

Analyse de Séries Chronologiques

J.J. Daudin, C. DUBY, S. Robin & P. Trécourt
(INA-PG, Mathématiques)

Mai 1996

Table des matières

1	Introduction	5
2	Etude de la partie déterministe	7
2.1	Généralités	7
2.1.1	Définition, exemple	7
2.1.2	Que veut-on en faire?	7
2.1.3	Point de vue descriptif graphique	7
2.1.4	Modèles	8
2.1.5	Plan du chapitre	9
2.2	Modèle linéaire généralisé	9
2.2.1	Erreurs indépendantes	9
2.2.2	Erreurs dépendantes	11
2.3	Lissage exponentiel pour la prévision	12
2.3.1	Lissage exponentiel simple	13
2.3.2	Lissage exponentiel double	14
2.3.3	Méthode de Holt et Winters	14
2.3.4	Conclusion	15
2.4	Méthode des moyennes mobiles	15
2.4.1	Principe, définitions et exemples	16
2.4.2	Moyennes mobiles arithmétiques	18
2.4.3	Autres moyennes mobiles	20
2.4.4	Census-X11	21
2.4.5	Traitement des extrémités d'une série	23
2.4.6	Avantages et inconvénients	23
3	Modélisation de la partie aléatoire	25
3.1	Processus stochastique	25
3.1.1	Définition et exemples	25
3.1.2	Processus stationnaire	26
3.1.3	Bruit blanc	27
3.1.4	Autocorrélations	28
3.2	Opérateurs B et Δ	29
3.2.1	Opérateur retard	29

3.2.2	Opérateur différence	30
3.3	Modèle autorégressif	31
3.3.1	Processus autoregressif AR(1)	31
3.3.2	Processus autoregressif AR(2)	33
3.3.3	Processus autoregressif AR(p)	34
3.4	Modèle moyenne mobile	35
3.4.1	Processus moyenne mobile MA(1)	35
3.4.2	Processus moyenne mobile MA(2)	36
3.4.3	Processus moyenne mobile MA(q)	37
3.5	Modèle autorégressif – Moyenne mobile	38
3.5.1	Processus ARMA(p, q)	38
3.5.2	Modèles ARIMA et SARIMA	38
3.6	Identification et estimation des paramètres	40
3.6.1	Identification du modèle: méthode de Box & Jenkins	40
3.6.2	Détermination du type et de l'ordre du modèle	41
3.6.3	Estimation des paramètres et prévision	41
3.6.4	Validation	43
4	Choix de modèles	45
4.1	Analyse spectrale: Recherche des périodicités	45
4.1.1	Rappels sur les fonctions déterministes	45

Chapitre 1

Introduction

Mon intérêt réside dans le futur car je me prépare à y passer le reste de ma vie.

C.F. Kettering

Une série chronologique est la réalisation d'un processus aléatoire indicé par le temps. On modélise un processus par la somme d'une partie déterministe et d'une partie aléatoire (modèle additif), ou par le produit d'une partie déterministe et d'une partie aléatoire (modèle multiplicatif). La séparation en partie déterministe et partie aléatoire est arbitraire.

L'étude d'un processus aléatoire à partir d'une série chronologique a généralement deux objectifs :

- expliquer les variations,
- prédire les valeurs futures.

Les deux objectifs sont souvent liés.

Pour réaliser ces objectifs on peut décrire la série, qui est la trajectoire du processus, et la modéliser. La modélisation comporte deux parties :

- celle de la partie fixe,
- celle de la partie aléatoire.

Suivant le type de la série, une des deux parties peut être prépondérante (cas des courbes de croissance où il suffit généralement de modéliser la partie fixe avec un modèle non linéaire).

Dans les deux cas, on recherche une structure. La recherche de structure pour la partie aléatoire se fera sur une série à laquelle on aura enlevé la partie fixe, on travaille sur un processus dit "stationnarisé".

Le cours se décompose en deux grandes parties. La première consiste à passer en revue les méthodes graphiques (empiriques) et la modélisation de la partie déterministe. La deuxième consiste à présenter les modèles ARIMA qui permettent la modélisation de la partie aléatoire.

Chapitre 2

Etude de la partie déterministe

2.1 Généralités

2.1.1 Définition, exemple

Une série chronologique est la réalisation d'un processus aléatoire indicé par le temps, noté $\{X_t\}$. Pour chaque t , X_t est une variable aléatoire dont on a une réalisation, x_t .

Exemple : la série des consommations mensuelles de viande bovine de 1980 à Octobre 1995 est (en tec) : $x_1 = 129.060$, $x_2 = 129.075$, $x_3 = 131.632$, ... $x_{189} = 119.381$, $x_{190} = 118.565$.

Dans ce cas l'indice t représente le numéro du mois compté à partir du début (Octobre 1995 correspond à $t = 190$). On peut avoir des données journalières, hebdomadaires, mensuelles ou trimestrielles. Dans tous les cas, t est le numéro de l'observation, mais l'unité de temps est différente. Lorsqu'on dispose de plusieurs séries (par exemple la série des consommations mensuelles et celle des prix mensuels) on parle de série chronologique multidimensionnelle.

2.1.2 Que veut-on en faire?

L'étude de la série peut avoir les objectifs suivants non exclusifs :

- Comprendre le passé : expliquer les variations observées
- Prédire les valeurs futures (proches)
- Etudier le lien avec d'autres séries.

2.1.3 Point de vue descriptif graphique

Il est toujours indispensable de tracer l'évolution de la série au cours du temps en utilisant des procédures de lissage diverses et variées. Par exemple le programme SAS :

```
symbol i=sm(nn);  
proc gplot;
```

```

    plot x*t;
run;

```

(où nn est un entier compris entre 0 et 99)

permet d'obtenir une courbe spline (c'est à dire une courbe constituée de polynômes de degré 3 par morceaux dont les dérivées premières et secondes s'accordent aux points de jonctions) d'autant plus lissée que nn se rapproche de 100. Pour $nn=0$, on obtient une courbe spline passant par chaque point. Pour $nn=99$ on obtient une droite.

On peut aussi utiliser "symbol i=r1" (respectivement i=rq et i=rc) qui ajuste un polynôme de degré 1 (respectivement 2 et 3). Ces lissages sont beaucoup moins souples que les splines car il s'agit d'un ajustement global (un seul polynôme pour toute la série).

2.1.4 Modèles

Modèle de base

Le modèle de base auquel on se réfère toujours plus ou moins clairement est le suivant :

$$X_t = f_t + s_t + E_t + p_t$$

où les petites lettres désignent des éléments déterministes et les majuscules désignent des variables aléatoires.

- f_t est la tendance qui peut se décomposer en une tendance "lourde" (souvent représentée par un polynôme) plus un cycle (de périodicité supérieure à un an). Par exemple la consommation de bovins a tendance à diminuer avec un cycle de 3 à 4 ans.
- s_t représente la composante saisonnière de période 4, 12, 52 ou 365 selon qu'il s'agit de données trimestrielles, mensuelles, hebdomadaires ou journalières.
- E_t représente la variation aléatoire due à de nombreuses causes pas forcément bien identifiées, mais de répercussion limitée. Dans le vocabulaire conventionnel des séries chronologique, on emploie les termes "innovation", "irrégularité", "erreur" ou "résiduelle". Une particularité nouvelle des modèles de séries chronologiques par rapport au modèle linéaire classique est l'importance accordée à la modélisation correcte de E_t . On ne peut pas se contenter de considérer que ce sont des variables aléatoires indépendantes ; il faut donc prendre en compte la structure de corrélation par des modèles appropriés.
- p_t représente les perturbations majeures liées à des événements importants mais qui ne se répètent pas et dont l'influence est limitée dans le temps : grève des transports, catastrophe naturelle... Dans le langage de la statistique on parle de points aberrants.

Par contre, à la suite d'un événement particulier (Mai 68) ou d'une évolution profonde qui n'est pas liée à un événement particulier, il peut y avoir une rupture dans le temps : le modèle fonctionne jusqu'à un temps t_0 , mais n'est plus valable après cette date. Il existe des méthodes de détection de rupture de modèle dont nous ne parlons pas dans ce cours.

Modèle multiplicatif

Une variante du modèle de base est le modèle suivant (dit "multiplicatif") :

$$X_t = (f_t + s_t + p_t)E_t$$

Quelques fois on considère les modèles suivants : $X_t = f_t s_t p_t E_t$ ou $X_t = f_t s_t p_t + E_t$. La transformation logarithme permet de passer de l'avant dernier modèle au modèle de base.

2.1.5 Plan du chapitre

Dans ce chapitre, on présente d'abord comment on peut étendre le modèle linéaire au cas des séries chronologiques, puis deux méthodes empiriques de lissage (lissage exponentiel orienté vers la prévision et lissage par les moyennes mobiles pour mettre en évidence la tendance et la saisonnalité).

