchage (stem and leafs), histogramme, test, …] et si de plus les résidus peuvent être considérés comme nuls en moyenne, la zone de confiance autour du modèle proprement dit est, au coefficient de sécurité 95%:

$$(z_t + s_t) \pm 2\ s_e,\ \forall\ t.$$

En effet, pour une variable centrée gaussienne X, $P(X > 1{,}96\ s_X) = .025$, et souvent 1,96 est arrondi à 2. Cette valeur provient ainsi de la loi normale et du coefficient de sécurité fixé à 95%.

Dans le cas d'une distribution symétrique, on pourrait aussi travailler sans hypothèse de loi sur les résidus en remplaçant la valeur $2s_e$ par la moitié de l'étendue des valeurs obtenues après avoir éliminé 2.5% des valeurs les plus petites et 2.5% des valeurs les plus grandes.
Sans hypothèse de loi, sans hypothèse sur la symétrie, comment faites-vous?

4 PRÉVISION PAR LISSAGE EXPONENTIEL

Ces lissages ont peu de chose à voir avec les lissages précédents, n'étant pas faits pour traduire un comportement tendanciel au sens précédent, mais pour effectuer un calcul de prévision à un moindre coût, ou adapté à un comportement tout à fait spécifique de la série étudiée.
Les plus élémentaires sont simples à programmer, à faire tourner vite, à mettre à jour. Ils peuvent être adaptés à une gestion de stocks pas trop coûteuse de très nombreux articles, sur un ordinateur de taille réduite, et adaptés à la prévision de séries de taille très réduite.

Une f**ormule de mise à jour** est une formule qui permet de prendre en compte des valeurs nouvellement observées, pour faire un réajustement des prévisions anciennement calculées, sans avoir à réestimer le modèle de prévision.
Un modèle additif pourrait être utilisé comme modèle de prévision en prolongeant la tendance et le saisonnier. Mais il n'est pas vraiment conçu pour cela, et le plus souvent cela conduit à une imprécision de la prévision, par rapport à des techniques plus adaptées. De plus, il n'y a pas de formule de mise à jour qui l'accompagne et chaque fois qu'une nouvelle observation est recueillie, il faudrait recalculer le modèle pour prendre en compte la nouvelle valeur.

4.1 Le lissage exponentiel simple

C'est une technique de prévision adaptée aux séries non saisonnières de tendance horizontale (constante en moyenne).

On veut trouver la valeur réelle **a** qui s'approche le mieux, aux sens des moindres carrés pondérés, de l'ensemble des valeurs d'une série $\{y_t\}$.
Soit $\beta \in]0,1[$, une constante arbitraire fixée.
Les poids β^j permettent d'attribuer plus d'importance aux dernières valeurs observées et peu d'importance aux valeurs du passé lointain. Il s'agit ainsi de minimiser

$$f(a) = \sum_{j=0}^{n-1} \beta^j (y_{n-j} - a)^2$$

La solution est obtenue pour

$$a = l_n = (1-\beta) \sum_{j=0}^{n-1} \beta^j y_{n-j}$$

Si on veut que les valeurs du passé lointain influence très peu cette solution, on choisit β assez éloigné de 1, sinon, assez proche de 1 (la valeur 0,7 est souvent adoptée). On montre que le calcul est plutôt peu sensible aux faibles variations de ce paramètre. Des études ont été faites pour optimiser la valeur de β. L'optimisation de β est cependant un problème qui dépend du critère à optimiser (confer § 4.4).

Notons $\hat{y}_n(h)$, h=1,2,…, **la prévision faite au temps n à l'horizon h**, c'est à dire l'estimation de la valeur prévue pour l'instant du futur n + h, calculée au moment présent n. Le mouvement tendanciel de la série étant *horizontal*, on convient de réaliser le calcul de prévision à l'aide de la formule:

$$\hat{y}_n(h) = l_n \ , \ \forall \ h = 1, 2,\ldots$$

Mise à jour

On suppose maintenant que l'on connaît aussi y_{n+1}, pour l'avoir dernièrement observée. Il faut en tenir compte pour une réévaluation des nouvelles prévisions, désormais faites à l'instant n+1.

$$l_{n+1} = (1-\beta) \sum_{j=0}^{n} \beta^j y_{n+1-j} = (1-\beta)\, y_{n+1} + (1-\beta)\, \beta \sum_{i=0}^{n-1} \beta^i y_{n-i} = (1-\beta)\, y_{n+1} + \beta\, l_n$$

$$l_{n+1} = l_n + (1-\beta)\,(y_{n+1} - l_n)$$

$$l_{n+1} = l_n + (1-\beta)\, e_n(1).$$

Donc la valeur de la prévision, faite au temps n+1, est la prévision précédente (faite au temps n) à laquelle est ajouté un terme correctif proportionnel à l'erreur de prévision $e_n(1)$ commise au temps n pour l'horizon 1. Ainsi, la formule de mise à jour des prévisions peut s'écrire:

$$\hat{y}_{n+1}(h) = \hat{y}_n(h+1) + (1-\beta)\, e_n(1) \ , \ \forall \ h = 1, 2,\ldots.$$

4.2 Lissage exponentiel double

C'est une technique de prévision adaptée lorsque le mouvement tendanciel est linéaire (droite non horizontale).

On veut trouver les valeurs a_1 et a_2 telles que la droite d'équation $y = a_1 + a_2 t$, s'approche au mieux, aux sens des moindres carrés pondérés, de l'ensemble des valeurs d'une série $\{y_t\}$ non saisonnière. Il s'agit ainsi de minimiser

$$f(a_1, a_2) = \sum_{j=0}^{n-1} \beta^j [y_{n-j} - (a_1 + a_2(n-j))]^2, \text{ pour n fixé.}$$

On note $l_1(n) = l_n = \hat{y}_n(1)$, $\forall$ n, les valeurs d'un premier calcul par lissage exponentiel simple. Ainsi, dès la valeur n=1, la formule de mise à jour (ou la formule de définition des prévisions) permet le calcul successif des valeurs $l_1(2), l_1(3),\ldots, l_1(n)$.

On pose $l_1(1) = y_1$ et on considère la nouvelle série de taille n notée l_1. Toujours avec la même formule de mise à jour, on calcule alors les valeurs $l_2(n)$, par une deuxième série de lissages exponentiels simples effectués cette fois sur la série l_1

$$l_2(n) = (1-\beta) \sum_{j=0}^{n-1} \beta^j \ l_1(n-j).$$

Les coefficients de la droite cherchée peuvent alors s'exprimer à partir du double lissage:

$$a_1 = a_1(n) = 2\, l_1(n) - l_2(n)$$

$$a_2 = a_2(n) = [(1-\beta)/\beta] \ [l_1(n) - l_2(n)],$$

et les prévisions faites à l'instant n sont donc données par

$$\hat{y}_n(h) = a_1(n) + h\, a_2(n)\ ,\ \ \forall\ h = 1, 2,\ldots.$$

On initialise ces formules avec $a_1(2) = y_2$ et $a_2(2) = y_2 - y_1$.

Les formules de mise à jour sont

$$a_2(n+1) = a_2(n) + (1-\beta)^2\,[y_{n+1} - \hat{y}_n(1)]$$

soit encore:

$$a_2(n+1) = a_2(n) + (1-\beta)^2\ e_n(1)$$

$$a_1(n+1) = a_1(n) + a_2(n) + (1-\beta^2)e_n(1)$$

de sorte que

$$\hat{y}_{n+1}(h) = \hat{y}_n(h+1) + [1-\beta^2 + h(1-\beta)^2]\ e_n(1),\quad \forall\ h = 1, 2,\ldots,\ \text{et}\ \ \forall\ n > 1.$$

4.3 Lissages exponentiels généralisés

Ils se construisent avec des hypothèses de mouvements tendanciels variés.

Ils correspondent au problème de moindres carrés pondérés minimisant:

$$f(a_1, a_2, \ldots, a_k) = \sum_{j=0}^{n-1} \beta^j\,[y_{n-j} - \Sigma_{i=1,\ldots,k}\ a_i\, g_i(j)]^2,\quad \text{pour n fixé,}$$

les fonctions connues g_i étant des fonctions puissance, ou des fonctions sinusoïdales, ou autres (confer **Brown**, 1962).

Le lissage exponentiel de **Holt** est un lissage ayant le même objectif que le lissage exponentiel double, avec cette fois deux paramètres arbitraires à choisir, pour plus de souplesse.

Si on désigne ces deux paramètres par $\alpha \in\ ]0,1[$ et $\gamma \in\ ]0,1[$, les formules de mise à jour sont :

$$a_1(n+1) = (1-\alpha)\ y_{n+1} + \ \alpha\ (a_1(n) + a_2(n))$$

$$a_2(n+1) = (1-\gamma)\ (a_1(n+1) - a_1(n)) + \ \gamma\, a_2(n).$$

Winters adapte ce modèle aux séries saisonnières de période connue, introduisant additivement le saisonnier. Les coefficients saisonniers c_j sont mis à jour, avec un troisième paramètre arbitraire à choisir

$$a_1(n+1) = (1-\alpha)\ (y_{n+1} - c_{n+1-T}) + \ \alpha\ (a_1(n) + a_2(n))$$

$$a_2(n+1) = (1-\gamma)\ (a_1(n+1) - a_1(n)) + \ \gamma\, a_2(n)$$

$$c(n+1) = (1-\delta)\ (y_{n+1} - a_1(n+1)) + \ \delta\, c_{n+1-T}$$

Il propose aussi une autre variante, où le saisonnier se multiplie avec la droite au lieu de s'ajouter (modèle mixte, dit par l'auteur modèle multiplicatif).
La complexité introduite rend alors ces techniques moins attrayantes. Le choix des paramètres peut être optimisé, selon l'un ou l'autre des critères ci-dessous, ce qui les rend encore plus lourdes.
C'est pourquoi les techniques de Box et Jenkins exposées par la suite seront souvent préférées, d'autant qu'on montre (**sauf pour Holt-Winters multiplicatif** qui ne peut être mis en relation avec ces techniques) qu'elles sont supérieures, quand le critère d'appréciation est celui de la variance minimale de l'erreur (cf. 4.4). Cependant la maîtrise acquise par certains dans la pratique de ces méthodes anciennes, accompagnée d'une grande connaissance des données, est un facteur d'appréciation qui ne peut être négligé. Quoiqu'il en soit, toutes les méthodes de prévision exposées dans ce cours sont des méthodes de prévision à court terme, où le phénomène tendanciel est seulement extrapolé depuis son comportement terminal.

4.4 Critères usuels de mesure de l'erreur de prévision (court terme)

On suppose que l'on dispose de n prévisions $\hat{y}_1$, $\hat{y}_2$, …, $\hat{y}_n$, faites à l'horizon 1, à partir des observations $y_1, y_2, \ldots, y_n$.

Par exemple, avec le lissage exponentiel simple (noté l_1), dès que la première valeur x_1 est connue, on a la prévision à l'horizon 1 et l'erreur de prévision évaluable quand la deuxième valeur est connue:

$$\hat{y}_2 = \hat{y}_1(1) = l_1(2) \text{ , et } \hat{e}_2 = x_2 - l_1(2).$$

Puis, connaissant les deux premières valeurs x_1 et x_2,

$$\hat{y}_3 = \hat{y}_2(1) = l_1(3) \text{ , et } \hat{e}_3 = x_3 - l_1(3).$$

Jusqu'à la connaissance des n valeurs:

$$\hat{y}_n = \hat{y}_{n-1}(1) = l_1(n) \text{ , } \hat{e}_n = x_n - l_1(n).$$

On peut aussi poser $\hat{y}_1 = \hat{y}_0(1) = x_1$. Ou prendre une valeur "post vue" en usant de la méthode de lissage utilisée, mais en ayant renversé le sens du temps...
Finalement, on dispose d'une évaluation de l'erreur de prévision faite avec une méthode donnée, pour chacunes des valeurs du temps, *prévisions toutes faites à l'horizon 1*.

On s'intéresse ici à une notion de prévision à très court terme et d'autres critères que ceux ci-après pourraient être envisagés s'il s'agissait de prévoir plus loin.
Pour mesurer ou comparer l'efficacité de différentes approches, on choisit l'un des critères, ce choix dépendant de l'objectif recherché.

* La variance des erreurs, ou l'écart-type.

* La moyenne des écarts à la médiane, ou parfois la moyenne des écarts à la moyenne "mean absolute deviation".

* La moyenne des carrés des erreurs (moment d'ordre 2), sensible au fait que les erreurs ne sont pas de moyenne nulle (ou sa racine carrée, erreur quadratique moyenne, "root mean square error"): ainsi par exemple le fait d'avoir plus d'erreurs par excès que par défaut augmente la valeur de ce critère, plus facilement qu'avec le critère variance.
Dans le même ordre d'idée, on peut proposer la moyenne des valeurs absolues des erreurs (sensible au fait que le nombre d'erreurs positives est différent du nombre d'erreurs négatives).

Chercher des pénalités pour dissymétrie de distributions centrées sur zero, demanderait d'user d'autres critères, à base de quantiles par exemple.
* On use aussi du critère **AIC** d'Akaike ("I" comme Information):

$$AIC= n \, Log(var(e)) + 2p$$

où p est le nombre de paramètres estimés de la formule de prévision et var(e) est la variance empirique de l'erreur de prévision (diviser par n et non par n-1 pour estimer cette variance).

* ou bien du SBIC=**BIC**=SCB ("Bayesian information criterion"), mathématiquement mieux justifié que AIC:

$$BIC= n \, Log(var(e)) + p \, Log(n)$$

Les deux critères précédents prennent cependant des formes un peu différentes lors de contextes précis (confer les prévisions par modèles arima).

Il est utile d'observer le corrélogramme des erreurs: s'il est intéressant de comparer entre eux les résultats de différents modèles de lissages exponentiels, il est au moins aussi intéressant de savoir si un modèle de prévision est susceptible de pouvoir être amélioré ou non.
Quand les erreurs sont assimilables à un "bruit", c'est à dire quand le corrélogramme est nul, cela signifie le plus souvent qu'il n'y a pas espoir de faire mieux. Quand le corrélogramme n'est pas nul, souvent la méthode de prévision pourra être améliorée (en modifiant la valeur d'un paramètre à choisir, en usant d'une autre technique, en transformant, etc.…).
De plus, les mesures précédentes sont peu valides statistiquement quand le corrélogramme n'est pas nul, et ne sont alors à considérer qu'à titre indicatif (confer 11.3).

On compare aussi parfois l'efficacité d'une méthode de prévision donnant les valeurs $\hat{y}_{i-1}(1)$, à celle de la **prévision dite naïve** qui consiste à évaluer la prévision au temps i par la valeur observée y_{i-1}. Les critères sont alors fondés sur l'écart entre les deux types d'erreur:
* avec le critère MAPE :

$$(n-1) \; MAPE=\Sigma_{i = 2\ldots n} \left| (\hat{y}_{i-1}(1) - y_{i-1}) / y_i \right|$$

* ou mieux avec le critère U:

$$U = [(\Sigma_{i = 2\ldots n} \left| (y_i - \hat{y}_{i-1}(1)) / y_{i-1} \right|^2) / (\Sigma_{i = 2\ldots n} \left| (y_i - y_{i-1}) / y_{i-1} \right|^2)]^{1/2}$$

qui vaut 1 si la méthode de prévision est aussi mauvaise que la méthode naïve, et zéro si la prévision atteint la perfection.

5 LES MODÈLES STOCHASTIQUES ou PROCESSUS

5.1 Définitions

Le modèle additif est un modèle déterministe, et non un modèle probabiliste. Il n'a pas été introduit en vue d'un calcul de prévisions, bien que les renseignements apportés par sa mise en œuvre peuvent être utiles pour mener à bien une prévision.
Ici les modèles probabilistes sont des processus stochastiques construits pour fournir une explication du présent par le passé.

Dans un premier temps, nous définissons une famille de processus à l'intérieur de laquelle seront considérées par la suite certaines classes caractérisantes de processus (celles de Box et Jenkins, 1976).

Les lettres majuscules sont réservées à la désignation de variables aléatoires et les minuscules le sont pour des valeurs prises (observées ou estimées) par ces variables aléatoires, c'est à dire par ces fonctions.

On nomme ici processus les processus aléatoires discrets suivants: une famille de variables aléatoires $\{X_t\}_{t \in Z}$ définies sur un même espace probabilisé (Ω, a, P), et à valeurs dans (R, B_R).