2.2 Modèle linéaire généralisé

2.2.1 Erreurs indépendantes

Il arrive qu'une bonne modélisation de la partie fixe élimine toute corrélation entre les erreurs successives. On est alors placé dans le cadre habituel du modèle linéaire (supposé connu). On en donne ici 2 exemples :

Modèle trimestriel de "Buys-Ballot"

$$X_t = a + bt + \sum_{j=1}^4 c_j s_{tj} + E_t$$

où $s_{tj} = 1$ si la date t correspond au trimestre j et vaut 0 sinon, $\sum c_j = 0$, $E_t \sim N(0, \sigma^2)$ et les E_t sont indépendantes.

$a + bt$ représente une tendance linéaire et c_j est l'effet du trimestre j . La méthode usuelle des moindres carrés ordinaires donne les estimateurs suivants :

$$A = \bar{X} - \left(\frac{5}{2} + 2(N-1) \right) B$$

$$B = 3 \frac{\sum_{n=1}^{n=N} n \tilde{X}_n - \frac{N(N+1)}{2} \bar{X}}{N(N^2 - 1)}$$

$$C_j = \bar{X}_j - \bar{X} - B \left(j - \frac{5}{2} \right)$$

où n est le numéro de l'année, N est le nombre d'années et \tilde{X}_n est la moyenne de l'année n .

On a les variances des estimateurs des paramètres $a, b, (c_j)$ et σ^2 , et donc des intervalles de confiance pour ces derniers.

On peut également tester les hypothèses du type $\mathbf{H}_0 = \{a = a_0\}$, désaisonnaliser la série par la formule : $X_t^{cvs} = X_t - s_t$ et faire de la prédiction de valeurs futures comme en regression avec un intervalle de confiance pour la valeur prédite.

On peut également tester et prendre en compte une perturbation p_t en considérant le modèle :

$$X_t = a + bt + \sum_{j=1}^4 c_j s_{tj} + dp_t + E_t$$

avec $p_t = 1$ si $t = t_0$ et $p_t = 0$ sinon, où t_0 est la date de la perturbation.

Modèle sur la consommation de viande bovine

On considère un modèle avec une tendance quadratique agrémentée d'un cycle de 4 ans et d'un effet saisonnier mensuel :

$$X_t = a + bt + ct^2 + d \sin\left(\frac{t}{48}\right) + e \cos\left(\frac{t}{48}\right) + \sum_{j=1}^1 f_j s_{tj} + E_t$$

où $s_{tj} = 1$ si la date t correspond au mois j et vaut 0 sinon, $\sum f_j = 0$ $E_t \sim N(0, \sigma^2)$ et les E_t sont indépendantes.

$a + bt + ct^2$ représente une tendance quadratique, $d \sin\left(\frac{t}{48}\right) + e \cos\left(\frac{t}{48}\right)$ représente le cycle de 48 mois et f_j est l'effet du mois j .

Etude de la corrélation des erreurs

Le modèle précédent suppose l'indépendance des erreurs. On doit donc vérifier cet a priori. On utilise la méthode suivante : si les erreurs sont indépendantes, deux erreurs successives doivent être non corrélées. C'est pourquoi on définit le coefficient d'autocorrélation d'ordre 1, noté ρ :

$$\rho = \frac{\text{Cov}(E_t, E_{t-1})}{\sigma^2}$$

On suppose que ρ ne dépend pas de t (processus stationnaire). On estime ρ par :

$$\hat{\rho} = \frac{\sum_{t=2}^{t=T} \hat{E}_t \hat{E}_{t-1}}{\sum_{t=1}^T \hat{E}_t^2}$$

Un test de $\mathbf{H}_0 = (\rho = 0)$ est le test de Durbin-Watson dont la statistique est :

$$DW = \frac{\sum_{t=2}^{t=T} (\hat{E}_t - \hat{E}_{t-1})^2}{\sum_{t=1}^T \hat{E}_t^2}$$

Si T est grand, on a $DW = 2(1 - \hat{\rho})$. La règle de décision est la suivante :

Si $DW < d_1(T)$ on rejette \mathbf{H}_0 ,

Si $DW > d_2(T)$ on accepte \mathbf{H}_0 ,
 Si $d_1(T) < DW < d_2(T)$ on ne peut rien conclure.

Les valeurs de $d_1(T)$ et de $d_2(T)$ sont donnés dans une table (table en annexe pour un risque de 5%).

2.2.2 Erreurs dépendantes

Modèle d'autocorrélation d'ordre 1

Les erreurs ont des tas de façon d'être dépendantes comme on le verra dans le chapitre ARIMA. On considère un des modèles les plus simples de dépendance, le modèle d'autocorrélation d'ordre 1 :

$$E_t = \rho E_{t-1} + \varepsilon_t$$

où ε_t sont des variables aléatoires indépendantes et de même loi $N(0, \gamma)$, et $\rho \in [-1, 1]$

On montre facilement les propriétés suivantes :

$$E(E_t) = 0$$

$$V(E_t) = \frac{\gamma}{1 - \rho^2}$$

$$\text{Cov}(E_t, E_{t-h}) = \gamma \frac{\rho^h}{1 - \rho^2}$$

La matrice de variance des E_t est donc :

$$\mathbf{V} = \frac{\gamma}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix}$$

Estimation des paramètres

Soit \mathbf{H} la matrice :

$$\mathbf{H} = \begin{bmatrix} \sqrt{1 - \rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix}$$

On montre que $V(\mathbf{H} \mathbf{E}) = \gamma \mathbf{I}$, où \mathbf{I} est la matrice identité d'ordre T et \mathbf{E} est le vecteur colonne des E_t . En particulier, il est facile de vérifier que les variables $E_t - \rho E_{t-1}$ sont non corrélées. Si on applique la matrice \mathbf{H} à toutes les lignes du modèle :

$$X_t = f_t + s_t + E_t$$

on obtient :

$$\mathbf{HX} = \mathbf{Hf} + \mathbf{Hs} + \mathbf{HE}$$

où \mathbf{X} est le vecteur colonne des X_t , \mathbf{f} est le vecteur colonne des f_t et \mathbf{s} est le vecteur colonne des s_t . On est ramené au cas d'erreurs non corrélées. En pratique, on est amené à faire la regression usuelle de $X_t - \rho X_{t-1}$ sur $f_t - \rho f_{t-1}$, $s_t - \rho s_{t-1}$. De plus comme on ne connaît pas ρ en général, on procède en deux étapes :

- première étape : on utilise les moindres carrés ordinaires sur X_t et on en déduit $\hat{\rho}$.
- deuxième étape : on utilise la valeur de ρ estimée à l'étape précédente pour faire la regression ordinaire de $X_t - \hat{\rho}X_{t-1}$ sur $f_t - \hat{\rho}f_{t-1}$, $s_t - \hat{\rho}s_{t-1}$.

Prévision

On peut prédire $X_t - \rho X_{t-1}$ en appliquant les résultats du modèle linéaire ordinaire :

$$X_{T+1} - \hat{\rho}X_T = \hat{f}_{T+1} - \hat{\rho}\hat{f}_T + \hat{s}_{T+1} - \hat{\rho}\hat{s}_T + \varepsilon_{T+1}$$

d'où on déduit :

$$X_{T+1} = \hat{f}_{T+1} + \hat{s}_{T+1} + \hat{\rho} [X_T - \hat{f}_T - \hat{s}_T] + \varepsilon_{T+1}$$

$$\hat{X}_{T+1} = \hat{f}_{T+1} + \hat{s}_{T+1} + \hat{\rho}U_T$$

où $U_t = X_T - \hat{X}_T$

Autres modèles d'autocorrélation

On peut utiliser d'autres modèles d'autocorrélation des erreurs avec des corrélations d'ordre supérieur à 1. Par exemple on peut avoir pour des données trimestrielles des autocorrélations d'ordre 4 : $E_t = \rho E_{t-4} + \varepsilon_t$. La démarche reste la même. La procédure de sas qui permet de faire du modèle linéaire avec des erreurs corrélées s'appelle "proc autoreg".

Avantages et inconvénients du modèle linéaire avec erreurs corrélées

Une caractéristique essentielle de cette méthode est qu'elle utilise un modèle global pour la tendance et la saisonnalité. Cela facilite l'interprétation des résultats, mais crée une rigidité trop forte dans certains cas, car ce modèle ne peut pas s'adapter à des évolutions. La prévision est possible à court terme grâce à la modélisation de la partie aléatoire et à plus long terme si la modélisation déterministe est correcte et raisonnablement extrapolable. La mise à jour est un peu laborieuse car il faut tout recalculer à chaque nouvelle observation.

2.3 Lissage exponentiel pour la prévision

Les défauts du modèle linéaire pour la prévision cités précédemment incitent à utiliser a contrario une méthode souple adaptative et facile à mettre à jour. Ce sont justement les qualités du lissage exponentiel.

On se place à la date T (le présent) et veut prédire la série à la date $T + k$. On note $\hat{X}_T(k)$ la prévision ; k est appelé l'horizon de la prévision. Si $k = 1$, on note par convention $\hat{X}_T = \hat{X}_T(1)$

2.3.1 Lissage exponentiel simple

Définition

La prévision fournie par le lissage exponentiel simple est :

$$\hat{X}_T = \lambda \sum_{j=0}^{T-1} (1 - \lambda)^j X_{T-j}$$

$\lambda \in [0, 1]$ s'appelle la constante de lissage. L'idée est que le processus dépend plus des résultats récents de la série que des résultats anciens. (Autrement dit on prévoit le temps de demain par celui d'aujourd'hui). On utilise donc une moyenne pondérée des observations passées avec un poids décroissant vers le passé. La décroissance est d'autant plus rapide que λ est proche de 1. Dans ce dernier cas on a une prévision "souple et réactive", alors que si λ est proche de 0 on a une prévision rigide et peu sensible aux dernières fluctuations.

Propriétés

On a une équation de récurrence :

$$\hat{X}_T = \lambda X_T + (1 - \lambda) \hat{X}_{T-1}$$

qui peut encore s'écrire :

$$\hat{X}_T = \hat{X}_{T-1} + \lambda(X_T - \hat{X}_{T-1})$$

La première équation montre que la prévision au temps T est une moyenne pondérée de la prévision au temps $T - 1$ et du résultat de la série au temps T . Elle permet de faire une mise à jour **très simple**. L'initialisation est $\hat{X}_1 = X_1$. La deuxième relation définit la prévision au temps T comme la prévision au temps $T - 1$ corrigée de l'erreur de prévision commise au temps $T - 1$ amortie d'un facteur λ .

On montre facilement que \hat{X}_T est solution du problème de minimisation suivant :

$$\text{Trouver } a \text{ tel que } \sum_{j=0}^{T-1} (1 - \lambda)^j (X_{T-j} - a)^2 \text{ soit minimum.}$$

\hat{X}_T est donc la constante la plus proche de la série au voisinage de T .

Choix de λ

On peut choisir λ arbitrairement selon que l'on veut une prévision réactive ou rigide ; on peut également choisir la valeur de λ qui minimise la somme des carrés des erreurs de

prévision quand on applique la méthode aux données passées. λ est alors la solution du problème de minimisation :

Trouver λ tel que

$$\sum_{t=1}^{T-1} \left[X_{t+1} - \lambda \sum_{j=0}^{t-1} (1-\lambda)^j X_{t-j} \right]^2$$

soit minimum.

2.3.2 Lissage exponentiel double

On suppose qu'il existe une tendance linéaire : $f_t = a + bt = a_1 + (t-T)a_2$. On cherche les valeurs de a_1 et a_2 qui minimisent :

$$\sum_{j=0}^{T-1} (1-\lambda)^j (X_{T-j} - a_1 + a_2 j)^2$$

Une solution approchée est :

$$A_1(T) = 2s_1 - S_2$$

$$A_2(T) = \frac{\lambda}{1-\lambda} (S_1 - S_2)$$

où

$$S_1 = \lambda \sum_{j=0}^{T-1} (1-\lambda)^j X_{T-j}$$

$$S_2 = \lambda^2 \sum_{j=0}^{T-1} j(1-\lambda)^j X_{T-j} + \lambda S_1$$

Il existe des formules de mise à jour plus simples :

$$A_1(T) = A_1(T-1) + A_2(T-1) + \lambda(2-\lambda)(X_T - \hat{X}_{T-1})$$

$$A_2(T) = A_2(T-1) + \lambda^2(X_T - \hat{X}_{T-1})$$

2.3.3 Méthode de Holt et Winters

Méthode non saisonnière

Il s'agit d'un lissage exponentiel double avec des relations de mises à jour modifiées qui utilisent des paramètres de "réglage" permettant de choisir une prévision plus ou moins réactive :

$$A_1(T) = (1-\alpha) [A_1(T-1) + A_2(T-1)] + \alpha X_T$$

$$A_2(T) = (1-\gamma)A_2(T-1) + \gamma [A_1(T) - A_1(T-1)]$$

Le choix de α et γ est arbitraire dans l'intervalle $[0, 1]$; si on veut une prévision réactive, on les choisira proches de 1.

Méthode saisonnière additive

On considère le modèle suivant :

$$X_t = a_1 + a_2(t - T) + s_t + E_t$$

A la différence du modèle linéaire, les paramètres a_1 et a_2 changent à chaque instant t . Les formules de mise à jour sont :

$$A_1(T) = (1 - \alpha) [A_1(T - 1) + A_2(T - 1)] + \alpha(X_T - S_{T-\tau})$$

$$A_2(T) = (1 - \gamma)A_2(T - 1) + \gamma [A_1(T) - A_1(T - 1)]$$

$$S_T = (1 - \delta)S_{T-\tau} + \delta [X_T - A_1(T)]$$

L'initialisation dans le cas de données trimestrielles est :

$$A_1(3) = \frac{1}{8}(X_1 + 2X_2 + 2X_3 + 2X_4 + X_5)$$

$$A_1(4) = \frac{1}{8}(X_2 + 2X_3 + 2X_4 + 2X_5 + X_6)$$

$$A_2(4) = A_1(4) - A_1(3)$$

$$S_1 = X_1 - A_1(3) + 2A_2(4) \quad S_2 = X_2 - A_1(3) + A_2(4), \quad S_3 = X_3 - A_1(3) \quad \text{et} \quad S_4 = X_4 - A_1(4)$$

Il existe une méthode saisonnière multiplicative.

2.3.4 Conclusion

La plupart des méthodes de lissages sont des cas particuliers de prévisions associées à des modèles ARIMA (sauf le modèle de Holt-Winters avec saisonnalité multiplicative). Leur simplicité conceptuelle et la facilité de la mise à jour en font un grand classique de la prévision. Cependant le choix arbitraire des coefficients n'est pas toujours facile à faire; d'autre part la prédiction est moins bonne que celle faite par un modèle ARIMA bien choisi quand le modèle vrai des erreurs n'entre pas dans le cadre des méthodes de lissage exponentiel. La procédure de SAS qui permet d'utiliser les méthodes de lissage exponentiel est "proc Forecast".

2.4 Méthode des moyennes mobiles

Ces méthodes permettent d'obtenir la tendance et la saisonnalité d'une série ainsi qu'une série corrigée des variations saisonnières. Il n'y a pas de modèle explicitement posé et la méthode est essentiellement empirique. Les moyennes mobiles sont à la base de la méthode "Census-X11" développée par le bureau de recensement du département US du commerce, méthode souvent utilisée par les administrations.

2.4.1 Principe, définitions et exemples

Principe

Le principe consiste à utiliser des opérateur "moyennes mobiles" comme des filtres qui annulent certaines composantes et en laissent d'autres invariantes. Si tout va bien ces filtres font donc apparaître exclusivement la composante saisonnière (ou exclusivement la tendance) et réduisent le "bruit" associé à la composante aléatoire.

Par exemple si on considère le modèle :

$$X_t = f_t + s_t + E_t + p_t$$

Si on a un filtre M tel que $M(s_t) = 0$ et $M(f_t) = f_t$ et qu'on sait se débarrasser des données aberrantes p_t on a :

$$M(X_t) = f_t + M(E_t)$$

Si $M(E_t)$ est de variance faible on pourra assimiler $M(X_t)$ avec f_t .

Définitions

- La série $Y_t = \sum_{i=-m_1}^{m_2} \alpha_i X_{t+i}$ est une moyenne mobile d'ordre m_1+m_2+1 de la série X_t . m_1 et m_2 sont des entiers positifs appartenant à l'intervalle $[0, T]$ et les α_i sont des réels.
- On note B "l'opérateur retard" défini par $B(X_t) = X_{t-1}$.
- L'opérateur moyenne mobile M est un polynôme en B : $M = \sum_{i=-m_1}^{m_2} \alpha_i B^{-i}$
- On dit que M est une moyenne mobile centrée si $m_1 = m_2$
- On dit que M est une moyenne mobile symétrique si elle est centrée et si $\forall i, \alpha_i = \alpha_{-i}$
On utilise alors la notation $M = [2m + 1; \alpha_m, \alpha_{m-1}, \dots, \alpha_0]$
- Le rapport $\frac{V(M(E_t))}{V(E_t)}$ s'appelle le facteur de réduction du bruit. Il est égal à $\sum_{i=-m_1}^{m_2} \alpha_i^2$

Exemples

- $M = [3; \frac{1}{3}, \frac{1}{3}]$ donne

$$Y_t = M(X_t) = \frac{1}{3}(X_{t-1} + X_t + X_{t+1})$$

C'est une moyenne mobile centrée d'ordre 3.