L'indice t désigne le temps. Pour $t \in Z$ fixé, x_t est valeur issue de la variable X_t de la famille $\{X_t\}$.
Des valeurs observées $x_1, x_2, \ldots, x_n$ seront considérées comme issues d'un processus $\{X_t\}_{t \in Z}$, aux temps $t = 1, 2, \ldots, n$.
L'intervalle entre deux valeurs successives est donc toujours le même, d'étendue standard égale à 1.

Les processus considérés seront des processus du $2^{ième}$ ordre, c'est à dire en particulier tels que $E(X_t)$ et $var(X_t)$ existent, $\forall\, t \in Z$.

On note L^2 l'espace de Hilbert des variables aléatoires réelles de carré intégrable, définies sur (Ω, a, P), muni du produit scalaire $\langle X, Y \rangle = E(XY)$ et de la norme $\| X \| = [\, E(X^2)]^{1/2}$.
Cette norme permet la mesure de la distance entre deux variables; elle permet aussi par extension d'évaluer les distances entre processus formés de variables de L^2.
La notion de voisinage entre processus ainsi conçue impliquera celle de modèles plus ou moins proches entre eux, extensible à celle de processus plus ou moins bien ajusté à des observations.

Définir un processus revient ici à définir une application de Z dans L^2.

Dans un premier temps, ne seront considérés que des processus stationnaires.
La stationnarité d'un processus est relative à la loi du processus. La stationnarité forte est une propriété recherchée parce que c'est l'hypothèse fondamentale (dans ce cours) pouvant rendre possible un calcul de prévision.

On dit qu'un processus $\{X_t\}$ est **fortement stationnaire** si
$\forall\ n \geq 1, \in N$, la loi marginale du n-uplé $(X_{i1}, X_{i2}, \ldots, X_{in})$ est identique à la loi marginale du n-uplé $(X_{i1+k}, X_{i2+k}, \ldots, X_{in+k})$, pour tout $k \in Z$.

C'est une propriété très forte, qui implique en particulier
- qu'une même loi régit chacune des variables X_t , $\forall\ t \in Z$.
- que la loi de $(X_1, X_2, \ldots, X_n)$ est aussi celle de $(X_{1+k}, X_{2+k}, \ldots, X_{n+k})$, $\forall\ k \in Z$.

Cette dernière ligne implique:
- qu'observer un échantillon de mesures $x_1, x_2, \ldots, x_n$ aux temps 1, 2, …, n, plutôt que de faire plus tard cette observation de n mesures successives (ou en ayant commencé avant en supposant que cela puisse être), n'est pas très important: la différence proviendra seulement de la notion de fluctuation d'échantillonnage, les variables aléatoires dont sont issues ces observations ayant leurs lois mutuelles inchangées.
- que les lois mutuelles déterminées à partir d'observations du passé pourront être légitimement utilisées dans le futur, puisqu'elles sont invariantes par une translation du temps.

En pratique, cette propriété de stationnarité forte (ou stricte) sera supposée sans pouvoir être vérifiable ou testée.

Celle de stationnarité faible (moins forte que la précédente, large) ne le sera pas beaucoup plus. Seulement quelques approches de type confirmatoire sont possibles en pratique.
Par exemple, si l'hypothèse de stationnarité est raisonnable pour mes données, alors le processus stationnaire estimé à partir de la première moitié de la série sera voisin de celui estimé à partir de la seconde moitié; en faisant les deux calculs et la comparaison, j'infirmerai ou non la réalité de l'hypothèse, selon que ce voisinage traduit une forte proximité ou non (et donc peu ou beaucoup de changement par translation dans le temps).
Un processus $\{X_t\}$ du $2^{ième}$ ordre sera dit dans ce cours **stationnaire**, s'il est **faiblement stationnaire**, soit:

$$\forall\ t, \forall\ s, (t, s) \in Z^2,\ \ cov(X_t, X_s)\ \text{ne dépend que de l'écart } |t - s|, \text{ et non de t ou de s.}$$

Seule la stationnarité "forte des variables prises deux à deux ou une à une" est visée (sans être forcément atteinte) par cette stationnarité faible. Nous sommes loin de la stationnarité forte concernant la loi d'un processus. Cependant, comme les modèles sont estimés à partir de telles covariances, la stationnarité faible est l'hypothèse minimale à considérer.

Une série observée sera dite stationnaire quand un processus stationnaire modélise convenablement cette série.
*Un **processus gaussien** (normal) est un processus tel que la loi de tout k-uplé de variables issus du processus est gaussienne (normale).*

Théorème:
Pour un processus gaussien, la stationnarité au sens large équivaut à la stationnarité stricte.

On aura donc intérêt à avoir plutôt une série observée modélisable par un processus gaussien. Sinon, on peut envisager par exemple une substitution de la série par une série transformée qui s'avérerait modélisable par un processus gaussien. Ou bien on peut penser à user d'une technique de prévision ne faisant pas appel à l'hypothèse de la stationnarité.
De plus, en pratique, l'hypothèse de normalité d'une série ne peut se tester facilement; l'introduction d'hypothèses supplémentaires, comme par exemple celle de stationnarité, est nécessaire. Pour effectuer des travaux la supposant vraie, on se contente donc le plus souvent de quelques procédures confirmatoires. Quant au contraire on a constaté que l'hypothèse gaussienne n'est pas vérifiée, on veille à ne réaliser que les calculs n'exigeant pas cette hypothèse, ou encore on émet des réserves quand aux calculs produits grâce à l'hypothèse gaussienne.
Quand une série est de taille trop courte et sans autre information, on est bien incapable de pouvoir affirmer que l'hypothèse gaussienne est valide, le risque de se tromper étant élevé. Ce qui n'empêche pas de faire des calculs la supposant**:** les valeurs calculées sous cette hypothèse doivent alors être considérées avec une certaine réserve.

Un processus **centré** $\{X_t\}$ est un processus tel que $E(X_t) = 0, \forall\ t$.

Dans un premier temps, tous les processus seront centrés et stationnaires.

Ils vérifient donc aussi la propriété dite d'**homoscédasticité**:

$$var(X_t) = E(X_t^2) = \sigma_X{}^2, \text{ constante } \forall\ t.$$

La variance du processus correspond alors à cette valeur.
On a de plus:

$$\forall\ k\ ,\ \ cov(X_t, X_{t+k}) = cov(X_t, X_{t-k}) = E(X_tX_{t-k}) = c_k\ ,\ \text{ constante } \forall\ t,$$

et

$$\rho_k = c_k\ /\ \sigma_X{}^2.$$

Le corrélogramme d'un tel processus alors existe.
Notez que $\rho_o = 1$.

Confer § 2.4.2 pour les estimations, qui supposent l'ergodicité du processus.

Un processus est dit **ergodique** quand les limites des moments empiriques convergent en moyenne quadratique vers les espérances correspondantes, lorsque la taille de la série s'accroît.
Ce sont des processus identifiables à partir d'une trajectoire infinie.

En particulier, en observant le processus suffisamment longtemps, les estimations des paramètres des modèles considérés dans ce cours faites à partir des observations sont alors valides, ... ce que nous supposerons de toutes façons à priori quand on voudra modéliser.
Cela implique aussi que, lorsqu'on observe suffisamment longtemps le phénomène, les estimations ne seront pas influencées par des conditions initiales particulières (par exemple de mise en route des premières observations, le régime de croisière étant atteint "assez" vite). Les procédures d'estimation de paramètres de modèle prennent souvent en compte l'ergodicité (on parle aussi de "propriété d'ergodisme" au lieu d'ergodicité).

Ainsi par exemple la moyenne d'un processus stationnaire est définie par $E(X_t)\ \forall\ t$, mais pour t fixé nous n'avons aucune répétition d'échantillon pour estimer cette moyenne. L'ergodicité du processus stationnaire suppose que, si le temps d'observation du processus est assez long, la moyenne temporelle va brasser des échantillons de t qui vont être approximativement indépendants et qui ainsi permettront une estimation depuis l'égalité suivante, qui suppose donc l'ergodicité:

$$E(X_t) = \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} x_i$$

Théorème

Tout processus stationnaire gaussien $\{X_t\}$ tel que

$$lim\ cov(X_t, X_s) \to 0 \qquad quand\ |t\text{-}s| \to +\infty$$

est ergodique.

Pour les processus stationnaires considérés dans ce cours, ρ_k tend vers zéro quand k tend vers l'infini.

Donc ils sont ergodiques s'ils sont gaussiens.

Opérateurs linéaires dans L^2

Ce sont les opérateurs définissant des nouveaux processus à partir de combinaisons linéaires de processus. En fait, nous nous intéressons ici essentiellement à une façon particulière de combiner les processus entre eux, via l'opérateur retard B (comme Backward).

Soit un processus $\{X_t\}$ quelconque.
On appelle **opérateur retard** l'opérateur linéaire B défini par:

$$B(X_t) = X_{t-1},\ \forall\ t\ ,$$

ce qui permet de définir un nouveau processus $B(\{X_t\})$ noté $\{X_{t-1}\}$.

On note **1** l'opérateur identité et B^2 l'opérateur BoB:`

$$1(X_t) = X_t,\ \forall\ t, \text{ et } \ B^2(X_t) = X_{t-2},\ \forall\ t.$$

Plus généralement, on note P(B) le polynôme en B (comme on parle de polynôme en "x") de degré n, en confondant la notation a_0 avec celle de $a_0 1$.

$$P(B) = a_0 + a_1 B + a_2 B^2 + a_3 B^3 + \ldots + a_n B^n.$$

Et le processus $P(B)(\{X_t\})$ est ainsi celui défini par la famille des variables

$$P(B)(X_t) = (a_0 + a_1B + a_2B^2 + a_3B^3 + \ldots + a_nB^n)\, X_t, \quad t \in Z$$

$$P(B)(X_t) = a_0X_t + a_1X_{t-1} + a_2X_{t-2} + a_3X_{t-3} + \ldots + a_nX_{t-n}, \quad t \in Z.$$

Les opérations habituelles sur les polynômes sont valides dans la mesure où, par exemple,

$$(1-B)^2\ (\{X_t\}) \equiv (1-2B+B^2)\ (\{X_t\}).$$

On fera même usage de fractions rationnelles, de décomposition en éléments simples de ces fractions, de développements polynomiaux infinis. Par exemple, x étant ici la variable analytique et λ un nombre complexe inférieur à 1 en module, on sait que:

$$(1-\lambda x)^{-1} = 1 + \lambda\, x + \lambda^2\, x^2 + \lambda^3\, x^3 + \ldots$$

de sorte que, avec $|\lambda| < 1$, on note:

$$(1-\lambda B)^{-1} = \Sigma_{0\ldots\infty}\ \lambda^i\, B^i$$

ce qui permet de définir un nouveau processus par la notation

$$(1-\lambda B)^{-1}\ (\{X_t\}).$$

Théorème 5.1

Les processus précédents tels que $P(B)(\{X_t\})$ et $(1-\lambda B)^{-1}\ (\{X_t\})$ sont stationnaires si le processus $\{X_t\}$ est stationnaire.

Plus généralement un opérateur f est un **filtre linéaire** quand c'est un opérateur linéaire invariant par changement de l'origine du temps. Le processus filtré $\{Y_t\}$ défini par

$$Y_t = f(X_t) \quad \forall\ t,$$

vérifie alors:

$$Y_{t-\tau} = f(X_{t-\tau}) \quad \forall\ \tau, \forall\ t,$$

et est alors stationnaire si $\{X_t\}$ est stationnaire.

5.2 Fonction d'autocorrélation partielle

5.2.1 Introduction

La fonction d'autocorrélation ρ estimée a été définie au § 2.4.2 par $\rho(k) = \rho_k$, pour k=1, 2, …,. Sa représentation graphique est le corrélogramme ρ. Il est toujours calculable en pratique sur une série sans mouvement tendanciel, mais il peut n'avoir aucun sens.

Ici nous en donnons la définition théorique, associée à un processus stationnaire.

A cause de la stationnarité, $\rho_k = \rho(X_t, X_{t-k})$, $\forall$ t.

En particulier, $\rho_k = \rho(X_n, X_{n-k})$, et quand n est la dernière valeur observée, associable au temps présent, n-k est un temps du passé, décalé de k unités par rapport au présent n.
Nous dirons souvent que X_{n-k} est la variable associée au passé d'ordre k et que la variable X_{t-k} représente le passé par rapport à t d'ordre k (t étant ici assimilé au présent).
Ainsi ρ_k représente l'expression du lien linéaire entre le présent et le passé d'ordre k.
Quand ρ_k est proche de 1, c'est qu'une grande (resp. petite) valeur observée au temps t-k est suivie le plus souvent d'une grande (resp. petite) valeur au temps t, et cela $\forall$ t.
Quand ρ_k est proche de -1, c'est qu'une grande (resp. petite) valeur observée au temps t-k est suivie le plus souvent d'une petite (resp. grande) valeur au temps t, et cela $\forall$ t.
Quand ρ_k est proche de 0, c'est qu'une grande ou une petite valeur observée au temps t-k est suivie le plus souvent d'une valeur quelconque au temps t, et cela $\forall$ t.

Nous savons que ces valeurs ρ_k, k=1, 2, …, sont insuffisantes pour expliquer comment le présent est relié au passé, puisque les variables du passé sont elles-mêmes liées entre elles (confer la régression mutiple au sens usuel).
Un autre descriptif de ces liens n'est donc pas inutile et, sous-jacente ici, est la préoccupation de savoir jusqu'où dans le passé il est nécessaire de remonter, pour obtenir une bonne explication du présent.

5.2.2 Définition de Φ

On considère cette fois la fonction d'**autocorrélation partielle** Φ, définie plus loin par $\Phi(k) = \phi_{kk}$, pour les décalages k=1, 2, ….
Sa représentation graphique est le corrélogramme Φ. Cette fonction n'est définissable que pour un processus stationnaire, même si en pratique un corrélogramme peut toujours être (soi-disant) estimé à partir d'observations qui ne supportent pas cette hypothèse.

ϕ_{11} est défini dans la régression linéaire de X_t en X_{t-1}. A cause de la stationnarité, ce problème est identique à lui-même $\forall$ t.

$$X_t = \phi_{11} X_{t-1} + R_{t,1} , \ \forall \ t$$

où $\phi_{11} \in \Re$, et la variable résidu $R_{t,1}$ est non corrélée à la variable explicative X_{t-1}.

ϕ_{22} est défini dans la régression linéaire de X_t en (X_{t-1}, X_{t-2}).

$$X_t = \phi_{12} X_{t-1} + \phi_{22} X_{t-2} + R_{t,2} ,$$

où $\phi_{22} \in \Re$, et $R_{t,2}$ est non corrélée aux variables explicatives X_{t-1} et X_{t-2}.

ϕ_{kk} est défini dans la régression linéaire de X_t en $(X_{t-1}, X_{t-2}, …, X_{t-k})$.

$$X_t = \phi_{1k} X_{t-1} + \phi_{2k} X_{t-2} + … + \phi_{kk} X_{t-k} + R_{t,k} ,$$

où $\phi_{kk} \in \Re$, et $R_{t,k}$ est non corrélée aux variables explicatives.

Quand le rajout d'une nouvelle variable d'un passé plus lointain (X_{t-k+1}) n'apporte plus rien, le calcul donne un coefficient ($\phi_{k+1,k+1}$) nul.
Il arrive ainsi que les coefficients ϕ_{kk} des nouvelles variables explicatives soient tous nuls à partir d'un certain rang, quand les valeurs k (valeurs du décalage entre présent et passé) sont considérées assez grandes.
Cela implique que les variables du passé proche, associées au premières variables explicatives, pourraient suffire pour expliquer totalement le présent.