Elle laisse invariante les polynômes de degré 1 : si $X_t = a + bt$, il est facile de voir que $M(X_t) = X_t$.

Elle annule la série $X_t = \cos(\frac{2\pi t}{3})$, c'est à dire la série sinusoïdale de période 3.

On considère la série

$$X_t = a + bt + \cos(\frac{2\pi t}{3}) + E_t$$

On a

$$M(X_t) = a + bt + M(E_t)$$

avec $V(M(E_t)) = \frac{\sigma^2}{3}$, c'est à dire que $M(E_t)$ représente la tendance linéaire perturbée par un bruit réduit d'un facteur $\frac{1}{3} = 0.33$

$$Y_t = M(X_t) = \frac{1}{2}(X_{t-1} + X_t)$$

C'est une moyenne mobile non centrée d'ordre 2. En pratique elle est "centrée" ou "positionnée" au temps $t - \frac{1}{2}$. On a toujours ce problème avec les moyennes mobiles d'ordre pair : elle ne peuvent pas être centrées. Pour éliminer ce problème, on fait la moyenne des deux moyennes mobiles d'ordre pair décalées d'une unité de temps :

$$\begin{aligned} Y_t &= \frac{1}{2}(M(X_t) + M(X_{t+1})) = \frac{1}{4}[(X_{t-1} + X_t) + (X_t + X_{t+1})] \\ &= \frac{1}{4}(X_{t-1} + 2X_t + X_{t+1}) \end{aligned}$$

On voit qu'il s'agit finalement de la moyenne mobile centrée d'ordre 3 : $M = [3; 1, 2]$. Cette moyenne mobile conserve les polynômes de degré 1 et annule la série $X_t = (-1)^t$. Par conséquent, si on a une série

$$X_t = a + bt + (-1)^t + E_t$$

alors on a :

$$M(X_t) = a + bt + M(E_t)$$

avec $V(M(E_t)) = \frac{3}{8} = 0.375$.

- $M = [5; \frac{1}{2}, 1, 1]$ est la moyenne mobile centrée d'ordre 5 dont on se sert pour éliminer la saisonnalité trimestrielle de période 4, avec un facteur de réduction du bruit de 0.22.
- $M = [13; \frac{1}{2}, 1, 1, 1, 1, 1, 1]$ est la moyenne mobile centrée d'ordre 13 dont on se sert pour éliminer la saisonnalité mensuelle de période 12 avec un facteur de réduction du bruit de 0.08.

Propriétés élémentaires

- Si M_1 et M_2 sont des moyennes mobiles, $M_1 \circ M_2$ et $M_1 + M_2$ le sont également. De plus si M_1 et M_2 sont centrées, $M_1 \circ M_2$ et $M_1 + M_2$ le sont également et si M_1 et M_2 sont symétriques $M_1 \circ M_2$ et $M_1 + M_2$ le sont également.
- On désigne par $\text{Ker}(M)$ l'ensemble des séries X_t telles que $M(X_t) = 0$. On a $\text{Ker}(M_1 \circ M_2) = \text{Ker}(M_1) + \text{Ker}(M_2)$.

Les séries appartenant au noyau de M vérifient l'équation de récurrence

$$\sum_{i=-m}^m \alpha_i X_{t+i} = 0$$

dont les solutions sont de la forme

$$\sum \beta_j c_j^t$$

où les c_j sont des nombres complexes solutions de l'équation caractéristique :

$$\sum_{i=0}^{2m} \alpha_{i-m} c^i = 0$$

- Les séries invariantes par M vérifient l' équation de récurrence

$$\sum_{i=-m}^m \alpha_i X_{t+i} = X_t$$

dont les solutions sont de la même forme que précédemment (avec des racines différentes car l'équation de récurrence n'est pas la même).

- $\sum_{i=-m}^m \alpha_i = 1 \Leftrightarrow$ M conserve la série constante $X_t = a$
- Si M est symétrique et conserve la série constante alors M conserve les polynômes de degré 1.
- Si X_t est vecteur propre de M ($M(X_t) = \lambda X_t$), la transformation M a un double effet :
 - multiplication de X_t par $|\lambda|$
 - effet de phase (c'est à dire décalage de la partie périodique de la série) si λ n'est pas un nombre réel.

- Effet de "Slutsky-Yule" : c'est l'effet de corrélation induit par la transformation M qui peut faire croire à une périodicité qui n'existe pas dans la série de départ.

Cet effet est dû aux corrélations entre les $M(E_t)$;

on a , si $-2m \leq h \leq 2m$ et si $\text{Cov}(E_t, E_{t+h}) = 0$:

$$\text{Cov}(E_t, E_{t+h}) = \gamma \sum_{i=-m}^m \alpha_i \alpha_{i-h}$$

La covariance est nulle si $h > |2m|$.

2.4.2 Moyennes mobiles arithmétiques

Définition, exemple

Si on veut une moyenne mobile symétrique telle que $\sum \alpha_i = 1$ et qui minimise $\sum \alpha_i^2$, on obtient :

$$\forall i \in [-m, m] \quad \alpha_i = \frac{1}{2m+1}$$

appelée moyenne mobile arithmétique. Par exemple $M = [3; \frac{1}{3}, \frac{1}{3}]$.

Noyau

L'équation caractéristique est

$$\sum_{i=0}^{2m} \lambda^i = \frac{1 - \lambda^{2m+1}}{1 - \lambda}$$

dont les solutions sont :

$$c_j = e^{\frac{i2\pi j}{2m+1}} \quad j = 1, \dots, 2m$$

Par suite les séries annulées par la moyenne mobile arithmétique d'ordre $2m + 1$ sont de la forme :

$$X_t = \sum_{j=1}^m \left[c_{1j} \cos\left(\frac{2\pi j}{2m+1}\right) + c_{2j} \sin\left(\frac{2\pi j}{2m+1}\right) \right]$$

qui sont les séries périodiques de période $2m + 1$, nulles en moyenne sur une période.

Séries invariantes

Ce sont les polynômes de degré 1 et d'autres séries liées aux autres racines de l'équation caractéristique.

Par exemple $M = \left[5; \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right]$ a un polynôme caractéristique pour l'invariance égal à

$$c(c^4 + c^3 - 4c^2 + c + 1)$$

dont les racines autres que 1 sont : $\frac{-3 \pm \sqrt{5}}{2}$. Par suite, en plus des polynômes de degré 1, les séries $X_t = \left(\frac{-3 \pm \sqrt{5}}{2}\right)^t$ sont invariantes par M.

Transformation du bruit blanc

Le facteur de réduction du bruit est égal à $\frac{1}{2m+1}$ et l'autocorrélation des $M(E_t)$ est égale à $\rho(h) = \frac{2m+1-h}{2m+1}$ si $-2m < h < 2m$.

Annulation des périodicités paires

On a vu qu'une série arithmétique annule les séries périodiques de période $2m + 1$. On ne peut donc pas annuler une périodicité paire avec une moyenne arithmétique. Par contre on peut utiliser une moyenne mobile "presque" arithmétique pour annuler une périodicité paire, ce qui est important, car en pratique les saisonnalités sont de période 4 ou 12. Pour ce faire, on utilise les moyennes mobiles

$$M = \left[2m + 1; \frac{1}{4m}, \frac{1}{2m}, \dots, \frac{1}{2m}\right]$$

Le polynôme caractéristique du noyau est

$$(1 + c) \frac{1 - c^{2m}}{1 - c}$$

M annule donc les séries de période $2m$ nulles en moyenne sur une période.

Par exemple pour $m = 6$ la moyenne mobile

$$M = \left[13; \frac{1}{24}, \frac{1}{12}, \dots, \frac{1}{12} \right]$$

annule les phénomènes saisonniers, a un facteur de réduction du bruit de 0.08 et une périodicité induite par "l'effet Slutsky-Yule" de l'ordre de 15.3.

Autres moyennes arithmétiques

– La moyenne mobile :

$$M = \left[7; \frac{1}{16}, \frac{2}{16}, \frac{3}{16}, \frac{4}{16} \right]$$

annule les saisonnalités de période 4 dont l'amplitude varie linéairement.

– La moyenne mobile de Spencer d'ordre 15 :

$$M = \left[15; \frac{1}{320} (-3, -6, -5, 3, 21, 46, 67, 74) \right]$$

annule les saisonnalités de période 4 et 5 (au cas où il y aurait un décalage) et conserve les polynômes de degré 3 ou moins. Il en existe une variante pour les séries mensuelles.

2.4.3 Autres moyennes mobiles

Il existe d'autres critères pour choisir une moyenne mobile :

– On peut vouloir conserver les polynômes locaux sur des intervalles de longueur $2m + 1$, comme par exemple :

$$M = \left[5; \frac{-3}{35}, \frac{12}{35}, \frac{17}{35} \right]$$

qui conserve les polynômes locaux de degré 3 sur des intervalles de longueur 5.

– On peut maximiser l'effet de lissage de la série, en imposant que les coefficients α_i ne varient pas trop brusquement. Pour cela on cherche à minimiser la quantité Q :

$$Q = \sum_i (\Delta^3 \alpha_i)^2 = \sum_i (\alpha_i - 3\alpha_{i-1} + 3\alpha_{i-2} - \alpha_{i-3})^2$$

Les moyennes mobiles de Henderson d'ordre $2m + 1$ minimisent Q sous la contrainte de conservation des polynômes de degré 2.

On peut résumer les mérites respectifs de chaque méthode par le tableau suivant (extrait du livre de Gourieroux et Monfort "Séries temporelles et modèles dynamiques") :

Type de moyenne mobile	Arithmétique	Conservation de polynômes locaux	Spencer	Henderson
Conservation de la tendance	0	++	+	+
Annulation de la saisonnalité	++	+	+	+
Réduction du bruit blanc	+	+	+	+
Lissage	-	-	+	+
Simplicité des coefficients	++	0	++	+

2.4.4 Census-X11

Il s'agit d'une méthode qui utilise de façon itérative (et intensive) toutes sortes de moyennes mobiles pour déterminer la tendance et la saisonnalité, en éliminant progressivement l'influence des perturbations. Il existe une version trimestrielle et une version mensuelle. L'effet saisonnier peut être multiplicatif ou additif. La procédure sas "proc X11" procède de la façon suivante dans sa version mensuelle additive :

4 grandes itérations

On considère le modèle :

$$X_t = f_t + s_t + E_t + p_t$$

- La première itération sert à calculer des poids préliminaires pour éliminer ou atténuer les perturbations p_t . Plus précisément, on calcule une première estimation \hat{f}_t et \hat{s}_t de f_t et de s_t d'où on déduit les valeurs estimées de $E_t + p_t$:

$$\widehat{(E_t + p_t)} = X_t - \hat{f}_t - \hat{s}_t$$

On calcule un poids $w_t \in [0, 1]$ pour chaque temps t et on constitue une série dite "corrigée" égale à

$$X_t^c = \hat{f}_t + \hat{s}_t + w_t(E_t + p_t)$$

Le poids w_t vaut 1 sauf si $\widehat{(E_t + p_t)}$ dépasse une valeur seuil qui dépend de l'écart-type de $\widehat{(E_t + p_t)}$; dans ce cas le poids est d'autant plus faible que la valeur de $\widehat{(E_t + p_t)}$ est élevée. Si $w_t = 1$ on a $X_t^c = X_t$

- La deuxième itération utilise la série corrigée et procède de la même façon qu'à l'étape 1 pour calculer de nouveaux poids, appelés "poids finaux". On espère qu'à cette étape on a éliminé l'élément p_t de $E_t + p_t$ qui ne contient plus que la "composante irrégulière" E_t .
- La troisième itération calcule la tendance et les effets saisonniers à partir de la série issue de la deuxième itération.
- La quatrième itération présente les résultats et donne la série corrigée des variations saisonnières et la série des E_t après un dernier "rabortage" des valeurs trop élevées de E_t .

Description d'une itération

Les 3 premières itérations sont construites sur le même modèle en 9 étapes :

1. On se donne la série de départ de l'itération qui peut être X_t si on en est à la première itération ou une série corrigée sinon. Cette série est notée B1 dans la première itération de la procédure X11 de sas.
2. On calcule une première estimation de la tendance par $\hat{f}_t = M_1(X_t)$ où $M_1 = [13; \frac{1}{2}, 1, 1, 1, 1, 1, 1]$. (notée B2)
3. On calcule une première estimation des coefficients saisonniers: $\hat{s}_t = X_t - \hat{f}_t$. (B3)
4. On corrige les coefficients s_t trop grands en les remplaçant par une moyenne mobile des coefficients correspondant au même mois. (B4)
5. On calcule les coefficients saisonniers mensuels par moyenne mobile :

$$\hat{s}_t = (I - M_1) \circ M_2(\hat{f}_t)$$

où $M_2(\hat{f}_t)$ est la moyenne mobile d'ordre 3 sur les coefficients saisonniers (pour un même mois). L'opération $(I - M_1)$ permet d'obtenir des coefficients saisonniers de somme approximativement nulle. (B5)

6. On calcule une première série corrigée des variations saisonnières: $X_t^{cvs} = X_t - \hat{s}_t$ (B6)
7. On calcule une deuxième estimation de la tendance par :

$$\hat{f}_t = M_3(X_t^{cvs})$$

où M_3 est une moyenne mobile de Henderson d'ordre 9. (B7)

8. On recommence la procédure à partir de l'étape 3 jusqu'à l'étape 6. On a donc à cette étape une nouvelle estimation de s_t notée $s_{3,t}$. (B8, B9, B10, B11)
9. On estime $E_t + p_t$ par $X_t - \hat{f}_t - s_{3,t}$. Ces valeurs permettent d'obtenir les poids w_t définis plus haut. (B13, B17)

Il existe des options pour le nombre de jours ouvrables dans le mois.

2.4.5 Traitement des extrémités d'une série

La transformation $M(X_t)$ ne fonctionne pas pour les valeurs de t proches de 1 et de T . On peut construire une moyenne mobile non centrée pour les données récentes :

$$M(X_T) = \sum_{i=-m}^0 \alpha_i X_{T+i}$$

On peut également prédire les valeurs futures et calculer la moyenne mobile centrée à l'aide de ces dernières.

2.4.6 Avantages et inconvénients

L'approche moyenne mobile est très souple et résiste mieux aux changements de régime que le modèle linéaire par exemple. Elle permet de prendre en compte et d'éliminer les perturbations. Par contre le choix de l'ordre et des coefficients est empirique et réclame de l'expérience.

Chapitre 3

Modélisation de la partie aléatoire

3.1 Processus stochastique

3.1.1 Définition et exemples

On appellera ici processus stochastique (à temps discret) une suite $\{X_t\}$ de variables aléatoires indicée par le temps.

Il faut bien noter que

- les X_t ne sont pas forcément indépendants les uns des autres,
- la loi de X_t dépend de t :

$$X_t \sim F_t.$$

La notion de processus généralise celle de variable aléatoire: la réalisation d'un processus est une série chronologique (on emploie également le terme de trajectoire).

Exemples :

- La suite $\{X_t\}$ dans laquelle les X_t sont i.i.d. gaussiens constitue un échantillon mais également un processus. On montre ici deux de ses trajectoires :

Echantillon gaussien standard

Dans cet exemple, on a

$$\begin{aligned} X_t &\sim \mathcal{N}(0, 1) \\ \text{Cov}(X_t, X_s) &= 0, \quad \forall (t, s). \end{aligned}$$

- Un exemple plus intéressant de processus est le processus Y_t des cumulants de l'échantillon $\{X_t\}$:

$$Y_t = \sum_{s=1}^t X_s, \quad Y_0 = 0.$$

Les trajectoires de ce processus correspondant à celles de $\{X_t\}$ présentées plus haut sont :

Cumul d'un échantillon gaussien standard

Dans cet exemple, on a

$$\left. \begin{array}{l} E(Y_t) = \sum_{s=1}^t E(X_s) = 0 \\ V(Y_t) = \sum_{s=1}^t V(X_s) = t \end{array} \right\} \Rightarrow Y_t \sim \mathcal{N}(0, t)$$

et

$$\begin{aligned} \text{Cov}(Y_s, Y_t) &= E(Y_s Y_t) = E\left(\sum_{r=1}^s X_r \times \sum_{r=1}^t X_r\right) = \sum_{r=1}^{\min(s,t)} E(X_r^2) \\ &= \min(s, t), \end{aligned}$$

$$\begin{aligned} E(Y_t | Y_s = y_s) &= E\left(\sum_{r=1}^t X_r \mid \sum_{r=1}^s X_r\right) = \sum_{r=1}^s x_r + E\left(\sum_{r=s+1}^t X_r\right) = \sum_{r=1}^s x_r \\ &= y_s \quad \text{pour } t > s. \end{aligned}$$

Les variables aléatoires Y_t ne sont évidemment pas indépendantes : la loi de Y_t dépend des valeurs prises par le processus aux dates antérieures.

3.1.2 Processus stationnaire

Dans la suite, on n'étudiera que des processus stationnaires c'est à dire des processus dont la loi ne varie pas au cours du temps.