5.2.3 Propriété

Les coefficients ϕ_{kk} sont des coefficients de corrélations:

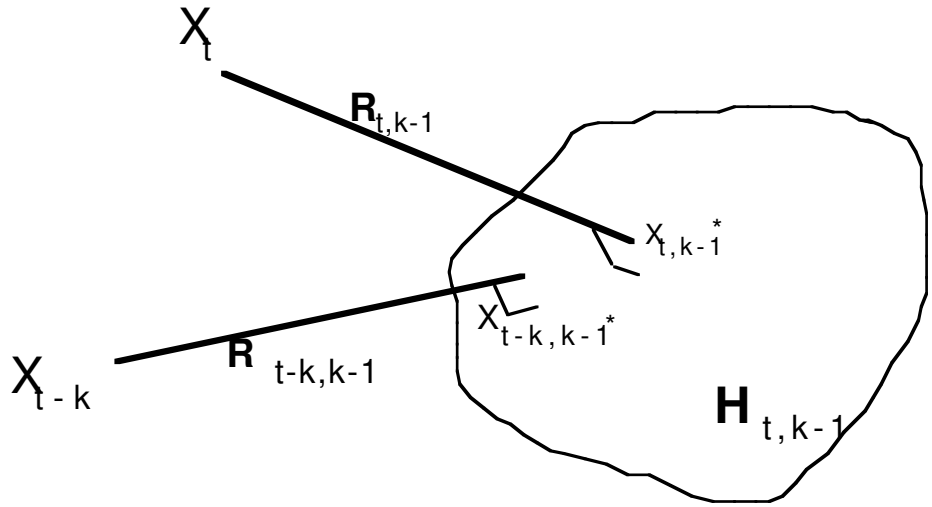
Pour k fixé, on considère la formule de régression multiple précédente:

$$X_t = \phi_{1k} X_{t-1} + \phi_{2k} X_{t-2} + \ldots + \phi_{kk} X_{t-k} + R_{t,k} \qquad (1)$$

On pose: $X_{t,k}^* = \phi_{1k} X_{t-1} + \phi_{2k} X_{t-2} + \ldots + \phi_{kk} X_{t-k}$.
Dans L^2, c'est la projection orthogonale de X_t sur le sous-espace de Hilbert $H_{t,k}$ engendré par les variables X_{t-1}, X_{t-2}, …, X_{t-k}.

Considérons le sous-espace de Hilbert $H_{t,k-1}$, engendré à partir des variables X_{t-1}, X_{t-2}, …, $X_{t-(k-1)}$.

Soient $X_{t,k-1}^*$ et $X_{t-k,k-1}^*$ les projections orthogonales respectives de X_t et X_{t-k} sur $H_{t,k-1}$, les variables résidus respectives de ces régressions étant notées $R_{t,k-1}$ et $R_{t-k,k-1}$.



A cause de la stationnarité, de mêmes coefficients de régression interviennent:

$$X_{t,k-1}^* = \phi_{1,k-1} X_{t-1} + \phi_{2k} X_{t-2} + \ldots + \phi_{k-1,k-1} X_{t-k+1}$$

$$X_{t-k,k-1}^* = \phi_{1,k-1} X_{t-k+1} + \phi_{2k} X_{t-k+2} + \ldots + \phi_{k-1,k-1} X_{t-1}$$

et

$$\phi_{kk} = \rho(R_{t,k-1}\ ,\ R_{t-k,k-1}).$$

Ainsi, la mesure ϕ_{kk} de l'intensité d'explication supplémentaire apportée par la dernière variable explicative de la régression (1) correspond à une mesure du lien résiduel entre la variable présent et la variable du passé d'ordre k, une fois enlevée l'explication produite par les variables du passé intermédiaire.
Donc, quand à partir d'une valeur k donnée le corrélogramme Φ est nul, c'est qu'il est inutile de rechercher plus loin dans le passé une explication linéaire du présent.

5.2.4 Relations entre corrélogrammes ρ et Φ.

Il existe des relations entre le corrélogramme ρ et le spectrogramme d'une série stationnaire. Le calcul d'un corrélogramme peut ainsi être plus rapide via la f.f.t. (le passage s'effectue par des formules de convolution).

On pourrait ensuite faire le calcul de Φ, sur la base des équations linéaires suivantes.

$$X_t = \phi_{11}\, X_{t-1} + R_{t,1}\ ,\ \forall\ t$$

D'où,

$$X_t\, X_{t-1} = \phi_{11}\, X_{t-1}\, X_{t-1} + R_{t,1}\ X_{t-1},\ \forall\ t$$

$$E(X_t\, X_{t-1}) = \phi_{11}\, E(X_{t-1}\, X_{t-1}) + E(R_{t,1}\ X_{t-1}),\ \forall\ t$$

$$\mathrm{cov}(X_t, X_{t-1}) = \phi_{11}\, \sigma_X^2,\ \forall\ t$$

et

$$\phi_{11} = \rho_1$$

De même, à partir de l'équation: $X_t = \phi_{12}\, X_{t-1} + \phi_{22}\, X_{t-2} + R_{t,2}$

$$X_t\, X_{t-1} = \phi_{12}\, X_{t-1}\, X_{t-1} + \phi_{22}\, X_{t-2}\, X_{t-1} + R_{t,2}\, X_{t-1}$$
$$X_t\, X_{t-2} = \phi_{12}\, X_{t-1}\, X_{t-2} + \phi_{22}\, X_{t-2}\, X_{t-2} + R_{t,2}\, X_{t-2}$$

$$E(X_t\, X_{t-1}) = \phi_{12}\, E(X_{t-1}\, X_{t-1}) + \phi_{22}\, E(X_{t-2}\, X_{t-1}) + E(R_{t,2}\, X_{t-1})$$
$$E(X_t\, X_{t-2}) = \phi_{12}\, E(X_{t-1}\, X_{t-2}) + \phi_{22}\, E(X_{t-2}\, X_{t-2}) + E(R_{t,2}\, X_{t-2})$$

$$\rho_1 = \phi_{12} + \rho_1\, \phi_{22}$$
$$\rho_2 = \rho_1\, \phi_{12} + \phi_{22}$$

D'où

$$\phi_{22} = (\rho_2 - \rho_1^2) / (1 - \rho_1^2).$$

En généralisant la démarche, à partir de

$$X_t = \phi_{1k}\, X_{t-1} + \phi_{2k}\, X_{t-2} + \ldots + \phi_{kk}\, X_{t-k} + R_{t,k}\ ,$$

on aboutit à un système de k équations linéaires à k inconnues ϕ_{1k}, ϕ_{2k}, ϕ_{kk}, nommées **équations de Yule-Walker**, pouvant permettre d'obtenir une estimation des inconnues à partir de l'estimation de la fonction ρ.

L'inconnue ϕ_{kk} vaut:

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & & & \rho_{k-3} & \rho_2 \\ \vdots & & \ddots & & & \\ \rho_{k-1} & & & & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & & \rho_{k-2} \\ & & & & & \\ \rho_{k-1} & & \cdots & & \cdots & 1 \end{vmatrix}}$$

5.3 Zéros d'un corrélogramme, bruit blanc

On dit qu'*un corrélogramme est nul quand toutes les corrélations sont nulles* (ρ_o **mis à part, qui vaut toujours 1**).

Ceci se produit notamment pour le processus bruit blanc.

On dit qu'un processus stationnaire centré est processus bruit blanc (est ***bruit blanc****) si et seulement si son corrélogramme ρ est nul.*
Alors son corrélogramme Φ l'est aussi.

Soit une série observée de taille n. On convient le plus souvent que le **corrélogramme estimé** à partir de cette série est **nul**, au coefficient de sécurité 95%, si chacune des valeurs calculées appartient à l'intervalle

$$[-2/\sqrt{n},\ 2/\sqrt{n}],$$

et si l'ensemble des valeurs est distribué au hasard autour de zéro.
Alors on dira que la **série est un bruit blanc.**

Le coefficient 2, pris pour 1.96, est issu d'une loi normale $N(0,\sqrt{n})$ qui est celle de l'estimateur d'un ρ_k (d'un ϕ_{kk}) sous l'hypothèse de nullité de ρ (de Φ). Il doit donc être remplacé par une autre valeur quand on veut changer la valeur du coefficient de sécurité.
Comme 5% = 1/20, en pratique on ne remet pas en cause la nullité d'un corrélogramme observé si environ 1 valeur sur 20 sort *peu* de l'intervalle de confiance, *et* si cet écart ne concerne pas les (disons 2, attachant ici encore plus d'attention à ce qui se passe pour la première que pour la deuxième) premières valeurs du corrélogramme.

Si ce faible écart se produit pour k = 1 (ou pour k = 2 ?), ou si l'ensemble des valeurs a une distribution autour de zéro visiblement très particulière, sans que ces valeurs ne sortent forcément de l'intervalle, ou si un peu trop de valeurs sortent un peu de l'intervalle, on dit que le corrélogramme est **presque nul.**
Un exemple de distribution particulière serait que les dix premières valeurs calculées soient toutes positives (et éventuellement toutes incluses dans la zone de nullité), au lieu d'offrir un désordre dans la succession des signes. C'est dire que, dans de tels cas, on soupçonne l'existence d'un signal faible mélangé au bruit.

On dit que le corrélogramme **s'annule à partir d'une valeur de k** (fixé), quand pour l'ensemble des valeurs ≥ k il se comporte comme un corrélogramme nul, et que cela n'aurait pu être dit pour une valeur de k plus petite.

Cependant, pour les MA(q) avec q≠0, l'intervalle de confiance associé à l'hypothèse nulle pour le corrélogramme ρ doit être pris plus large (confer § 6.2.2).

En pratique, on convient de faire le calcul des valeurs d'un corrélogramme pour les décalages k compris entre 1 et n/4. Cependant, si on le juge nécessaire, on peut prolonger le corrélogramme à des valeurs de k plus élevées, notamment quand la taille n est suffisamment grande pour que ces estimations supplémentaires restent raisonnables.

On peut simuler (engendrer successivement des valeurs soi-disant observées d') un bruit blanc, à partir d'un générateur de nombres au hasard. Dire cela ainsi, pourrait laisser entendre qu'un bruit blanc est distribué selon la loi uniforme. Ce n'est pas forcément cette loi, toute autre pouvant convenir, y compris des lois discrètes ou une distribution sans référence à une loi connue.

La notion de bruit blanc est indépendante de celle de loi de probabilité.

Un bruit blanc gaussien est un processus bruit blanc gaussien.

Par définition, lorsque le bruit blanc est gaussien, non seulement les valeurs successives sont d'autocorrélations nulles, comme l'implique la notion même de bruit blanc, mais sont ***indépendantes.***
Plus précisément, les variables aléatoires du processus gaussien bruit blanc sont mutuellement indépendantes.

Nous dirons qu'un processus $\{X_t\}$ est un processus stationnaire centré **engendré à partir d'un bruit blanc**, noté $\{\varepsilon_t\}$, si et seulement si on peut écrire

$$\forall\ t, \quad X_t = \Sigma_{0\ldots\infty}\ \psi_j\ \varepsilon_{t-j}$$

avec $\psi_o = 1$, $\psi_j \in \Re$, et

$$\Sigma_{0\ldots\infty}\ |\psi_j| < +\infty\ .$$

Tous les processus stationnaires considérés dans ce cours sont des processus qui peuvent être considérés comme engendrés à partir d'un bruit blanc, ce dernier pouvant alors être dit **processus des innovations**.

Théorème

Un processus stationnaire engendré à partir d'un bruit blanc gaussien est lui-même gaussien.

En pratique, cela signifiera qu'il suffira de tester la normalité des résidus issus d'un processus stationnaire pris pour modèle (ce modèle étant validé), pour rejeter ou non l'hypothèse de normalité de la série modélisée.

5.4 Taille de la série

On suppose le plus souvent que la durée d'observation de la série est suffisamment grande.

En effet, une série à tendance horizontale peut être une série qui croît quelques temps, puis se met à décroître, puis se remet à croître, puis reste à peu près constante avant de se remettre à croître… Si on observe une telle série pendant une période trop courte, où elle se trouve être constamment croissante, on conclurait que cette série est forcément non stationnaire, alors qu'en fait, observée sur une période plus importante, l'hypothèse de stationnarité est envisageable.
Quand sait-on que la série n'est pas trop courte? On ne sait pas répondre à cette question sans avoir des connaissances a priori sur les données.

On prend aussi une durée d'observation pas trop grande, en tout cas telle qu'il soit raisonnable de supposer que de mêmes lois régissent d'un bout à l'autre le phénomène étudié: par exemple en économie, il est exceptionnel de pouvoir envisager des périodes stables supérieures à dix ans.
Quand sait-on que la série est trop longue? Quand n est grand, alors que par exemple un modèle correct construit sur la première moitié des valeurs de la série, celle associée au passé le plus lointain, diffère trop du même modèle estimé à partir des valeurs observées de la deuxième moitié (ou ne permet plus d'obtenir un modèle correct).

Et si on veut avoir une technique de prévision souple, donnant plus de poids aux dernières valeurs observées, une façon de procéder est la suivante:
Bien qu'on dispose parfois de série de taille énorme, il arrive que l'on ne se serve que des (disons) 30 dernières valeurs observées pour estimer les paramètres d'un modèle, et, qu'à chaque observation nouvelle, on estime à nouveau les paramètres de ce modèle sur les 30 dernières observations, plutôt que d'user des formules de mise à jour.
Cette taille 30 donnée en exemple aurait pu être choisie parce que les autres tailles voisines s'avéreraient donner des résultats moins bons, une étude ayant été faite à ce sujet d'après un relevé d'erreurs de prévision.
Ce qui pourrait s'expliquer: une taille plus petite donnerait des estimations trop imprécises, une taille plus grande mènerait à une technique trop peu souple pour ce genre de données, l'optimum se situant entre les deux pour la taille 30. Rappelons que la comparaison des valeurs observées aux valeurs prévues peut être gérée selon une notion de coût attaché à l'erreur commise, proportionnelle aux écarts, aux carrés des écarts, ou selon une définition plus appropriée.

Cette question de taille est finalement bien délicate, dépend beaucoup du bon sens.

6 QUELQUES MODÈLES STATIONNAIRES

6.1 Modèles autorégressifs d'ordre p, AR(p)

6.1.1 Définition

On considère les processus $\{X_t\}$ dont les variables vérifient l'équation:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t, \quad \forall\, t \qquad (6.1.1)$$

où les coefficients réels ϕ_i sont fixés, avec $\phi_p \neq 0$ et où $\{\varepsilon_t\}$ est un processus bruit blanc donné.

On note $\Phi(B)$ le polynôme en B

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p.$$

L'équation (6.1.1) s'écrit aussi

$$\Phi(B)\,(X_t) = \varepsilon_t, \quad \forall\, t$$

Si toutes les racines du polynôme Φ sont supérieures à 1 en module, alors l'équation (6.1.1) définit un processus $\{X_t\}$ unique, ce processus est stationnaire et il appartient à la catégorie des modèles nommés AR(p).
p est l'**ordre** du modèle AR (Auto Régressif, régressions sur "soi-même", sur son propre passé).

La stationnarité du processus se vérifie à partir du théorème 5.1, en s'appuyant sur le fait qu'un bruit blanc est stationnaire par définition, et en décomposant en éléments simples la fraction rationnelle $1 / \Phi(B)$. Comme les racines sont supérieures à 1 en module, la C.N.S. de stationnarité du § 5.3 (processus stationnaire engendré à partir d'un bruit blanc) est en effet vérifiée.
Cela est une façon de dire que la notation suivante est sensée

$$X_t = [\Phi(B)]^{-1}\, \varepsilon_t, \quad \forall\, t,$$

et que le processus $\{X_t\}$ est engendré à partir du bruit blanc $\{\varepsilon_t\}$.

Dire que le processus AR(p) est complètement déterminé par l'équation (6.1.1) implique que toute l'explication (linéaire) du présent par le passé est contenue dans seulement p variables explicatives, celles du passé le plus proche.

La projection orthogonale dans L^2 de X_t, sur le sous-espace de Hilbert engendré par *toutes* les variables du passé par rapport à t, est ainsi égale à

$$\phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p},$$

tout ce qui échappe à cette explication étant du domaine du bruit ε_t, de l'imprévisible.

Une méthode de calcul d'un modèle AR(p) consiste à estimer le modèle de régression avec p variables explicatives, puis à vérifier si toutes les racines sont supérieures à 1 en module. Si cela a marché, on regarde si les résidus du modèle peuvent passer pour ceux d'un bruit blanc (si les résidus sont ceux d'un bruit blanc estimé ε_t, leur corrélogramme ρ doit être nul). Si oui, c'est qu'il s'est avéré possible d'expliquer tout ce que le passé peut expliquer du présent avec seulement les p variables du passé le plus proche, et que le modèle trouvé est acceptable, pourvu que p ne soit pas trop élevé par rapport à n .