Plus précisément, un processus est dit stationnaire si la loi de tout k -uplets est invariante par translation dans le temps :

$$\forall k, \quad \forall (t_1, t_2, \dots, t_k), \quad \mathcal{L}(X_{t_1}, X_{t_2}, \dots, X_{t_k}) = \mathcal{L}(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h}).$$

Cette hypothèse de stabilité temporelle de la loi du processus facilite évidemment bien des analyses. Pratiquement, cette hypothèse implique notamment que la covariance entre les valeurs prises en 2 dates distinctes par le processus ne dépend que de la longueur de la période qui les sépare (et pas de la date initiale) :

$$E(X_t), \quad V(X_t) \quad \text{et} \quad \text{Cov}(X_t, X_{t-h}) \quad \text{ne dépendent pas de } t,$$

soit

$$\begin{aligned} E(X_t) &= \mu, \\ V(X_t) &= \sigma^2, \\ \text{Cov}(X_t, X_{t-h}) &= \gamma_h. \end{aligned}$$

Dans les exemples vus plus haut, il est clair que l'échantillon $\{X_t\}$ est un processus stationnaire, mais pas $\{Y_t\}$ (puisque $\text{Cov}(Y_s, Y_t) = s \wedge t$).

La stationnarité est donc une hypothèse essentielle pour pouvoir estimer les paramètres caractéristiques du processus : elle implique notamment que le processus n'a pas de tendance.

En l'absence de stationnarité, chaque X_t est une variable aléatoire avec son espérance et sa variance propres et on dispose d'une seule réalisation (x_t) de cette variable. Les estimations de son espérance et de sa variance devraient donc se fonder sur une seule observation, ce qui est impossible.

Processus faiblement stationnaire : Un processus est dit faiblement stationnaire si ses moments d'ordre 1 et 2 sont stationnaires :

$$\mathbb{E}(X_t) \equiv \mu, \quad \mathbb{V}(X_t) \equiv \sigma^2, \quad \text{Cov}(X_t, X_{t-h}) = \gamma_h.$$

- La stationnarité entraîne la faible stationnarité.
- Un processus gaussien faiblement stationnaire est stationnaire.

3.1.3 Bruit blanc

Un processus $\{X_t\}$ est un bruit blanc s'il constitue un échantillon i.i.d. d'espérance nulle :

$$\begin{aligned} \forall t & : X_t \sim F, & \mathbb{E}(X_t) &= 0, \\ \forall t \neq s & : (X_t, X_s) & \text{indépendants.} \end{aligned}$$

Un tel processus n'a ni tendance ni mémoire : la connaissance de la valeur du processus à une date donnée n'apporte aucune information pour la prédiction de sa valeur à une date ultérieure :

$$\forall t > s : \mathcal{L}(X_t | X_s) = \mathcal{L}(X_t).$$

Bruit blanc faible : Un processus est un bruit blanc faible s'il constitue un échantillon non corrélé :

$$\begin{aligned} \forall t & : X_t \sim F, & \mathbb{E}(X_t) &= 0, \\ \forall t \neq s & : \text{Cov}(X_t, X_s) &= 0. \end{aligned}$$

- Un bruit blanc est un bruit blanc faible
- Un bruit blanc faible gaussien est un bruit blanc.

3.1.4 Autocorrélations

Définition 1 : On étudie la "mémoire" d'un processus en calculant son **autocorrélation** de retard h noté ρ_h :

$$\rho_h = \text{Corr}(X_t, X_{t-h}) = \frac{\text{Cov}(X_t, X_{t-h})}{\sqrt{V(X_t)V(X_{t-h})}}$$

qui mesure le lien entre les valeurs du processus à deux dates distantes de h .

Pour un processus stationnaire, ρ_h prend une forme plus simple :

$$\rho_h = \frac{\text{Cov}(X_t, X_{t-h})}{V(X_t)} = \frac{\gamma_h}{\gamma_0}.$$

On peut tracer la courbe $\rho_h = f(h)$ qui est appelée (auto-)corrélogramme.

Définition 2 : De même, on définit l'**autocorrélation partielle** de retard h comme la corrélation entre $(X_t - X_t^*)$ et $(X_{t-h} - X_{t-h}^*)$ où X_t^* désigne la régression de X_t sur les $(h-1)$ valeurs $\{X_{t-1}, X_{t-2}, \dots, X_{t-h+1}\}$:

$$\tau_h = \text{Corr}(X_t - X_t^*, X_{t-h} - X_{t-h}^*) = \frac{\text{Cov}(X_t - X_t^*, X_{t-h} - X_{t-h}^*)}{\sqrt{V(X_t - X_t^*)V(X_{t-h} - X_{t-h}^*)}}$$

avec

$$X_t^* = \sum_{k=1}^{h-1} \alpha_k X_{t-k}, \quad X_{t-h}^* = \sum_{k=1}^{h-1} \beta_k X_{t-h-k}$$

où les α_k et les β_k sont les coefficients des régressions.

Cette quantité rend compte de l'intensité de la liaison entre X_t et X_{t-h} en supprimant les liaisons induites par des variables intermédiaires $\{X_{t-1}, X_{t-2}, \dots, X_{t-h+1}\}$.

On peut ainsi remarquer que pour tout processus,

$$\rho_1 = \tau_1$$

puisque qu'il n'y a aucune variable intermédiaire entre X_t et X_{t-1} .

Comme en régression multiple, l'estimation des τ_h nous permet de mesurer le retard qu'il faut remonter pour trouver une information originale sur X_t .

Estimation : En pratique, les ρ_h sont évidemment inconnus mais sur une série $\{x_t, 1 \leq t \leq T\}$ on peut les estimer avec la formule naturelle

$$\hat{\rho}_h = \frac{\sum_{t=h+1}^{t=T} (x_t - \bar{x})(x_{t-h} - \bar{x})}{\sum_{t=1}^{t=T} (x_t - \bar{x})^2}.$$

où \bar{x} représente la moyenne des valeurs observées.

Il n'existe pas de forme simple des estimateurs des τ_h : les autocorrélations partielles sont généralement estimées par récurrence. Ainsi, on sait que τ_2 vaut

$$\tau_2 = \frac{\text{Corr}(X_t, X_{t-2}) - \text{Corr}(X_t, X_{t-1})\text{Corr}(X_{t-2}, X_{t-1})}{\sqrt{(1 - \text{Corr}(X_t, X_{t-1})^2)(1 - \text{Corr}(X_{t-2}, X_{t-1})^2)}}$$

et peut donc être estimé, en utilisant la stationnarité, par

$$\hat{\tau}_2 = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}.$$

Syntaxe SAS : Les corrélogrammes (ainsi que d'autres analyses) d'une série $\{X_t\}$ contenue dans un tableau "TAB" peuvent être obtenus par la procédure ARIMA de SAS :

```
proc ARIMA data=TAB;
  identify var=X;
```

3.2 Opérateurs B et Δ

Pour étudier les processus (et donc les séries chronologiques), on définit des opérateurs retard et de différenciation.

3.2.1 Opérateur retard

L'opérateur B décale le processus d'une unité de temps vers le passé :

$$B(X_t) = X_{t-1}.$$

Si on applique h fois cette opérateur, on décale le processus de h unités de temps :

$$\underbrace{B(B(\dots B(X_t)\dots))}_{h \text{ fois}} = B^h(X_t) = X_{t-h}.$$

Syntaxe SAS : Dans les étapes 'data' de SAS, l'opérateur retard est noté 'lag' :

$$\underbrace{X}_{X_t} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 9 \end{bmatrix} \Rightarrow \underbrace{\text{lag}(X)}_{B(X_t)} = \begin{bmatrix} \cdot \\ 1 \\ 3 \\ 5 \end{bmatrix} \Rightarrow \underbrace{\text{lag2}(X)}_{B^2(X_t)} = \begin{bmatrix} \cdot \\ \cdot \\ 1 \\ 3 \end{bmatrix}.$$

Dans le programme suivant :

```
data TAB;
  infile 'serie.don';
  input T X;
```

```

Y=lag(X);
proc Corr data=TAB;
var X;
with Y;

```

la sortie de la procédure Corr donne l'autocorrélation d'ordre 1 de la série $\{X_t\}$.

3.2.2 Opérateur différence

L'opérateur Δ fait la différence entre le processus et sa version décalée de une unité de temps. Cet opérateur se construit en utilisant l'opérateur précédent :

$$\begin{aligned} \Delta(X_t) &= X_t - X_{t-1} = X_t - B(X_t) \\ \iff \Delta &= I - B \end{aligned}$$

où I est l'opérateur identité :

$$I(X_t) = X_t.$$

Dans la littérature, cet opérateur est parfois noté ∇ .

Propriétés :

– Elimination de la tendance :

L'opérateur Δ élimine les tendances linéaires. Pour un processus de la forme

$$X_t = at + b + E_t$$

où E_t est stationnaire, on a

$$\Delta(X_t) = X_t - X_{t-1} = (at + b + E_t) - (a(t-1) + b + E_{t-1}) = a + (E_t - E_{t-1}).$$

De façon générale, l'opérateur Δ^d élimine les tendances polynomiales de degré d . Par exemple, pour une tendance de degré 2

$$X_t = at^2 + bt + c + E_t$$

on a

$$\Delta^2 = (I-B)^2 = I - 2B + B^2$$

et donc

$$\Delta^2(X_t) = X_t - 2X_{t-1} + X_{t-2} = \dots = 2a + (E_t - 2E_{t-1} + E_{t-2}).$$

Il faut bien noter que l'opérateur Δ^d raccourcit la série des x_t de d valeurs puisque $\Delta^d X_t$ n'est défini que pour $t > d$.