6.1.2 Propriétés des AR

Soit donné le modèle AR(p) d'équation

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t, \ \forall\ t.$$

Corrélogramme ρ

On a, pour $k \in N^*$,

$$X_t X_{t-k} = \phi_1 X_{t-1} X_{t-k} + \phi_2 X_{t-2} X_{t-k} + \ldots + \phi_p X_{t-p} X_{t-k} + \varepsilon_t X_{t-k}$$
$$E(X_t X_{t-k}) = \phi_1 E(X_{t-1} X_{t-k}) + \phi_2 E(X_{t-2} X_{t-k}) + \ldots + \phi_p E(X_{t-p} X_{t-k}) + E(\varepsilon_t X_{t-k}).$$

D'où ce système d'équations aux inconnues ρ_k, $k \in N^*$ (ρ_o=1):

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}, \quad k \in N^*. \qquad (6.1.2)$$

Notons x_i les p racines du polynômes Φ, qui sont de module > 1. Les solutions du système linéaire *récurrent* précédent sont de la forme, en notant a_i les coefficients réels,

$$\rho_k = a_1 (1/x_1)^k + a_2 (1/x_2)^k + \ldots + a_p (1/x_p)^k , \quad k \in N^*.$$

Quand une racine x_i est réelle, le terme correspondant, $a_i (1/x_i)^k$, est une exponentielle, qui décroît vers zéro quand k s'accroît.

Quand une racine x_i n'est pas réelle, sa conjuguée x_j est aussi une racine. Le terme correspondant à ces racines conjuguées, $a_i (1/x_i)^k + a_j (1/x_j)^k$, peut aussi s'écrire $a\, b^k \sin(2\pi f_k + \alpha)$, avec $|b|<1$, et décroît vers zéro selon une sinusoïde exponentiellement amortie quand k s'accroît.

Ainsi le corrélogramme ρ d'un processus AR est un corrélogramme dont les valeurs absolues diminuent, jusqu'à devenir nulles.

En pratique, lorsque la taille n de la série est trop faible, on ne pourra pas toujours avoir suffisamment de valeurs calculées pour vérifier que le zéro est effectivement atteint.

Corrélogramme Φ

Regardons l'expression d'un ϕ_{kk} en fonction des ρ_i, donnée au § 5.2.4 et écrivons-la pour k=p+1.

Si on regarde aussi les p+1 premières égalités du système 6.1.2, on constate qu'elles définissent l'écriture de la dépendance linéaire de la dernière colonne du déterminant du numérateur de $\phi_{p+1,p+1}$, en fonction

des autres colonnes. Cette valeur est donc nulle, et, pour les mêmes raisons sont nulles les valeurs de ϕ_{kk} pour les valeurs de $k > p+1$.

L'ordre p d'un AR(p) *peut donc se lire sur le corrélogramme* ***Φ*** *des autocorrélations partielles, ce corrélogramme étant nul à partir de k=p+1.*

Supposant la stationnarité, c'est ainsi que l'on identifie un AR(p).

Estimations des paramètres du modèle

Il existe plusieurs méthodes d'estimation. Citons la plus simple, celle de Yule-Walker, qui n'est pas la plus précise, mais qui est peu coûteuse en temps calcul. Elle est basée sur la résolution du système linéaire des p équations (dites de Yule-Walker) à p inconnues ϕ_1, ϕ_2, $\ldots\phi_p$, constitué des p premières égalités du système 6.1.2, sachant que les valeurs ρ_i sont les valeurs estimées.

La méthode MLE (maximum likehood estimator) ne converge pas toujours.

La méthode de Burg conduit souvent à des estimateurs de prévision qui seront les plus faibles, bien que le nombre des estimateurs des paramètres paraît souvent excessif au premier abord.

6.2 Modèles moyennes mobiles d'ordre q, MA(q)

On pourrait a priori se contenter des AR(p). Cependant, d'une façon générale, on sait que l'estimation simultanée des p paramètres d'un modèle, faite à partir de n observations, est d'autant plus fiable (c'est à dire correspond à des estimateurs de variance d'autant plus faible), que le nombre de paramètres est faible (même si cela doit être pondéré par des considérations sur l'indépendance des estimateurs).

Donc, quand l'ordre de l'AR est grand relativement à la taille de la série (disons, pour donner une idée, supérieur ou égal à 5 quand n=100 alors que le bruit serait important), on est en difficulté quant à la précision des estimateurs.

On souhaite assez naturellement pouvoir expliquer le présent en remontant plus loin dans le passé, tout en ayant peu de paramètres à estimer. Les processus MA (Moving Average) sont une façon d'y arriver.

6.2.1 Définition

On considère les processus $\{X_t\}$ dont les variables vérifient l'équation:

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}, \quad \forall\ t \qquad (6.2.1)$$

où les coefficients réels θ_i sont fixés, avec $\theta_q \neq 0$ et où $\{\varepsilon_t\}$ est un processus bruit blanc donné.

On note $\theta(B)$ le polynôme en B

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q.$$

L'équation (6.2.1) s'écrit alors:

$$X_t = \theta(B)\ \varepsilon_t, \quad \forall\ t$$

Ces processus sont stationnaires (processus engendré à partir d'un bruit blanc, § 5.3).

Si toutes les racines du polynôme θ sont supérieures à 1 en module, alors l'équation (6.2.1) définit un processus $\{X_t\}$ unique, ce processus est dit **inversible** et il appartient à la catégorie des modèles nommés MA(q).

q est l'**ordre** du modèle MA.

Le fait que les racines de θ soient supérieures à 1 en module permet d'écrire l'équation du modèle sous la forme (inverse):

$$\theta(B)^{-1} X_t = \varepsilon_t, \quad \forall\ t$$

C'est donc une façon de définir un AR d'ordre infini, si on développe $\theta(B)^{-1}$ sous la forme d'un développement polynomial infini. Le fait que θ ait ses racines supérieures à 1 permet alors d'avoir les termes du passé lointain négligeables et d'avoir une explication du présent par le passé qui pourrait être approchée d'aussi près que l'on veut en prenant un nombre suffisant de termes du passé le plus proche.

6.2.2 Propriétés

Fonction d'autocorrélation ρ

$$\sigma_X^2 = E(X_t X_t) = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q})(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q})]$$
$$= E(\varepsilon_t^2 + \theta_1^2 \varepsilon_{t-1}^2 + \theta_2^2 \varepsilon_{t-2}^2 + \ldots + \theta_q^2 \varepsilon_{t-q}^2),$$

puisque $E(\varepsilon_t \varepsilon_{t'}) = 0$ quand $t \neq t'$. Ainsi:

$$\sigma_X^2 = \sigma\varepsilon^2 (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2).$$

$$\mathrm{cov}(X_t, X_{t-1}) = E(X_t X_{t-1})$$
$$= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-3} - \ldots - \theta_q \varepsilon_{t-q-1})]$$
$$= E(-\theta_1 \varepsilon_{t-1}^2 + \theta_2\theta_1 \varepsilon_{t-2}^2 + \theta_3 \theta_2 \varepsilon_{t-3}^2 \ldots + \theta_{q-1}\theta_q \varepsilon_{t-q}^2).$$
$$= \sigma\varepsilon^2 (-\theta_1 + \theta_1\theta_2 + \ldots + \theta_{q-1}\theta_q).$$

D'où

$$\rho_1 = (-\theta_1 + \theta_1\theta_2 + \ldots + \theta_{q-1}\theta_q) / (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2).$$

En considérant $\mathrm{cov}(X_t, X_{t-2})$, on obtient de façon analogue

$$\rho_2 = (-\theta_2 + \theta_1\theta_3 + \ldots + \theta_{q-2}\theta_q) / (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2).$$

$$\rho_k = (-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q) / (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2), \qquad \text{pour } 0 \leq k \leq q.$$

En particulier,

$$\rho_q = -\theta_q / (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)$$

et de plus,

$$\rho_k = 0, \quad \forall\ k > q.$$

*Donc l'**ordre q d'un MA(q)** se lit sur le corrélogramme ρ des autocorrélations, se comportant comme un corrélogramme nul à partir des décalages supérieurs ou égaux à q+1, et non nul en deçà.*
C'est ainsi que l'on peut identifier un MA(q).

On démontre que l'**intervalle de confiance** associé à l'hypothèse nulle d'une valeur ρ_k, au coefficient de sécurité 95%, doit être pris non pas égal à $\pm 2/\sqrt{n}$, mais égal à

$$\pm 2(1+2\rho_1^2+2\rho_2^2+ \ldots+2\rho_q^2)^{1/2} / \sqrt{n}.$$

Ceci n'est pas toujours mis en œuvre dans les logiciels.
Il est assez rare en pratique d'envisager un ordre q plus grand que 2 (sauf pour des phénomènes physiques).

Fonction d'autocorrélation partielle Φ

Elle est dominée par des exponentielles décroissantes vers 0 quand k s'accroît, ou/et des mouvements sinusoïdaux exponentiellement amortis, son allure est de même type que celle du corrélogramme ρ d'un AR.

En pratique, lorsque la taille n de la série est trop faible, on ne pourra pas toujours avoir suffisamment de valeurs calculées pour vérifier que le zéro est effectivement atteint.

Par exemple, pour un MA(1), vérifiant par définition $|\theta_1| < 1$, on a:

$$\phi_{kk} = -\theta_1^k (1+\theta_1^2) / 1- \theta_1^{2(k+1)}.$$

Pour les estimations des paramètres du modèle MA, il existe plusieurs méthodes, dont celle de Cleveland (1972) basée sur la résolution d'un système linéaire. Elles permettent d'éviter le calcul des ϕ_{kk} à partir des ρ_i, dont la mise en œuvre nécessiterait un algorithme particulier complexe pour chaque valeur de k, ce qui est impraticable.

6.3 Modèles d'ordre (p,q): ARMA(p,q) mixtes

Avec les modèles MA il est donc possible de fabriquer des "AR d'ordre infini", tout en ayant peu de paramètres à estimer. Cette famille de modèles peut être encore étendue avec les modèles mixtes, à la fois autorégressifs et moyennes mobiles.
Souvent on s'intéresse à cette possibilité de modélisation quand celle par AR ou MA a été jugée peu satisfaisante ou impossible (corrélogramme des résidus non nul).

6.3.1 Définition

On considère les processus $\{X_t\}$ dont les variables vérifient l'équation:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}, \ \forall\, t \qquad (6.3.1)$$

où les coefficients réels ϕ_i et θ_i sont fixés, avec $\phi_p \neq 0$ et $\theta_q \neq 0$, et où $\{\varepsilon_t\}$ est un processus bruit blanc donné.

On note $\Phi(B)$ et $\theta(B)$ les polynômes en B

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p.$$
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q.$$

L'équation (6.3.1) s'écrit

$$\Phi(B)\ (X_t) = \theta(B)\ \varepsilon_t, \quad \forall\ t$$

Si les racines des polynômes Φ et θ sont supérieures à 1 en module, si aucune des racines de Φ n'est égale à une racine de θ, alors l'équation (6.3.1) définit un processus $\{X_t\}$ unique, ce processus est stationnaire et inversible, et il appartient à la catégorie des modèles nommés ARMA(p,q).

6.3.2 Corrélogrammes de l'ARMA mixte.

Les corrélogrammes ρ et ϕ d'un ARMA(p,q) se comportent plutôt comme le corrélogramme ρ d'un AR et le corrélogramme ϕ d'un MA. Quand k s'accroît, ils décroissent vers zéro selon un mélange d'exponentielles.

Ils ne permettent donc pas d'identifier l'ordre (p,q) *d'un ARMA.*

En particulier, quand un corrélogramme observé ρ s'annule à partir de q_1+1, alors que dans le même temps le corrélogramme ϕ s'annule à partir de p_1+1, cela signifie qu'on peut envisager comme modèles soit un AR(p_1), soit un MA(q_1), *soit aussi un ARMA(p,q) avec* $0<p \leq p_1$ ***et*** $0<q \leq q_1$.

6.4 Identification du modèle mixte stationnaire

L'identification de l'ordre (p,q) du modèle mixte est délicate dans la mesure où les méthodes à mettre en œuvre sont imprécises, ou deviennent si coûteuses quand la précision est recherchée, qu'elles ne sont pas toujours portées sur les logiciels.
Citons la méthode dite "**du coin**", où l'ordre est indiqué par le nombre de lignes et de colonnes délimitant, dans un tableau de valeurs de déterminants, un coin supérieur gauche, en dehors duquel toutes les déterminants calculés sont sensiblement nuls.

Une autre approche consiste à ne s'intéresser aux modèles mixtes qu'après avoir décidé qu'il ne peut s'agir ni d'un AR ni d'un MA. On modélise alors la série par un pseudo AR(p) et les résidus obtenus (non bruit blanc) par un MA(q), q étant cette fois déterminé à l'aide du corrélogramme ρ des nouveaux résidus. Si p est le bon ordre, alors q est aussi le bon. On entame la procédure avec p = 1.

Quand la taille de la série n'est pas très élevée, le nombre de paramètres égal à p+q ne peut être bien grand. De plus il est rare que q soit pris supérieur à 2. Le nombre des ordres à envisager dans cette procédure est donc le plus souvent plutôt réduit.

Approche interactive ou approche automatique: étude dans la revue Statistique et Analyse des données, Juin 1990, Vol.15, n°1, p.61.

Une estimation précise des paramètres d'un ARMA mixte est celle par la **méthode du maximum de vraisemblance** (MLE, maximum likehood estimator), utilisable aussi pour les AR et MA. Cette méthode assez lourde n'est pas toujours celle à mettre en œuvre, car elle suppose à priori le processus gaussien. Mais elle est indépendante des initialisations faites sur le bruit.
L'hypothèse gaussienne doit donc être vérifiée à posteriori sur le bruit blanc estimé.

Si le processus est gaussien, si n est grand, et l'ordre de l'ARMA étant fixé, on montre que la recherche des estimateurs des paramètres par MLE coïncide avec la recherche du modèle ayant une variance des résidus minimale.

Cela veut dire que, les corrélogrammes des bruits blancs étant nuls, les histogrammes pouvant passer pour ceux d'échantillons gaussiens, il suffit alors de comparer les variances estimées des bruits blancs pour faire son choix entre plusieurs modèles d'un ordre donné et estimés par MLE (soit entre modèles ARp, soit entre modèles MAq, soit entre modèles ARMA(p,q)).
Cette remarque présente un intérêt dans la mesure où la procédure MLE est initialisée à partir d'une première évaluation des paramètres faite en résolvant un système linéaire, ou obtenue a partir d'une autre méthode d'estimation préliminaire de ces paramètres. La solution ensuite obtenue est souvent un optimum local, dépendant de cette initialisation (comme souvent dans les procédures MLE).

On peut aussi converger vers des valeurs de paramètres non admissibles. Le modèle en cours est alors rejeté. Pour une initiation aux techniques d'estimation de modèle ARMA, lire Mélard (1989).

Un AR(0) ou un MA(0), sont des notations permettant de désigner un bruit blanc comme cas limite d'AR ou de MA, et un AR(1) ou un MA(1) avec une valeur du paramètre proche de zéro, sont des processus proches du processus bruit blanc.
C'est ainsi qu'un ARMA(0,q) est aussi une façon de désigner un MA(q), et qu'un ARMA(p,0) est une façon de désigner un AR(p).

Finalement, avec cette convention, les *ARMA désignent la classe de tous les modèles stationnaires* considérés dans ce cours, et les *ARMA mixtes désignent la sous-classe des ARMA(p,q) avec* $p \neq 0$ *et* $q \neq 0$.

7 MODÈLES NON STATIONNAIRES

7.1 Introduction

Un ARMA n'est plus un ARMA si les racines du polynôme Φ ne sont pas toutes supérieures à 1 en module, la stationnarité n'étant plus vérifiée.
Ainsi par exemple l'équation

$$(1 - 2B)\, Y_t = \varepsilon_t,$$

correspond à un polynôme Φ de degré 1, ayant pour racine 1/2. Cette équation, où $\{\varepsilon_t\}$ est bruit blanc, définit un processus $\{Y_t\}$ non stationnaire.

A chaque étape on double la valeur précédente, au bruit blanc près:

$$y_1 = 2y_0 + \varepsilon_1$$
$$y_2 = 2y_1 + \varepsilon_2$$
$$\dots$$

Même si on initialise par une valeur $y_0 = 0$, il arrivera, avec une probabilité 1, si on n'impose pas de limite dans le temps, que la valeur doublée soit grande (en valeur absolue) de sorte que, les valeurs doublées suivantes étant encore plus grandes, le bruit en devienne très rapidement négligeable. Le résultat en sera une croissance ou une décroissance exponentielle avec une probabilité nulle ou presque nulle de retournement de cette tendance. Le bruit devient alors tellement insignifiant au bout d'un certain temps, qu'un modèle déterministe exponentiel décrit très bien le comportement monotone.

Nous n'irons pas si loin dans la non stationnarité, voulant proposer des modèles stochastiques, où le bruit conserve son importance.

On a déjà signalé qu'une croissance (stricte), ou une décroissance, était un symptôme de non stationnarité. Par exemple, considérons la succession de valeurs régie par

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

On remarque que

$$z_t = y_t - y_{t-1} = at + b - a(t-1) - b + x_t - x_{t-1} = a + x_t - x_{t-1}.$$

Ce qui signifie que le processus $\{Z_t\} = \{(1-B)\, Y_t\}$ serait un processus stationnaire, si $\{X_t\}$ était lui-même stationnaire.

Considérons cette fois un polynôme de degré 2 et non plus de degré 1:

$$y_t = at^2 + bt + c + x_t, \quad \forall\ t.$$

$$z_t = y_t - y_{t-1} = at^2 + bt + c - a(t-1)^2 - b(t-1) - c + x_t - x_{t-1} = 2at - a + b + x_t - x_{t-1}$$

Ce qui signifie que le processus $\{Z_t - Z_{t-1}\} = \{(1-B)^2\, Y_t\}$ serait stationnaire, si $\{X_t\}$ l'était.
Les polynômes $(1-B)$ ou $(1-B)^2$ sont des polynômes en B ayant une racine égale à 1.

Il apparaît donc concevable de pouvoir se ramener (souvent) au cas stationnaire en ne travaillant pas sur la série initiale mais sur des séries obtenues par différences successives (par **différenciations**).

Cela pourra être vraie aussi dans le cas où il n'y a pas de monotonie à considérer, la non stationnarité pouvant exister aussi quand la tendance est horizontale (moyenne du processus constante).

7.2 Modèles ARIMA

On considère les processus $\{X_t\}$ dont les variables vérifient l'équation:

$$\Phi(B)\ (1-B)^d\ X_t = \theta(B)\ \varepsilon_t, \quad \forall\ t \qquad (6.2)$$

avec $\Phi(B)$ et $\theta(B)$ les polynômes en B,

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q,$$

et d ($\neq 0$) l'ordre de multiplicité de la racine 1 du polynôme $(1-B)^d$.

Les coefficients réels ϕ_i et θ_i sont fixés et $\{\varepsilon_t\}$ est un processus bruit blanc donné.

Si les racines des polynômes Φ et θ sont supérieures à 1 en module, si aucune des racines de Φ n'est égale à une racine de θ, alors l'équation (6.2) définit un processus $\{X_t\}$ unique, ce processus est non stationnaire, et il appartient à la catégorie des modèles nommés ARIMA(p,d,q).
(I comme "integrated", fonction réciproque de "differenciated").

(p,d,q) est l'**ordre** de l'ARIMA.

Les corrélogrammes ρ et ϕ d'un ARIMA n'existent pas, et le périodogramme non plus.

La famille ARIMA désigne parfois la classe de tous les modèles, stationnaires et non stationnaires, en convenant que les ARIMA(p,0,q) sont les ARMA(p,q).
On parle aussi d'ARI(p,d) ou d'IMA(d,q)...

7.3 Identification du nombre d de différenciations

Une approche fréquemment pratiquée consiste à considérer les séries successivement différenciées $(1-B)^d$ pour d=0,1,2,3,4, et 5, en estimant chaque fois la variance. La liste des variances possède un minimum, ce minimum correspond à une valeur d=k qui est souvent celle qui conviendra; le nombre maximum de différenciations considérées pris égal à 5 suffit le plus souvent.
Si on n'arrive pas à modéliser avec l'hypothèse d=k précédente, on envisage la possibilité que la valeur de d associée à la variance la plus proche de la plus petite variance puisse mieux convenir. En partant d'une variance basse on a "plus de chance" d'aboutir à des erreurs de prévision faibles, et ceci explique la méthode sans la justifier. En effet, il arrive que les premières valeurs de d ainsi considérées ne soient pas les bonnes.
Il est assez exceptionnel d'avoir une valeur de d supérieure à 3.

Les hésitations sur le choix de d peuvent aussi provenir de ce que deux modèles associés à deux valeurs différentes de d sont en fait associés à des processus voisins, comme le montre par exemple les deux équations suivantes:
En notant P(B) un polynôme de degré p qui a ses p racines supérieures à 1 en module, avec les équations

$$P(B)\ (1-B)^2\ X_t = \theta(B)\ \varepsilon_t \quad \text{et} \quad P(B)\ (1-.97B)\ (1-B)\ Y_t = \theta(B)\ \varepsilon_t,$$

on définit un processus $\{X_t\}$ ARIMA(p, 2, q) et un processus très voisin $\{Y_t\}$ ARIMA(p+1, 1, q).

Dans un tel cas, il est conseillé de choisir le modèle "le moins stationnaire", c'est à dire le premier, avec la valeur de d la plus élevée. En effet, ce genre de voisinage induit parfois des difficultés dans l'estimation des paramètres du second processus, notamment pour une procédure MLE.

Des hésitations sur le choix de d peuvent aussi provenir de ce qu'aucun modèle ARIMA ne peut vraiment bien convenir à la série étudiée.

Une fois la valeur de d fixée, on recherche un processus $\{Z_t\} = \{(1-B)^d X_t\}$, ARMA(p,q), ajustée à la série $(1-B)^d x_t$.

7.4 Mouvement tendanciel déterministe: exemple de traitement.

Le praticien peut savoir d'avance que la tendance est de loi donnée. Il peut s'agir alors d'estimer les paramètres de la loi connue à priori. Ensuite, la série détendancialisée pourra être modélisée (éventuellement) par AR(I)MA.
On peut désirer une méthode résistante d'estimation des paramètres de la loi à la méthode des moindres carrés non pondérés, d'autant que le système de pondération, réitéré et évalué chaque fois d'après les résidus obtenus, ne fournit pas toujours une procédure convergente.
Ainsi, la méthode ci-dessous dite "des 3 points", de Tukey, est une méthode intéressante.

Supposons que le mouvement tendanciel soit régi par la courbe de **Gompertz** (β étant positif entre zero et 1, décroissance asymptotique vers e^γ si $\alpha > 0$, ou croissance en S, forte au début et moins forte ensuite, si $\alpha < 0$, $|\alpha|$ assez grand; courbes amusantes quand β est négatif et t entier):

$$y = \exp(\alpha\beta^t + \gamma)$$

soit encore

$$z = \mathrm{Log}(y) = \alpha\beta^t + \gamma$$

On découpe les données $(t_i, z_i)_{i=1,\dots n}$ en trois paquets j=1,2,3, chacun de temps moyen $\bar{t}_j$ et de taille égale τ ($\tau = n/3$, si n est multiple de 3; si $n=3\tau+1$, le paquet du milieu est pris de taille $\tau+1$; si $n=3\tau+2$, les paquets extrêmes sont pris de taille $\tau+1$). On calcule les 3 médianes des valeurs z_i des 3 paquets: m_1, m_2, m_3. Résoudre alors le système des 3 équations à 3 inconnus:

$$m_j = \alpha\ \beta^{\bar{t}_j} + \gamma, \quad j = 1,2,3.$$

On a en effet $\bar{t}_2 - \bar{t}_1 = \bar{t}_3 - \bar{t}_2 = \delta$ et $\mathrm{Log}(\beta) = \mathrm{Log}[(m_3-m_2)/(m_2-m_1)] / \delta$.
D'où β, puis α, puis γ.

8 CRITÈRES DE CHOIX, OU DE REJET, DE MODÈLES ARMA

Une série observée qui a une tendance non horizontale, ou une variabilité monotone croissante ou décroissante, est à priori une série non stationnaire: il est inutile de regarder les corrélogrammes de ces séries, et le calcul de la moyenne n'a aucun sens. Si la tendance n'est pas déterministe, c'est la technique des différenciations qui est adoptée.

On suppose maintenant que l'observation du graphe de la série, ou une connaissance à priori de propriétés de cette série, ne permettent pas d'affirmer que les hypothèses de moyenne constante et de variance constante seraient à rejeter.
La série est peut-être stationnaire, peut-être pas.
Il faut différencier aussi et regarder la liste des variances des séries différenciées 0, 1, 2, …, 5 fois. La série initiale est-elle celle dont la variance est la plus basse? Si oui, il est probablement maladroit de différencier.
Dans tous les cas, en différenciant ou non, le but est d'être ramené au cas d'une série susceptible d'être stationnaire.
Le problème d'identification consiste alors à trouver le processus stationnaire de la classe ARMA qui s'ajusterait le mieux à ces données.
On cherche à identifier un ou plusieurs modèles en s'attachant à ce que soient satisfaits les critères suivants, le premier étant celui à satisfaire d'abord, car c'est le critère d'acceptation ou de rejet du modèle.

- *corrélogramme nul des résidus*

Soit on se contente d'apprécier la nullité en observant le corrélogramme, soit en faisant le **Q-test** nommé aussi test "**portemanteau**". Ce test n'est pas idéal, ne faisant pas intervenir les signes des valeurs ρ_k. Donc on attache dans le même temps de l'importance à l'observation du corrélogramme lui-même.

$$\hat{Q} = (n-d)(n-d-2)\sum_{i=1}^{K}(n-d-i)^{-1}\rho_i^2$$

K-(p+q) (K moins le nombre de paramètres; pour un SARIMA ce sera K-(p+q+P+Q)) est le nombre de degré de liberté de l'estimateur qui suit un χ^2_{K-p-q}.

On rejette **l'hypothèse d'indépendance des erreurs** (plus forte que celle de nullité du corrélogramme) si la valeur est trop grande. Le plus souvent, le test est répété pour différentes valeurs de K, de façon à vérifier s'il n'a pas donné une réponse trop particulière pour une seule valeur de K arbitrairement choisie. Dans Splus, ce rejet correspond à la situation où les points sont sous la ligne en pointillé représentant les valeurs limites au coefficient de sécurité 95%. Le premier test, celui qui englobe le moins de premières corrélations du corrélogramme, est à 1 degré de liberté.

Quand deux valeurs résiduelles sont nettement plus importantes que toutes les autres, elles induisent un ρ_i relativement important, i représentant la distance entre les 2 valeurs résiduelles. Ainsi donc un modèle résistant pourrait conduire au rejet du modèle alors qu'un modèle peu résistant n'aurait pas conduit à une telle décision.

- *variance faible des résidus.*

Ce n'est pas un critère d'acceptation ou de rejet du modèle: que le bruit soit important par rapport au signal est un phénomène qui est inhérent aux données, et la démarche visant à obtenir la plus petite variance ne coïncide pas totalement avec celle visant à obtenir un corrélogramme nul ou avec celle visant à proposer un modèle pertinent (sachant ce que l'on a à priori dans la tête au sujet des données).

C'est plutôt un critère de choix entre modèles n'ayant pu être départagés à partir de l'observation des corrélogrammes des résidus ou le test précédent.

Ce critère, comme le AIC et le BIC ci-dessous, peut être utilisé pour comparer des modèles produits à partir d'une même méthode d'estimation des paramètres. Ils deviennent douteux quand les méthodes diffèrent.

En particulier, quand on use d'une méthode résistante, le calcul dilate le plus souvent les résidus les plus importants déjà repérés avec une autre méthode non résistante. Un calcul de variance étant très sensible aux plus grands écarts mènera donc très souvent à une variance plus élevée pour les résidus de la méthode résistante, par construction. En toute logique, dans ce cas les 2 méthodes sont peu comparables, visant deux objectifs différents, et les écarts à la médiane pourraient être utilisés pour comparer entre elles des méthodes résistantes.

- Le critère d'Akaïke (basé sur la quantité d'information de Kullback) pondère le jugement de comparaison entre modèles par la variance des résidus, en faisant intervenir notamment le nombre de paramètres du modèle, compte tenu de la taille de la série. C'est le plus usité.

$$AIC = n \, Log(\sigma_\varepsilon^2) + 2n(p+q+1+\delta) / (n-d).$$

où $\delta = 1$ si $d=0$, et $=0$ si $d>0$.

Ou le BIC, proposé aussi par Akaïke, et associé à un estimateur convergent,

$$BIC = Log(\sigma_\varepsilon^2) + (p+q) \, Log(n-d) / (n-d),$$

ou encore la variante

$$Log(\sigma_\varepsilon^2) + (p+q) \, c \, Log(Log(n-d)) / (n-d), \quad \text{avec } c > 2.$$

- *nombre de paramètres peu élevé*

Ce souhait est lié aux variances des estimateurs de ces paramètres préférées faibles, et à la variance des résidus, souhaitée faible aussi. La notion du nombre de paramètres peu élevé est à considérer relativement à la taille de la série (les critères précédents peuvent permettre des comparaisons entre modèles construits sur des séries de taille différentes).

C'est un critère important de choix entre modèles. Ce peut être un critère de rejet de modèles.

Par exemple, un AR(7) peut sembler satisfaisant en usant des critères précédents; mais estimé sur une série de taille 60 il peut être mis fortement en doute; et tout le nécessaire doit être fait pour proposer plutôt une modélisation avec moins de paramètres. Certaines méthodes d'estimations plus fiables que d'autres peuvent cependant conduire à admettre un nombre d'estimateurs plus élevé.

- *autres indicateurs*

Quand on dispose des variances des estimateurs des paramètres (on souhaite ces variances faibles pour une bonne précision des estimations), on dispose aussi des corrélations linéaires entre les paramètres (souvent on les souhaite souvent plutôt faibles, pour avoir des estimateurs plutôt indépendants entre eux) et tout cela peut aussi aider à départager les modèles et/ou les méthodes d'estimation des paramètres du modèle.

Cependant le plus souvent variances et corrélations ne vont pas tous dans un même sens en faveur d'un seul modèle, et bien des hypothèses devraient être vraiment vérifiées pour s'y fier totalement (normalité, stationnarité).

Un autre repère consiste à se référer à la variance des estimateurs de prévision. Sans y voir un défaut de conception, on constate que la méthode d'estimation des paramètres de Burg (pour les AR) conduit souvent à un nombre de paramètres très important (bien que le critère d'Akaïke soit utilisé pour déterminer ce nombre), alors que dans le même temps les variances des estimateurs de prévision sont plus faibles pour les petits horizons que les variances des estimateurs associés aux autres méthodes d'estimation des paramètres qui utilisent moins de paramètres.

- *pour conclure*

Une difficulté de l'identification provient de ce qu'un processus est toujours voisin d'autres processus, stationnaires ou non, alors que vouloir satisfaire en même temps aux 3 critères est rarement possible. C'est dire que rarement on se trouve devant une situation nette, et on peut hésiter entre plusieurs modèles, voisins ou non. La situation évidemment s'aggrave lorsqu'aucun modèle ne semble tout à fait satisfaisant, parce qu'il n'en existe pas vraiment, les corrélogrammes des résidus de différentes modélisations étant presque nuls, plutôt que nuls.

Les procédures automatiques d'identification du modèle masquent les choix faits et ne conduisent pas toujours à la solution que l'utilisateur aurait préférée: cela n'est pas important s'il n'est pas très important d'obtenir toujours "la" "meilleure" prévision possible.

Exemples de modèles stationnaires voisins

$$X_t = .5\, X_{t-1} - .1\, X_{t-2} + \varepsilon_t, \ \forall\, t \qquad \text{et} \qquad Y_t = .5\, Y_{t-1} + \varepsilon_t, \ \forall\, t$$

$$X_t = \varepsilon_t - .75\, \varepsilon_{t-1} + .12\, \varepsilon_{t-2}, \ \forall\, t \qquad \text{et} \qquad Y_t = \varepsilon_t - .75\, \varepsilon_{t-1}, \ \forall\, t$$

$$X_t = .5\, X_{t-1} - .25\, X_{t-2} + .13\, X_{t-3} + \varepsilon_t, \ \forall\, t \qquad \text{et} \qquad Y_t = \varepsilon_t + .5\, \varepsilon_{t-1}, \ \forall\, t$$

9 ÉQUATION POUR LA POSTVISION

Considérons une équation générale définissant un processus ARMA mixte:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}, \ \forall\, t.$$

En particulier, la variable aléatoire du présent n peut s'exprimer à partir du passé, au bruit ε_n près (appelé aussi processus des innovations puisqu'il constitue la seule nouveauté entre le temps n et les temps du passé par rapport à n), par l'équation:

$$X_n = \phi_1 X_{n-1} + \phi_2 X_{n-2} + \ldots + \phi_p X_{n-p} + \varepsilon_n - \theta_1 \varepsilon_{n-1} - \theta_2 \varepsilon_{n-2} - \ldots - \theta_q \varepsilon_{n-q}.$$

Comme la loi du processus est totalement définie, cette loi régit aussi le futur, et par exemple,

$$X_{n+1} = \phi_1 X_n + \phi_2 X_{n-1} + \ldots + \phi_p X_{n-p+1} + \varepsilon_{n+1} - \theta_1 \varepsilon_n - \theta_2 \varepsilon_{n-1} - \ldots - \theta_q \varepsilon_{n-q+1}$$

est l'équation exprimant la variable du futur (horizon 1), en fonction du présent et du passé. Nous avons ainsi des équations qui permettront de construire un calcul de prévisions.