– **Élimination de la saisonnalité :**

L'opérateur $\Delta_\omega = (I - B^\omega)$ élimine une saisonnalité de période ω (ω entier). Si on a un processus de la forme

$$X_{k\omega+j} = s_j + E_{k\omega+j}$$

où E_t est stationnaire, on a

$$\begin{aligned} \Delta_\omega(X_{k\omega+j}) &= (I - B^\omega)(X_{k\omega+j}) = X_{k\omega+j} - X_{(k-1)\omega+j} \\ &= (s_j + E_{k\omega+j}) - (s_j + E_{(k-1)\omega+j}) = E_{k\omega+j} - E_{(k-1)\omega+j}. \end{aligned}$$

On utilise fréquemment les opérateurs Δ_{12} pour les données mensuelles, Δ_7 pour les données journalières, Δ_{24} pour les données horaires, *etc.*

Cet opérateur est également coûteux en information puisqu'il raccourcit la série de ω valeurs, les ω premières servant de références pour la saisonnalité.

Syntaxe SAS : Il n'existe pas de syntaxe spécifique pour l'opérateur Δ dans SAS mais il se reconstruit facilement avec l'opérateur 'lag' :

$$\underbrace{\mathbf{X}}_{X_t} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 9 \end{bmatrix} \Rightarrow \underbrace{\text{lag}(\mathbf{X})}_{B(X_t)} = \begin{bmatrix} \cdot \\ 1 \\ 3 \\ 5 \end{bmatrix} \Rightarrow \underbrace{\mathbf{X} - \text{lag}(\mathbf{X})}_{\Delta(X_t) = X_t - B(X_t)} = \begin{bmatrix} \cdot \\ 2 \\ 2 \\ 4 \end{bmatrix}.$$

3.3 Modèle autorégressif

Dans l'étude d'une série chronologique, il est naturel de penser que la valeur de la série à la date t peut dépendre des valeurs prises aux dates précédentes :

$$X_t = f(X_{t-1}, X_{t-2}, \dots).$$

Il n'est généralement pas nécessaire de prendre en compte tout le passé de la série et on peut le plus souvent se limiter à p valeurs :

$$X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-p}) + E_t$$

où $\{E_t\}$ est un bruit blanc.

3.3.1 Processus autoregressif AR(1)

Le processus autorégressif d'ordre 1, noté AR(1), est stationnaire et vérifie l'équation

$$X_t = \varphi X_{t-1} + E_t = \varphi B(X_t) + E_t$$

où E_t est un bruit blanc de variance σ^2 .

On a donc, à toute date t ,

$$E(X_t) = \mu, \quad V(X_t) = \sigma_X^2.$$

Espérance, variance, stationnarité : On a

$$\begin{aligned} \text{E}(X_t) &= \varphi \text{E}(X_{t-1}) + \text{E}(E_t) = \varphi \text{E}(X_{t-1}), \\ \text{et } \text{V}(X_t) &= \varphi^2 \text{V}(X_{t-1}) + \text{V}(E_t) = \varphi^2 \text{V}(X_{t-1}) + \sigma^2. \end{aligned}$$

La stationnarité implique que

$$\begin{cases} \mu = \varphi\mu \\ \sigma_X^2 = \varphi^2\sigma_X^2 + \sigma^2 \end{cases} \Rightarrow \begin{cases} \mu = 0 \text{ ou } \varphi = 1 \\ (1 - \varphi^2)\sigma_X^2 = \sigma^2 \end{cases} \Rightarrow \begin{cases} \mu = 0 \\ |\varphi| < 1 \end{cases} .$$

Remarque : La contrainte $\mu = 0$ n'est pas très forte puisqu'on passe d'un processus d'espérance μ à un processus d'espérance nulle par simple translation.

Autocorrélation : On a

$$\begin{aligned} X_t &= \varphi X_{t-1} + E_t = \varphi(\varphi X_{t-2} + E_{t-1}) + E_t = \varphi^2 X_{t-2} + \varphi E_{t-1} + E_t \\ &= \varphi^h X_{t-h} + \sum_{k=0}^{h-1} \varphi^k E_{t-k} \end{aligned}$$

donc

$$\begin{aligned} \gamma_h &= \text{Cov}(X_t, X_{t-h}) = \text{E}(X_t, X_{t-h}) = \text{E}\left(\left(\varphi^h X_{t-h} + \sum_{k=0}^{h-1} \varphi^k E_{t-k}\right) X_{t-h}\right) \\ &= \varphi^h \text{E}(X_{t-h}^2) = \varphi^h \text{V}(X_{t-h}) = \varphi^h \sigma_X^2. \end{aligned}$$

Le coefficient d'autocorrélation vaut donc

$$\rho_h = \frac{\gamma_h}{\gamma_0} = \frac{\varphi^h \sigma_X^2}{\sigma_X^2} = \varphi^h.$$

Ce dernier résultat implique que, puisque $|\varphi| < 1$, les corrélogrammes auront des allures décroissantes amorties :

$$\varphi = 0.9 \quad \varphi = 0.3$$

$$\varphi = -0.9 \quad \varphi = -0.3$$

Le processus "oublie" progressivement les valeurs passées.

Autocorrélation partielle : En ce qui concerne les autocorrélations, la définition même du processus nous montre que

$$- \tau_1 = \rho_1 = 1 \text{ (voir plus haut),}$$

– $\tau_h = 0$ pour $h > 1$: la régression de X_t sur $\{X_{t-1}, X_{t-2}, \dots, X_{t-h+1}\}$ vaut

$$X_t^* = \varphi X_{t-1}$$

et donc

$$X_t - X_t^* = E_t,$$

or les E_t sont non corrélés par hypothèse.

Les corrélogrammes partiels auront donc la forme suivante :

$$\varphi = 0.9 \quad \varphi = -0.9$$

Cette forme de corrélogramme partiel rappelle que, conditionnellement à X_{t-1} , la connaissance de X_{t-2} n'apporte aucune information sur X_t .

3.3.2 Processus autoregressif AR(2)

Un processus AR(2) est un processus stationnaire qui vérifie une équation de la forme

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + E_t = \varphi_1 B(X_t) + \varphi_2 B^2(X_t) + E_t.$$

Dans un tel modèle, l'influence du passé se manifeste par une régression linéaire sur les deux valeurs antérieures.

Selon les valeurs de φ_1 et φ_2 , il n'est pas toujours possible de trouver un processus stationnaire vérifiant cette équation.

Dans le cas des processus stationnaires, on peut montrer que

$$\rho_h \xrightarrow{h \rightarrow \infty} 0.$$

Plus précisément, l'autocorrélation décroît à vitesse exponentielle ; on a par exemple

$$\rho_1 = -\frac{\varphi_1}{1 + \varphi_2}, \quad \rho_2 = -\frac{\varphi_1^2 - \varphi_2(1 + \varphi_2)}{1 + \varphi_2}$$

Le corrélogramme d'un AR(2) est assez semblable à celui d'un AR(1).

D'autre part, la régression sur les deux dates précédentes donne

$$\begin{aligned} X_t^* &= \varphi_1 X_{t-1} + \varphi_2 X_{t-2} \\ \text{et donc } X_t - X_t^* &= E_t. \end{aligned}$$

L'autocorrélation partielle τ_h est donc nulle dès que h est supérieur à 2 :

$$h > 2 \quad \Rightarrow \quad \tau_h = 0.$$

3.3.3 Processus autoregressif AR(p)

De façon générale, un processus AR(p) est un processus qui dépend linéairement des p valeurs antérieures :

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + E_t$$

où $\{E_t\}$ est un bruit blanc.

On peut, sans restriction de généralité, supprimer le terme constant et obtenir un modèle de la forme

$$X_t = \sum_{1 \leq k \leq p} \varphi_k X_{t-k} + E_t \quad \Leftrightarrow \quad \Phi(B)X_t = E_t$$

où Φ est le polynôme de degré p dont les coefficients sont $(1, -\varphi_1, \dots, -\varphi_p)$.

Autocorrélation : On montre que les autocorrélations sont solutions des équations Yule et Walker

$$\rho_h + \sum_{1 \leq k \leq p} \varphi_k \rho_{h-k} = 0.$$

Comme pour les processus AR(1) et AR(2), on montre que ρ_h décroît exponentiellement.