En fait, dans le même temps, nous disposons aussi des **équations de postvision**. C'est à dire les équations de prévisions quand le sens du temps est inversé. Il suffit pour cela, dans l'équation d'un modèle ARMA ou ARIMA, de remplacer l'opérateur retard B par l'opérateur avance F (F pour Forward, avec $F = B^{-1}$):

$$\Phi(F)\,(1-F)^d X_t = \theta(F)\,\eta_t, \quad \forall\, t \qquad \text{avec } F(X_t) = X_{t+1}\ , \ \forall\, t$$

On a cependant changé le nom du bruit blanc, car il ne peut être identifié à celui des équations de prévision, même si ses propriétés sont inchangées (comme "la" loi si une identification à une loi connue a été faite). Notamment on a $\sigma_\eta^2 = \sigma_\varepsilon^2$.

On a donc, pour un ARMA,

$$X_t = \phi_1 X_{t+1} + \phi_2 X_{t+2} + \ldots + \phi_p X_{t+p} + \eta_t - \theta_1 \eta_{t+1} - \theta_2 \eta_{t+2} - \ldots - \theta_q \eta_{t+q}, \ \forall\, t$$

En particulier, si X_1 est la variable aléatoire associée à la première observation, X_0 est celle associée à la postvision d'horizon 1 et on a

$$X_0 = \phi_1 X_1 + \phi_2 X_2 + \ldots + \phi_p X_p + \eta_0 - \theta_1 \eta_1 - \theta_2 \eta_2 - \ldots - \theta_q \eta_q$$

L'estimation des paramètres du modèle ne dépend pas du sens choisi pour le temps.

En pratique, on fait de la postvision avec le modèle de prévision trouvé, mais appliquée à la série dont le sens du temps a été inversé, c'est à dire à la série des valeurs: $y_t = x_{n+1-t}$, t = 1, 2, …, n.

10 ESTIMATION DES RÉSIDUS (Box et Jenkins)

On estime le bruit blanc à partir de toute l'information, et à partir du modèle stationnaire calculé, sur la série ou sur la série différenciée.

On estime chacune des valeurs du résidu par l'espérance conditionnelle sachant que $X_1 = x_1$, $X_2 = x_2$, …, et $X_n = x_n$, et on note ces espérances avec "[ ]":

$$\varepsilon_t = [\mathcal{E}_t] = E(\mathcal{E}_t \mid X_1 = x_1, X_2 = x_2, \ldots, \text{et } X_n = x_n) , \quad \forall\ t.$$

Supposons que le modèle choisi soit l'ARMA(1,1):

$$X_t - .3\ X_{t-1} = \mathcal{E}_t - .7\ \mathcal{E}_{t-1} , \ \forall\ t.$$

On a donc,

$$\mathcal{E}_t = X_t - .3\ X_{t-1} + .7\ \mathcal{E}_{t-1} , \ \forall\ t.$$

D'où les relations entre espérances conditionnelles:

$$[\mathcal{E}_t] = [X_t] - .3\ [X_{t-1}] + .7\ [\mathcal{E}_{t-1}] , \ \forall\ t.$$

Ainsi

$$\varepsilon_2 = x_2 - .3\ x_1 + .7\ [\mathcal{E}_1]$$
$$\varepsilon_3 = x_3 - .3\ x_2 + .7\ \varepsilon_2$$
$$\varepsilon_4 = x_4 - .3\ x_3 + .7\ \varepsilon_3$$
$$\ldots$$
$$\varepsilon_n = x_n - .3\ x_{n-1} + .7\ \varepsilon_{n-1}$$

Si $[\mathcal{E}_1]$ était connu, alors de proche en proche on pourrait en déduire les estimations du résidus pour t allant de 2 à n. Mais cela n'est pas le cas et initialiser le calcul en posant par exemple $[\mathcal{E}_1] = 0$, reviendrait à commettre une erreur qui se répercute de proche en proche. Cependant, ces erreurs diminuent de proche en proche, jusqu'à devenir négligeables assez rapidement (ergodicité du modèle).

Ne serait-ce que pour avoir une meilleure estimation de la variance du bruit blanc, on adopte plutôt la méthode qui suit, mise en œuvre sur une série d'école, avec le modèle précédent supposé lui convenir.
On se sert d'abord du modèle en mode postvision pour estimer les résidus η_t en initialisant au temps n de façon arbitraire, mais sachant que l'erreur s'atténue fortement de proche en proche, jusqu'à devenir négligeable pour les dernières valeurs calculées η_2, η_1.
On en déduit les postvisions x_0, x_{-1}, x_{-2}, …, jusqu'à $x_{-s} \cong 0$.

En effet, *la prévision à long terme donne toujours la moyenne avec un ARMA*.

On termine le calcul par les estimations ε_t, en utilisant le modèle en mode prévision pour t= -s+1, -s+2, …, n.

Ainsi, le modèle de postvision s'écrivant

$$X_t - .3\, X_{t+1} = \eta_t - .7\, \eta_{t+1}\ ,\ \forall\ t$$

on a

$$[\eta_t] = [X_t] - .3\, [X_{t+1}] + .7\, [\eta_{t+1}]\ ,\ \forall\ t.$$

En initialisant par $[\eta_{12}] = 0$ (on se réfère au fait que $E(\eta_t) = 0$ pour faire ce choix plutôt qu'un autre), on calcule η_{11}, η_{10}, …

t	1	2	3	4	5	6	7	8	9	10	11	12
x_t	1.10	.56	-.24	.17	-.32	.21	-.20	.08	.22	-.17	.02	.21
η_t	1.26	.47	-.24	.08	-.27	.16	-.15	.10	.13	-.21	-.04	0

Sachant que

$$[X_t] = [\eta_t] + .3\, [X_{t+1}] - .7\, [\eta_{t+1}]\ ,\ \forall\ t$$

et que $[\eta_t] = 0$ pour t = 0, -1, …, car *on suppose l'indépendance* des variables η_t (ou le bruit est gaussien), on a les estimations

$x_0 = 0 + .3*1.1 - .7*1.26 = -.55$
$x_{-1} = -.17$; $x_{-2} = -.05$; $x_{-3} = -.01$, $x_{-4} = -.00$ (s=3).

Avec l'équation: $[\varepsilon_t] = [X_t] - .3\, [X_{t-1}] + .7\, [\varepsilon_{t-1}]\ ,\ \forall\ t,$

ayant posé au départ $[\varepsilon_{-4}] = 0$, on obtient pour t = -3, puis -2, …:

t	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12
ε_t	1.10	.56	-.24	.17	-.32	.21	-.20	.08	.22	-.17	.02	.21				

L'erreur commise ainsi au départ, en posant $[\eta_{12}] = 0$, a finalement une influence négligeable sur les estimations ε_t, $\forall$ t = 1, 2,…, n.

L'estimation de la variance du bruit blanc est alors faite avec les s+n valeurs estimées (en divisant par n et non par n+s).

Quand la série est de grande taille, cette démarche peut être évitée, et l'estimation de la variance du bruit peut se faire sur série tronquée.

11 PRÉVISION

11.1 Définitions

On considère ici les processus $\{Y_t\}$ ARMA (mixtes ou non), *modèles stationnaires* de séries initiales ou de séries différenciées.

Pour un instant présent t, y_t étant la dernière valeur observée, on veut prévoir la valeur y_{t+h} du processus stationnaire $\{Y_t\}$.
Y_{t+h} est la variable du processus associée à la vraie valeur non connue y_{t+h} et $\hat{\mathbf{Y}}_t(h)$ est l'estimateur associé à la prévision cherchée $\hat{y}_t(h)$, faite au temps t et à l'horizon h.
Le processus $\{Y_t\}$ est engendré par un bruit blanc ε_t et en particulier,

$$Y_{t+h} = \sum_{j=0\ldots\infty} \psi_j \varepsilon_{t+h-j}.$$

Pour la prévision faite au temps t, on recherche un estimateur engendré par le même bruit blanc, et s'écrivant aussi en fonction du passé jusqu'à t:

$$\hat{\mathbf{Y}}_t(h) = \sum_{j=0\ldots\infty} \alpha_{h+j} \varepsilon_{t-j}.$$

On détermine cet estimateur en calculant les réels α_{h+j} de façon à minimiser dans L^2 la distance entre Y_{t+h} et $\hat{\mathbf{Y}}_t(h)$

$$\begin{aligned} E[(Y_{t+h} - \hat{\mathbf{Y}}_t(h))^2] &= E[(\sum_{0\ldots\infty} \psi_j \varepsilon_{t+h-j} - \sum_{0\ldots\infty} \alpha_{h+j} \varepsilon_{t-j})^2] \\ &= E[(\varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \ldots + \psi_{h-1} \varepsilon_{t+1} + \sum_{0\ldots\infty} (\psi_{h+j} - \alpha_{h+j}) \varepsilon_{t-j})^2] \\ &= \sigma_\varepsilon^2 (1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{h-1}^2) + E[(\sum_{0\ldots\infty} (\psi_{h+j} - \alpha_{h+j}) \varepsilon_{t-j})^2]. \end{aligned}$$

Le minimum est atteint pour $\alpha_{h+j} = \psi_{h+j}$ $\forall$ j $\in$ N. (ψ_0=1).

D'où la **définition de l'estimateur**:

$$\hat{\mathbf{Y}}_t(h) = \sum_{0\ldots\infty} \psi_{h+j} \varepsilon_{t-j}.$$

Si le bruit blanc est gaussien, l'estimateur est donc aussi gaussien.

La **définition de l'estimation** est ainsi**:**

$$\hat{y}_t(h) = \psi_h \varepsilon_t + \psi_{h+1} \varepsilon_{t-1} + \psi_{h+2} \varepsilon_{t-2} + \ldots$$

Pour un horizon h, la variable aléatoire représentant l'**erreur de prévision** est:

$$e_t(h) = Y_{t+h} - \hat{\mathbf{Y}}_t(h) = \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \ldots + \psi_{h-1} \varepsilon_{t+1}.$$

$E[e_t(h)] = 0$. L'estimateur $\hat{\mathbf{Y}}_t(h)$ est sans biais.

$$\text{var}[e_t(h)] = \sigma_\varepsilon^2 (1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{h-1}^2).$$

La précision des estimations est donc d'autant plus forte que la variance du bruit blanc est faible. Et cette précision diminue (assez naturellement) quand on veut prévoir plus loin dans le futur.

Dans le cas particulier où le processus $\{\mathcal{E}_t\}$ est gaussien, et pour un coefficient de sécurité de 99%, $e_t(h)$, somme finie de variables gaussiennes non corrélées, suit la loi gaussienne $N(0, \sigma_e^2)$.
Donc, $P(|Y_{t+h} - \hat{\mathbf{Y}}_t(h)| < 2{,}6\ \sigma_e) = .99$ et l'intervalle de confiance autour de l'estimation d'horizon h est

$$\hat{y}_t(h) \pm 2{,}6\ \sigma_\varepsilon\ (1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{h-1}^2)^{1/2}.$$

11.2 Calcul

L'espérance conditionnelle, conditionnée par la connaissance de tout le passé du processus dont on connaît l'équation, y compris à l'instant présent t, est notée

$$E_t(Y_{t+h}) = E(Y_{t+h} / Y_t=y_t, Y_{t-1}=y_{t-1}, Y_{t-2}=y_{t-2}, \ldots) = E(Y_{t+h} / \mathcal{E}_t=\varepsilon_t, \mathcal{E}_{t-1}=\varepsilon_{t-1}, \mathcal{E}_{t-2}=\varepsilon_{t-2}, \ldots).$$

Donc, **si on suppose l'indépendance mutuelle** des variables $\mathcal{E}_t$ et non plus seulement les autocorrélations nulles, on obtient

$$E_t(Y_{t+h}) = E_t(\hat{\mathbf{Y}}_t(h) + e_t(h)) = \psi_h\, \varepsilon_t + \psi_{h+1}\, \varepsilon_{t-1} + \psi_{h+2}\, \varepsilon_{t-2} + \ldots$$

Autrement dit, on obtient la **formule générale du calcul des prévisions d'un ARIMA**

$$\hat{y}_t(h) = E_t(Y_{t+h}).$$

Par exemple, pour un AR(1) d'équation $Y_t = .7\ Y_{t-1} + \mathcal{E}_t$, on a:

$$\hat{y}_t(1) = E_t(Y_{t+1}) = E_t(.7\ Y_t + \mathcal{E}_{t+1}) = .7\ E_t(Y_t) + E_t(\mathcal{E}_{t+1}) = .7\ y_t.$$

En effet $E(Y_t / Y_t=y_t) = y_t$, dernière valeur observée.
Vue l'indépendance supposée, on a aussi: $E_t(\mathcal{E}_{t+1}) = E(\mathcal{E}_{t+1}) = 0$.
Par suite,

$$\hat{y}_t(2) = E_t(Y_{t+2}) = E_t(.7\ Y_{t+1} + \mathcal{E}_{t+2}) = .7\ E_t(Y_{t+1}) + E_t(\mathcal{E}_{t+2}) = .7\ \hat{y}_t(1).$$

$$\hat{y}_t(h) = .7\ \hat{y}_t(h-1) \text{ pour } h > 1.$$

Pour un IMA(1,1) d'équation $Y_t = Y_{t-1} - .5\ \mathcal{E}_{t-1} + \mathcal{E}_t$,

$$\hat{y}_t(1) = E_t(Y_{t+1}) = E_t(Y_t + .5\ \mathcal{E}_t + \mathcal{E}_{t+1}) = E_t(Y_t) + .5\ E_t(\mathcal{E}_t) + E_t(\mathcal{E}_{t+1}) = y_t + .5\ \varepsilon_t.$$

$$\hat{y}_t(2) = E_t(Y_{t+2}) = E_t(Y_{t+1} + .5\ \mathcal{E}_{t+1} + \mathcal{E}_{t+2}) = E_t(Y_{t+1}) = \hat{y}_t(1).$$

$$\hat{y}_t(h) = \hat{y}_t(h-1) \text{ pour } h > 1.$$

L'indépendance est donc indispensable au calcul, qui se fait en la supposant.

Lorsque le **bruit blanc est gaussien** - ce qui se contrôle graphiquement ou par test, sur les résidus - alors **cette indépendance supposée est vérifiée** (normalité + non corrélations induisant l'indépendance).
Sinon, on ne sait rien de cette hypothèse d'indépendance, et il peut arriver qu'elle ne soit pas vérifiée. La formule de calcul serait alors fausse, malheureusement.

Remarque:
Les logiciels donnent les valeurs des paramètres d'un ARMA centré. On obtient la vraie équation du processus non centré associé à une série de moyenne non nulle m, en remplaçant dans l'équation des modèles Y_t par $Y_t + m$, *quelque soit t* . On évalue ainsi la constante du modèle. Cette question ne se pose pas pour un ARIMA où $d > 0$.
Une valeur prévue avec un processus centré à laquelle on ajoute la moyenne donne l'estimation recherchée.

11.3 Formules de mise à jour

La vraie valeur y_{t+1} ayant été observée, il s'agit de mettre à jour les prévisions précédemment faites au temps t, de façon à prendre en compte cette information supplémentaire.

$$e_t(1) = \varepsilon_{t+1}, \forall\ t.$$

Donc les erreurs de prévisions à l'horizon 1 sont non corrélées entre elles, et supposées indépendantes. Ce n'est pas le cas des prévisions aux horizons supérieurs.

Cela explique, quand on compare les résultats de calculs de prévision issus de différentes méthodes pour une évaluation de l'efficacité de ces méthodes, que les mesures (variances, somme d'écarts, etc.) soient basées uniquement sur les erreurs de prévision à l'horizon 1. On parle ainsi de méthodes de **prévision à court terme**, la qualité d'une méthode s'évaluant d'après les erreurs de prévision à l'horizon 1.

Cette identité indique pourquoi $\{\varepsilon_t\}$ est appelé **processus des innovations**. Le bruit blanc étant processus d'erreurs de prévision dans ces modèles, il est associable à ce qui est imprévisible, innovant.

Cette identité nous donne la **mise à jour du bruit blanc**, auparavant estimée jusqu'au temps t:

$$\varepsilon_{t+1} = y_{t+1} - \hat{y}_t(1).$$

Pour la mise à jour des prévisions, puisque l'on a:

$$\hat{\mathbf{Y}}_{t+1}(h) = \psi_h \varepsilon_{t+1} + \psi_{h+1} \varepsilon_t + \psi_{h+2} \varepsilon_{t-1} + \ldots.$$

et $$\hat{\mathbf{Y}}_t(h+1) = \psi_{h+1} \varepsilon_t + \psi_{h+2} \varepsilon_{t-1} + \psi_{h+2} \varepsilon_{t-1} + \ldots.$$

on a donc

$$\hat{\mathbf{Y}}_{t+1}(h) = \hat{\mathbf{Y}}_t(h+1) + \psi_h \varepsilon_{t+1}.$$

Ainsi la correction à apporter aux anciennes prévisions pour avoir les nouvelles ne dépend que de l'innovation et correspond à la formule de **mise à jour des prévisions**:

$$\hat{y}_{t+1}(h) = \hat{y}_t(h+1) + \psi_h \varepsilon_{t+1}.$$

Calcul des coefficients ψ_h

$\psi_0 = 1$. Pour un ARIMA d'équation générale

$$\Phi(B)\ (1-B)^d\ X_t = \theta(B)\ \varepsilon_t,$$

le processus stationnaire défini par $Y_t = (1-B)^d\ X_t$ pouvant s'écrire

$$Y_t = \sum_{0...\infty}\ \psi_j\ \varepsilon_{t-j},$$

on a:

$$\Phi(B)\ (\sum_{0...\infty}\ \psi_j\ \varepsilon_{t-j}) = \theta(B)\ \varepsilon_t,\ \forall\ t.$$

En usant de l'écriture polynomiale, l'égalité entre polynômes en B qui s'en déduit,

$$\Phi(B)\ (\sum_{0...\infty}\ \psi_j\ B^j) = \theta(B),$$

permet, par identification des termes de mêmes degré et en commençant par les degrés les plus faibles, de connaître la valeur de ψ_1, puis de ψ_2, etc.

Par exemple, si le modèle est

$$(1 - .85\ B + 1.56\ B^2)\ (1 - B)\ X_t = (1 - .7\ B)\ \varepsilon_t,$$

et si on a besoin des valeurs ψ_1 et ψ_2 seulement, on écrit

$$(1 - .85B + 1.56\ B^2)\ (1 + \psi_1\ B^1 + \psi_2\ B^2 + \ldots)\ =\ (1 - .7\ B).$$

Soit:

$$\begin{array}{llll} 1 & - .85\ B & + 1.56\ B^2 & \\ & + \psi_1\ B & - .85\ \psi_1\ B^2 & + 1.56\ \psi_1\ B^3 \\ & & + \psi_2\ B^2 & - .85\ \psi_2\ B^3 + 1.56\ \psi_2\ B^4 \end{array}$$

$$= 1 + (\psi_1 - .85)B + (\psi_2 - .85\ \psi_1 + 1.56)B^2 + \ldots \quad = 1 - .7\ B$$

D'où: $\psi_1 - .85 = -.7$ et $\psi_1 = -1.55$; $\psi_2 - .85\ \psi_1 + 1.56 = 0$ et $\psi_2 = ?$

L'amiral LARIMA
Larima quoi
La rime à rien
L'amiral LARIMA
L'amiral RIEN
Jacques Prévert

12 MODÈLES SAISONNIERS SARIMA

12.1 Introduction

La méthode de décomposition d'une série en différentes composantes tendancielles, saisonnières, calendaires, cycliques, … permet souvent un très bon ajustement aux observations. Cependant elle n'est pas conçue en vue d'une prévision et l'extrapolation d'un tel modèle dans le futur n'est pas le plus souvent la meilleure approche pour prévoir.

Quand le saisonnier est absent, le modèle de décomposition additive ($x_t = z_t + s_t + e_t$, $\forall$ t) s'écrit

$$x_t = z_t + e_t , \quad \forall \text{ t.}$$

Un processus ARIMA peut être recherché, adapté aux observations x_t.
Mais s'il est reconnu pour cette série que le mouvement tendanciel z_t est déterministe de loi connue, pour prévoir mieux vaut prolonger cette loi analytique dans le futur et ajouter aux valeurs de la tendance ainsi prévues les prévisions issues d'un ARIMA construit sur la série e_t.
Quand le saisonnier est présent, on peut de même se proposer de faire la prévision en prolongeant dans le futur le saisonnier modélisé de la décomposition additive, si on juge que c'est une attitude raisonnable pour la série étudiée (dans le cas le plus simple, on aurait vérifié que l'hypothèse des coefficients saisonniers constants est tout à fait justifiée et peut être prolongée dans le futur). On serait alors ramené au cas précédent.

Mais si on ne peut justifier la décomposition additive (multiplicative, …), ou si on ne peut extrapoler de façon évidente les valeurs des coefficients saisonniers, il est souhaitable de disposer d'une méthode de prévision intégrant le phénomène saisonnier (et intégrant ou non le tendanciel).

12.2 Définition

Pour faciliter l'exposé, on suppose que la série y_t est constituée de valeurs mensuelles de période 12.

On pourrait chercher à modéliser l'évolution des seuls mois de Janvier par un ARIMA(P, D, Q).
Puis celle des mois de Février par un autre ARIMA, etc. En fait, on modélise l'évolution *d'année en année* par un seul processus $\{X_t\}$, en traduisant le comportement moyen des 12 ARIMA par un seul pseudo-ARIMA, sans contraindre le résidu α_t de ce processus commun à être bruit blanc.
L'équation générale de cette évolution est:

$$\Phi_a(B^{12})\ (1-B^{12})^D\ X_t = \theta_a(B^{12})\ \alpha_t, \quad \forall \text{ t}$$

L'évolution *de mois en mois* non traduite dans le modèle précédent est modélisée par un ARIMA(p, d, q) construit sur la série résiduelle α_t:

$$\Phi_m(B)\ (1-B)^d\ \alpha_t = \theta_m(B)\ \varepsilon_t, \quad \forall \text{ t}$$

On montre alors que le report de α_t ainsi défini, dans l'équation précédente, aboutit à l'équation générale du processus $\{Y_t\}$ modélisant la série y_t:

$$\Phi_a(B^{12})\,\Phi_m(B)\,(1-B)^d\,(1-B^{12})^D\,Y_t = \theta_a(B^{12})\,\theta_m(B)\,\varepsilon_t, \quad \forall\ t.$$

Cette équation définit la classe de modèles nommés **SARIMA d'ordre (p, d, q) x (P, D, Q)$_{12}$**

ε_t est bruit blanc. Les polynômes en B, Φ_m et θ_m, et en B^{12}, Φ_a et θ_a, ont leurs racines supérieures à 1 en module. Une racine de polynôme Φ_m et Φ_a ne doit pas être égale à une racine des polynômes respectifs θ_m et θ_a.

On désigne aussi cette classe comme celle des **ARIMA multiplicatifs**, car l'équation générale est en fait presque l'équation d'un ARIMA, particulier parce que dans la liste des paramètres coefficients des différentes puissances de B certains coefficients sont nuls, et parce qu'il peut exister des racines multiples 1 de polynômes en B^{12}.

Par exemple, un SARIMA $(1,0,1)x(2,1,0)_{12}$ est défini par l'équation:

$$(1 - \phi_{m1}\,B)(1 - \phi_{a1}\,B^{12} - \phi_{a2}\,B^{24})(1 - B^{12})\,Y_t = (1 - \theta_{m1}\,B)\,\varepsilon_t$$

Soit

$$[1 - \phi_{m1}\,B - (\phi_{a1} + 1)\,B^{12} + \phi_{m1}\,(\phi_{a1} + 1)\,B^{13} - (\phi_{a2} - \phi_{a1})\,B^{24} + \phi_{m1}\,(\phi_{a2} - \phi_{a1})\,B^{25} + \phi_{a2}\,B^{36} - \phi_{m1}\,\phi_{a2}\,B^{37}]Y_t = (1 - \theta_{m1}\,B)\,\varepsilon_t.$$

Ce qui pourrait passer pour l'équation d'un ARIMA(37, 0, 1), mais 4 paramètres seulement sont à estimer.

Il y a donc ici peu de choses à ajouter sur ce qui a été dit sur les ARIMA eux-mêmes et, en pratique, c'est surtout l'identification de l'ordre du modèle qui est un nouveau problème.

12.3 Identification

On cherche d'abord la stationnarité par différenciations. Une méthode souvent utilisée consiste à repérer la valeur du couple (d, D) associée à la plus faible variance des séries $(1-B)^d\,(1-B^{12})^D\,y_t$ calculées pour $(d, D) \in \{1,2,3,4,5\}x\{1,2,3\}$ par exemple. Le meilleur couple sera celui-là ou un couple proche de celui-là.
La valeur du couple (d, D) étant choisie, il s'agit alors de trouver un SARMA$(p,q)x(P, Q)_{12}$ pour la série différenciée.
Box et Jenkins proposent une démarche intéressante, mais qui n'est pas facile à mettre en œuvre sur tous les logiciels (c'était plus facile par exemple avec le vieux logiciel TSA de la NAG, qu'avec le plus moderne logiciel Splus).
On peut aussi s'aider du corrélogramme ρ.

Ainsi par exemple, pour un SARMA$(0,1)_x(0, 1)_{12}$, les ρ_k sont non nuls pour k=1, 11, 12 et 13 et nuls pour les autres décalages.
L'identification de l'ordre d'un ARIMA n'étant pas chose si aisée, on comprend que du temps soit nécessaire pour l'identification du "bon SARIMA", quelque soit la démarche préconisée.

Il existe de nombreux outils graphiques de description du saisonnier. A partir de là il est parfois possible de remarquer que les mois pourraient se répartir disons en 2 classes. Par exemple, une classe de mois ayant un certain type de corrélogrammes, et une seconde d'un autre type.
Au lieu de rechercher un SARMA pour la série de période 12, peut être vaut-il mieux alors envisager un SARMA de période T pour la sous-série des T mois de la première classe et un autre SARMA de période 12-T pour la sous-série des 12-T autres mois.
En production, souvent le mois d'août est à part, et la série à modéliser est celle des 11 autres mois, de période 11.

La démarche empirique suivante sur série différenciée est envisageable, sans prétendre être la plus judicieuse. Elle exige une estimation des résidus pour toute les valeurs de temps considéré, sans troncature.

* Observer graphiquement les 12 sous-séries mensuelles en privilégiant les outils d'observation simultanée et observer leurs corrélogrammes (opérer éventuellement une classification).

* Se proposer un ordre (P, Q) commun et estimer les 12 ARMA issus de cette hypothèse. Estimer l'ARMA commun en prenant la moyenne ou la médiane des 12 valeurs trouvées pour chaque paramètre. Calculer les 12 sous-séries résiduelles issues de cet ARMA commun.

* Rechercher un ordre (p,q) pour la série résiduelle reconstituée à partir des 12 sous-séries résiduelles précédentes.

* Estimer les paramètres du SARIMA $(p,d,q)_x(P,D,Q)_{12}$ alors considéré pour la série initiale. Observer le corrélogramme ρ des nouveaux résidus.

* Comparer les résultats de cette modélisation aux résultats issus de modèles ayant des hypothèses de quadruplés (P,Q,p,q) proches du quadruplé obtenu, en faisant varier un seul terme à la fois du quadruplé, et en tenant compte de ce qui est observé sur le corrélogramme précédent, pour réduire les valeurs ρ_k pouvant sortir de la zone de confiance du zéro.

Toute procédure automatique d'identification est ici intéressante, si on la joint à cette dernière étape, sachant qu'en pratique la variance minimale des résidus (ou un critère AIC) est plus souvent mise en œuvre (vue la facilité de cette mise en œuvre) que la recherche d'un corrélogramme nul.

Les logiciels exigent que les valeurs p+q et P+Q soient chacune au moins égale à 1. Dans le cas où une de ces valeurs pourraient être envisagée comme nulle, Box et Jenkins conseillent de choisir la partie MA non nulle plutôt que la partie AR.

13 TRANSFORMATIONS ET PREVISIONS

On suppose que la série initiale est $\{x_t\}$ et que c'est la série définie par $y_t = f(x_t)$, où f est fonction monotone, qui a été modélisée par un processus $\{Y_t\}$ permettant les estimations $\hat{y}_t(h)$. La fonction f peut avoir été introduite parce que la stationnarité ne pouvait être obtenue par différenciation des données initiales $\{x_t\}$.

Alors, $f^{-1}(\hat{y}_t(h))$ n'est pas la bonne estimation de x_{t+h}.
Cela est dû au calcul par les espérances, ici conditionnelles:

$$E_t(X_{t+h}) = E_t[f^{-1}(Y_{t+h})] \neq f^{-1}[E_t(Y_{t+h})].$$

Tout dépend de la concavité de la fonction f^{-1}: si cette concavité est tournée vers les ordonnées positives, $f^{-1}[E_t(Y_{t+h})]$ est une sous-estimation de la valeur cherchée. Si elle est tournée vers les ordonnées négatives, c'est une surestimation. Si f est linéaire, c'est la bonne estimation.

Par exemple, supposons que f=Log et f^{-1} = exp. $X_t = \exp(Y_t)$. Supposons que Y soit normale, de moyenne μ et de variance σ^2.
Alors X = exp(Y) est log-normale, de moyenne $m = \exp(\mu + \sigma^2 / 2)$.
Ici:

$$\mu = E_t(Y_{t+h}) = \hat{y}_t(h)$$

$$\sigma^2 = \mathrm{var}_t(Y_{t+h}) = \mathrm{var}[e_t(h)] = \sigma_\varepsilon^2(1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{h-1}^2).$$

Finalement la formule de prévision devient:

$$E_t(X_{t+h}) = E_t[\exp(Y_{t+h})] = \exp[\ \hat{y}_t(h) + \sigma_\varepsilon^2\ (1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{h-1}^2)/2].$$

Avec la classe des transformations de BOX-COX (§ 2.2), le paramètre λ de la transformation est estimé (par MLE ou par moindres carrés) en même temps que les paramètres de l'ARMA, de sorte que le choix de la transformation dans cette classe est optimisé (variance minimale des résidus).

14 GÉNÉRALISATIONS

Quand il s'agit, dans la généralisation, du passage d'une à plusieurs séries, en plus de prendre le temps de décrire une à une chacune des séries, il faut ensuite aussi user d'outils descriptifs, comme des **corrélogrammes croisés et spectrogramme croisé**, pour décrire les liaisons entre séries prises deux à deux. Ainsi, entre une série $x_{1,t}$ et une série $x_{2,t}$ on cherche à savoir les corrélations existantes entre l'une et le passé de l'autre, entre l'une et le futur de l'autre (on ne sait pas toujours d'avance laquelle peut influer sur l'autre), entre le présent de l'une et le présent de l'autre.
C'est à ce prix que l'on pourra prétendre pouvoir aborder la complexité introduite par la généralisation.

14.1 ARMA multivarié: introduction élémentaire

Ce qui a été fait pour une série est généralisable à plusieurs séries considérées simultanément, de façon à prendre en compte les liaisons entre ces séries.
En pratique, il semble bien que cette approche ait pour le moment ses limites, et qu'il soit peu envisageable de considérer plus de 5 séries simultanément, dans le cas d'un AR, et moins dans le cas général. On peut se tourner notamment vers des techniques d'analyse des données pour réduire le nombre des séries ou extraire les séries les plus pertinentes et la littérature est abondante sur les traitements multivariés. On doit toujours se demander si le traitement global d'une quantité importante d'information constitue vraiment le jeu de données à considérer ou s'il traduit plutôt une incapacité (provisoire) à se poser de bonnes questions. Dans la plupart des situations de ce genre, l'information relative à une question pertinente est fortement redondante.

Ce qui est le plus abordable, car contrôlable à l'aide d'outils comme le corrélogramme croisé ou le spectrogramme croisé, est le cas de 2 séries seulement.
On trouvera bien des précisions sur les ARMA multivariés dans Gourieroux et Monfort.

On considère maintenant seulement un processus bivarié $\{X_t\} = \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix}$.

Il est centré: $E(X_t) = 0$, $\forall$ t. C'est à dire $E(X_{1,t}) = 0$ et $E(X_{2,t}) = 0$.

On pose $\Gamma_{t+h,t} = E(X_{t+h}\, X_t) = E\left(\begin{bmatrix} X_{1,t+h} \\ X_{2,t+h} \end{bmatrix} \begin{bmatrix} X_{1,t} & X_{2,t} \end{bmatrix}\right) = \begin{pmatrix} E(X_{1,t+h}X_{1,t}) & E(X_{1,t+h}X_{2,t}) \\ E(X_{2,t+h}X_{1,t}) & E(X_{2,t+h}X_{2,t}) \end{pmatrix}$.

Un bruit blanc $\{\varepsilon_t\}$ bivarié est un processus stationnaire centré tel que $\Gamma_{t+h,t} = 0$ pour $h \neq 0$, et

$$\Gamma_{t,t} = \Gamma_\varepsilon = \begin{pmatrix} E(\varepsilon_{1,t}^{\,2}) & E(\varepsilon_{1,t}\varepsilon_{2,t}) \\ E(\varepsilon_{2,t}\varepsilon_{1,t}) & E(\varepsilon_{2,t}^{\,2}) \end{pmatrix}.$$

Un processus centré bivarié est stationnaire si $\Gamma_{t+h,t} = \Gamma_h$, $\forall$ t, $\forall$ h.

Il ne suffit donc pas que $\{X_{1,t}\}$ et $\{X_{2,t}\}$ soient stationnaires, il faut aussi que la liaison linéaire ne dépende que du décalage h entre les deux séries, indépendamment de t.

A étant matrice carrée de dimension 2, la norme de A, notée $\|A\|$, est la plus grande valeur propre de AA'.

Un processus bivarié engendré par un bruit blanc $X_t = \sum_{0...\infty} \Psi_j \varepsilon_{t-j}$, avec $\Psi_o = I$, Ψ_j matrices 2x2 telles que $\sum_{0...\infty} \|\Psi_j\| < +\infty$, *est stationnaire*.

L'équation générale d'un ARMA bivarié est

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \Theta_2 \varepsilon_{t-2} - \ldots - \Theta_q \varepsilon_{t-q}, \ \forall\, t$$

où Φ_i et Θ_i sont des matrices réelles 2x2, et ε_t est bruit blanc bivarié.

La matrice $\Phi(z) = I - \Phi_1 \begin{pmatrix} z & \\ & z \end{pmatrix} - \Phi_2 \begin{pmatrix} z^2 & \\ & z^2 \end{pmatrix} - \ldots - \Phi_p \begin{pmatrix} z^p & \\ & z^p \end{pmatrix}$ est une matrice 2x2 de polynômes en z (en B) de degré p (le plus grand trouvé).

Le déterminant $\det[\Phi(z)]$ est polynôme en z. Si les racines de ce polynôme sont supérieures à 1 en module, le processus bivarié $\{X_t\}$ est stationnaire…

On conçoit de façon analogue la notion d'inversibilité relative a la partie MA du processus.

p est l'ordre de la partie AR.
Cependant, certains auteurs précisent l'ordre d'un AR par la matrice carrée dont les termes sont les degrés des termes-polynômes de $\Phi(z)$; la somme des degrés de ces polynômes contitue alors le nombre de paramètres à estimer.

Exemple:
Avec

$$\Phi_1 = \begin{pmatrix} .5 & .5 \\ 0 & .5 \end{pmatrix} \text{ et } \Theta_1 = \begin{pmatrix} .5 & 0 \\ .5 & .5 \end{pmatrix},$$

l'équation

$$X_t - \Phi_1 X_{t-1} = \varepsilon_t - \Theta_1 \varepsilon_{t-1}$$

revient à écrire les 2 relations:

$$X_{1,t} = .5\, X_{1,t-1} + .5\, X_{2,t-1} + \varepsilon_{1,t} - .5\, \varepsilon_{1,t-1}$$

$$X_{2,t} = .5\, X_{2,t-1} + \varepsilon_{2,t} - .5\, \varepsilon_{1,t-1} - .5\, \varepsilon_{2,t-1}.$$

14.2 Modèles à représentation markovienne: introduction vraiment élémentaire

Ils sont plus généraux que les processus multivariés ARMA; on les appelle aussi modèles à espace d'états (cf. Gourieroux et Monfort).

$\{X_t\}$ est un processus k-multivarié associé aux observations, et $\{Z_t\}$ est un autre processus, associable à un ensemble de séries, appelées variables d'états, observables ou non.

L'équation générale est donnée par 2 relations; la première est l'expression du lien linéaire aux variables d'états, la seconde décrivant l'évolution temporelle de ces variables d'états. H et G sont des matrices réelles, $\{h_t\}$ et $\{g_t\}$ sont bruits blancs (indépendants) et h_t est indépendant de g_t:

$$X_t = H\, Z_t + h_t$$
$$Z_t = G\, Z_{t-1} + g_t$$

Par exemple, avec $H = [1,\ 0]$, $Z_t = \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix}$, $g_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$, la matrice $G = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}$, et avec $h_t = 0$, on reconstitue un AR(2) monovarié.

En fait, on montre que tout modèle à représentation markovienne où $h_t = 0$, est un ARMA (multivarié éventuellement). Mais un ARMA admet le plus souvent plusieurs représentations.

14.3 Modèles de transfert

$$Y_t = P(B)\, X_t + \varepsilon_t \quad \forall\, t\,, \quad X \text{ en entrée, } Y \text{ en sortie.}$$

L'**Armax** correspond à un modèle mixte, à la fois ARMA et modèle de transfert:

$$\Phi(B) X_t = D(B)\, Y_t + \Theta(B)\, \varepsilon_t \qquad \forall\, t$$

où Φ, D, Θ sont polynômes de degrés p, r et q, et ε_t bruit blanc

14.4 GARCH

Modélisation en cas d'hétéroscédasticité (hypothèse de variance non constante du processus).
CH comme hétéroscédasticité conditionnelle. G comme généralisé.
Permet de modéliser avec un bruit qui est processus ARCH.

14.5 ARH

Les points successifs d'une série sont remplacés par des morceaux de courbes successifs. Ici H comme Hilbert. Pour la prévision du morceau de courbe suivant.

14.6 Modèles économétriques, par exemple usant simultanément des techniques de prévision multivariée précédentes et des techniques explicatives de variables par d'autres variables décalées aussi dans le temps.

14.7 Modèles ARFIMA

Le F est donné pour "fractionnaire". Quand la tendance est linéaire, on sait que le choix d=1 suffit pour en faire la modèlisation. Quand la tendance est polynomiale de degré 2, d=2 suffit pour en faire la modélisation. Au lieu de d entier, on peut donc cherche à estimer une valeur de d réelle positive optimale pour traduire au mieux l'évolution à long terme de la série. C'est en général ce qui est visé par une modélisation ARFIMA.

14.8 Modèles non paramètriques

Ne sachant rien, je n'en dis rien. Mais leur efficacité n'est plus à démontrer.

14.9 Utilisation de l'approche "**Réseaux neuronaux**".

Etc. etc.

15 QUELQUES RÉFÉRENCES

Un très bon moyen de se procurer des références d'articles ou de livres à la base d'une technique programmée est de regarder celles citées dans les help ou dans les notices de logiciels. Par exemple, pour les séries, il existe une bonne bibliographie dans Systat-Sygraph. Cependant, quand le logiciel est anglais ou américain les références concernent uniquement (ou presque) des auteurs anglophones.

Anderson T.W. (1971) *The statistical analysis of Time Series,* Wiley, New York. (développé sur les tests).

Azencott R. et Dacunha-Castelle D. (1984) *Séries d'observations irrégulières.* Masson. Paris.

Barnett V. and Lewis T. (1978) *Outliers in Statistical data.* John Wiley & Sons Ltd.

Bensaber A. Bleuse-Trillon B (1989) *Pratique des chroniques et de la prévision à court terme*. Masson, coll. Méthodes et Programmes. (notamment méthodes Holt-Winters)

Bestougeff H., Ligozat G., (1989) *Outils logiques pour le traitement du temps. "de la linguistique à l'intelligence artificielle".* Masson.

Bloomfield P. (1976) *The Fourier analysis of Time Series: An introduction.* Wiley, New York.

Bosq D. Lecoutre J.P. (1992) *Analyse et prévision des séries chronologiques. Méthodes paramétriques et non paramétriques.* Masson. ISBN 2-225-83889-5. . Très intéressant

Bosq D. (1996) *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction.* Springerg-Verlag 26,rue des Carmes 75005 Paris. ISBN 0 387 94713 2.

Box G.E.P. and Cox D.R. (1964) *An analysis of transformations.* J.R.S.S. B 26 pp.211

Box G.E.P. and Jenkins G.M. (1976) *Time Series Analysis: Forecasting and Control,* San Francisco: Holden Day.

Brillinger D.R. (1975) *Time Series: Data Analysis and Theory.* Holt, Rinehart & Winston New York. (concerne surtout l'analyse harmonique).

Brillinger D.R. Krishnaiah P.R. (1991) *Time Series in the Frequency Domain.* North-Holland.

Brockwell P.J. and Davis (1987) *Time Series: Theory and Method.* Springer-Verlag.

Brown R. (1962) *Smoothing, forecasting and prediction,* Printice Hall.
Caperaa P., Van Cutsem B. (1988) *Méthodes et modèles en statistique non paramétrique.* Dunod. Paris.

Chambers J.M., Cleveland W.S., Kleiner B., Tukey P.A. (1983) *Graphic methods for data analysis.* Belmont CA:Wadsworth Internat. group.

Christensen R. (1990) *Linear Models for Multivariate, Time Series, and Spatial Data.* Springer-Verlag.

Cleveland W.S. (1979) *Robust locally weighted regression and smoothing scatterplots.* J.A.S.A. n°74, pp.829.

Cleveland W.S. (1981) *LOWESS: a program for smoothing scatterplots by robust locally weighted regression.* The American Statistican. pp.35

Cleveland W.S. IRMA J.T. (1982) *Graphical methods for seasonal adjustment.* J.A.S.A., Vol.77, n°377, theory and methods section.

Cleveland W.S. IRMA J.T. (1985) *The Elements of Graphing Data.* Wadsworth Advanced Books and Software.

Cleveland W.S. MacGill M.E.. (1988) *Dynamic Graphics for Statistics.* Wadsworth Advanced Books and Software.

Coley J.W., Tukey J.W. (1965) *An algorithm for the machine computation of complex Fourier series.* Math. Computation N°19, pp.297.

Coutrot B. et Droesbeke F. (1984) *Les méthodes de prévision.* coll. que sais-je? n° 2157 PUF.

Dacunha-Castelle D. , Duflo M. (1993) *Probabilités et Statistiques. Tome 2. Problèmes à temps mobile.* Masson. (2ème édit., martingales, chaîne de Markov, décisions pas à pas, processus de comptage, à temps continu,...)

David M., Michaud J.C. (1989) *La Prévision, approche empirique d'une méthode statistique.* Coll. INRA actualités scientifiques et agronomiques. Masson. (Sarima avec SAS ou BMDP)

Droesbeke J.J., Fichet B., Tassi P. (1989) *Séries chronologiques, théorie et pratique des modèles ARIMA.* Economica. (ASU)

J.J. Droesbeke, B. Fichet, P. Tassi (1994) *Modélisation ARCH. Théorie statistique et applications dans le domaine de la finance.* Ed. Ellipses. Ed. univ. Bruxelles.

Erickson B.H., Nosanchuck T.A. (1977) *Understanding Data: an introduction to exploratory and confirmatory data analysis for students in the social sciences.* Open University Press.

Faget de Casteljau P. (1990) *Le lissage.* éd. Hermès (Math. Appli.).
(lissage quadratique, algébrique, trigo, projectif ,...). bouquin plutôt théorique.

Gardner E.S. (1985) *Exponential smoothing: the state of the art.* Journal of forecasting n°4 pp.1.

Gouriéroux C. et Monfort A. (1983) *Cours de séries temporelles.* Economica, coll. économie et statistiques avancées (49 rue Héricart, 75015 Paris). Indispensable.

Guégan D. (1994) *Séries chronologiques non linéaires à temps discret.* Economica.

Härdle W. (1990) *Smoothing Techniques, with implementation in S.* Springer-Verlag.

Jolivet E., Legay J.M.,Tomassone R. (1980) *Biométrie et temps.* Société française de biométrie.

Laffargue J.P. (1981) *Interprétation économique des caractéristiques quasi-cycliques d'un modèle macroéconomique évoluant en environnement aléatoire.* Annales INSEE 43, pp.1.

Ladiray D. et Roth N. (1987) *Lissage robuste de séries chronologiques. Une étude expérimentale.* Annales d'Economie et de Statistique n°5, pp. 147-181.`

Laroque G.(1977) *Analyse d'une méthode de désaisonnalisation: le programme X11.* Annales de l'INSEE, 28, pp. 105

Lütkepohl H. (1991) *Introduction to Multiple Time Series Analysis.* Springer-Verlag.

Malinvaud E. (1978) *Méthodes statistiques de l'économétrie.* 3e éd. Dunod.

Mélard G. (1990) *Méthodes de prévision à cours terme.* Coll. Statistique et Mathématiques appliquées. Ed. Ellipses. Tres interessant.

(La revue semestrielle de) **MODULAD,** sur banques de programmes français et utilisations de méthodes statistiques. **INRIA** UCIS DIFFUSION Domaine de Voluceau Rocquencourt BP 105, 78153 Le Chesnay Cedex. tricot@vcnam.cnam.fr
Le numero 15 (Juin 1995) a un article simple sur une méthode non paramétrique de prévision via **régressogramme mobile et prédicteur à noyau,** un autre sur **réseaux neuronaux**.

Newton H. J. (1988) *TIMESLAB: a time series analysis laboratory.* Wadsworth & Brooks/Cole Publishing Company. Advanced Books and Software, Pacific Grove, California.

Ostrow C.W. (1982) *Time Series Analysis: Regression Techniques.* Série: Quantitative Applications in the social Sciences. Sage Publications.

Pankratz A. (1983) *Forecasting with univariate Box-Jenkins Models, Concepts and cases.* John Wiley & Sons.

Pham D.T. (1990) *Choix de modèle en analyse de séries chronologiques*. Revue Statistique et Analyse des données, Vol.15, n°1, p.61. (choix entre les modèles Box et Jenkins)

Priestley M.B. (1981) *Spectral Analysis and Time series,* Academic. NewYork. (tests, grosse bibliographie).

Rappachi Bernard (1990) *Lissons les données!* br@justine.grenet.fr (médianes mobiles).

Sanders D.H., Murph A.F., ENG R.J. (1984) *Les statistiques: une approche nouvelle.* Mac-Graw Hill.

Saporta G. (1981) *Méthodes exploratoires des données temporelles.* Cahiers du Bureau Universitaire de Recherche Opérationnelle n° 37-38..

Tenenhaus M. (1994) *Méthodes Statistiques en gestion.* ISBN 2100010611. Dunod.

Tukey J.W. (1977) *Exploratory data analysis.* Reading, Mass: Addison-Wiley.

Velleman P.F. et Hoaglin D.C. (1981) *Aplications, Basics, and Computing of Exploratory Data Analysis.* Duxbury Press.

Revues spécialisées

Journal of time series analysis ISSN 0143-9782.

Stochastic Processes and their applications. ISSN 0304-4149.

Empirical economics ISSN 0377-7332.

Etc.

Quelques Sites

http://lib.stat.cmu.edu/

http://www.stats.gla.ac.uk/cti/links_stats/software.html#packages

http://www.psychstat.smsu.edu/introbook/sbk27.htm

http://www.psychstat.smsu.edu/MultiBook/mlt02.htm

http://pbil.univ-lyon1.fr/ADE-4/ADE-4F.html

http://forrest.psych.unc.edu/research/ViStaF.html

http://www.datadesk.com/

http://stat.umn.edu/~rcode/

http://www-rocq.inria.fr/scilab/

http://fermivista.math.jussieu.fr/

http://www.sg.cnrs.fr/metiers/bap1/1_13.htm

http://www.clark.net/pub/sinecto/eispack3.htm

http://math.nist.gov/cgi-bin/gams-serve/class/D6.html

http://www.digital.com/info/hpc/software/lapack.html#descr

http://www.linux-center.org/fr/

http://www.adobe.com/prodindex/acrobat/readstep.html

http://www.ams.org/

http://cm.bell-labs.com/cm/ms/departments/sia/doc/index.html

http://www.sfds.asso.fr/

http://www.fundp.ac.be/%7Emremon/sfc/index.shtml

http://www.lsp.ups-tlse.fr/index.html

http://www.statsnet.net/