Autocorrélation partielle : Dans un tel processus, X_t et X_{t-p+1} sont indépendants conditionnellement aux valeurs intermédiaires $\{X_{t-1}, X_{t-2}, \dots, X_{t-p}\}$ et donc

$$h > p \quad \Rightarrow \quad \tau_h = 0.$$

La valeur à la date t dépend des p dates précédentes et pas des autres.

Cette propriété sert à l'identification des modèles et à déterminer l'ordre p d'un processus AR(p) au vu du corrélogramme partiel.

Estimation des coefficients : Par ailleurs, les coefficients φ_k du modèle peuvent s'estimer à partir de l'estimation des ρ_h en utilisant le système de Yule et Walker :

$$\rho_h = \sum_{k=1}^p \varphi_k \rho_{h-k} \quad \text{pour } h \geq 1$$

où par d'autres méthodes comme le maximum de vraisemblance ou les moindres carrés.

Syntaxe SAS : Pour l'analyse d'une série autorégressive d'ordre 2 par SAS, la syntaxe est :

```
proc ARIMA data=TAB;
  identify var=X;
  estimate p=2;
```

La valeur de p n'étant pas connue *a priori*, on exécute d'abord une fois la procédure ARIMA sans l'instruction '**estimate**' pour évaluer p au vu des corrélogrammes puis on exécute une seconde fois la procédure, cette fois avec l'instruction '**estimate**'.

3.4 Modèle moyenne mobile

Le terme E_t est souvent présenté comme un "choc" (une innovation pour les économètres) : il rend compte d'un élément nouveau dans l'évolution du processus. On peut envisager que ces chocs (non-correlés et d'espérances nulles) aient des effets sur l'évolution du processus non seulement à la date à laquelle ils se produisent mais aussi à des dates ultérieures.

3.4.1 Processus moyenne mobile MA(1)

Un processus moyenne mobile d'ordre 1, noté MA(1), est un processus stationnaire de la forme

$$X_t = E_t + \theta E_{t-1}$$

où $\{E_t\}$ est un bruit blanc de variance σ^2 .

Pour un tel processus, on a

$$\begin{aligned} E(X_t) &= E(E_t) + \theta E(E_{t-1}) = 0, \\ V(X_t) &= V(E_t) + \theta^2 V(E_{t-1}) = (1 + \theta^2)\sigma^2. \end{aligned}$$

Autocorrélation : On a

$$\gamma_1 = \text{Cov}(X_t, X_{t-1}) = \text{Cov}(E_t + \theta E_{t-1}, E_{t-1} + \theta E_{t-2}) = \theta V(E_{t-1}) = \theta \sigma^2$$

Le coefficient d'autocorrélation d'ordre 1 vaut donc

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta \sigma^2}{(1 + \theta^2)\sigma^2} = \frac{\theta}{1 + \theta^2}.$$

Pour $h > 1$, on a

$$\gamma_h = \text{Cov}(X_t, X_{t-h}) = \text{Cov}(E_t + \theta E_{t-1}, E_{t-h} + \theta E_{t-h-1}) = 0$$

donc les coefficients d'autocorrélations d'ordre supérieur à 1 sont nuls :

$$h > 1 \quad \Rightarrow \quad \rho_h = 0.$$

Si on inverse la formule de ρ_1 , on obtient

$$(1 + \theta^2)\rho_1 = \theta \quad \Leftrightarrow \quad \theta^2 \rho_1 - \theta + \rho_1 = 0$$

équation en θ qui n'admet de solution que si

$$\Delta = 1 - 4\rho_1^2 > 0 \quad \Leftrightarrow \quad \rho_1^2 < 1/4 \quad \Leftrightarrow \quad |\rho_1| < 1/2.$$

Pour un processus MA(1), l'autocorrélation d'ordre 1 est inférieure à 1/2 en valeur absolue.

Les corrélogrammes d'un processus MA(1) seront donc de la forme

$$\theta = 0.9 \quad \theta = -0.9$$

Autocorrélation partielle : Le calcul du coefficient d'autocorrélation partielle est plus complexe, il se résout en manipulant l'équation

$$X_t = (I + \theta B)E_t \Leftrightarrow \frac{1}{I + \theta B}X_t = E_t \Leftrightarrow \left(\sum_{k=0}^{\infty} (-\theta)^k B^k \right) X_t = E_t$$

ce qui permet de donner une autre équation d'un processus MA(1) sous la forme

$$X_t = E_t - \left(\sum_{k=1}^{\infty} (-\theta)^k B^k \right) X_t = E_t - \sum_{k=1}^{\infty} (-\theta)^k X_{t-k}.$$

On peut ainsi montrer que

$$\tau_h = \frac{(-\theta)^h (\theta^2 - 1)}{1 - \theta^{2(h+1)}}$$

qui nous donne des corrélogrammes partiels de la forme

$$\theta = 0.9 \quad \theta = 0.3$$

$$\theta = -0.9 \quad \theta = -0.3$$

3.4.2 Processus moyenne mobile MA(2)

Un processus MA(2) est défini par une équation du type

$$X_t = E_t + \theta_1 E_{t-1} + \theta_2 E_{t-2} = (I + \theta_1 B + \theta_2 B^2)E_t$$

on a

$$\begin{aligned} E(X_t) &= E(E_t) + \theta_1 E(E_{t-1}) + \theta_2 E(E_{t-2}) = 0, \\ V(X_t) &= V(E_t) + \theta_1^2 V(E_{t-1}) + \theta_2^2 V(E_{t-2}) = (\theta_2^2 + \theta_1^2 + 1)\sigma^2, \\ \text{Cov}(X_t, X_{t-1}) &= \text{Cov}(E_t + \theta_1 E_{t-1} + \theta_2 E_{t-2}, E_{t-1} + \theta_1 E_{t-2} + \theta_2 E_{t-3}) \\ &= \theta_1 V(E_{t-1}) + \theta_2 \theta_1 V(E_{t-2}) = \theta_1(1 + \theta_2)\sigma^2, \\ \text{Cov}(X_t, X_{t-2}) &= \text{Cov}(E_t + \theta_1 E_{t-1} + \theta_2 E_{t-2}, E_{t-2} + \theta_1 E_{t-3} + \theta_2 E_{t-4}) \\ &= \theta_2 V(E_{t-2}) = \theta_2 \sigma^2, \\ \text{Cov}(X_t, X_{t-h}) &= \text{Cov}(E_t + \theta_1 E_{t-1} + \theta_2 E_{t-2}, E_{t-h} + \theta_1 E_{t-h-1} + \theta_2 E_{t-h-2}) \\ &= 0 \quad \text{pour } h > 2. \end{aligned}$$

On a donc

$$\rho_1 = \frac{\theta_1(1 + \theta_2)}{(\theta_2^2 + \theta_1^2 + 1)}, \quad \rho_2 = \frac{\theta_2}{(\theta_2^2 + \theta_1^2 + 1)}, \quad \rho_h = 0 \text{ pour } h > 2.$$

Enfin, on peut montrer que l'autocorrélation partielle décroît de façon exponentielle :

$$\tau_h \xrightarrow{h \rightarrow \infty} 0.$$

3.4.3 Processus moyenne mobile MA(q)

Il s'agit d'un processus vérifiant l'équation

$$X_t = E_t + \theta_1 E_{t-1} + \dots + \theta_q E_{t-q} = \Theta(B)E_t$$

où Θ est le polynôme de degré q dont les coefficients sont $\{1, \theta_1, \dots, \theta_q\}$.

Dans un tel modèle, on suppose de plus que l'influence des chocs passés se manifeste au travers d'une fonction linéaire.

Un tel modèle est appelé moyenne mobile d'ordre q car X_t est une moyenne mobile (en anglais moving average) appliquée aux variables aléatoires $E_t, E_{t-1}, \dots, E_{t-q}$. Le terme moyenne est à prendre dans un sens très large dans la mesure où la somme des coefficients θ_k n'est pas nécessairement égale à 1.

Autocorrélation : Pour un tel processus, on peut montrer que l'autocorrélation ρ_h est nulle pour $h > q$:

$$\rho_h = \begin{cases} \frac{\theta_h + \sum_{k=1}^{k=q-h} \theta_k \theta_{h-k}}{1 + \sum_{k=1}^{k=h} \theta_k^2} & \text{si } h \leq q \\ 0 & \text{sinon} \end{cases}$$

Cette propriété est évidemment très précieuse pour l'identification du modèle et la détermination de l'ordre q d'un processus MA(q).

Autocorrélation partielle : Les autocorrélations partielles τ_h d'un processus moyenne mobile d'ordre q ont un comportement semblable à celui des autocorrélations ρ_h d'un processus autoregressif de même ordre : elle s'amortissent à vitesse exponentielle.

Remarque : Un processus autorégressif d'ordre 1 peut s'exprimer sous forme de moyenne mobile en inversant l'équation :

$$(I - \varphi B)X_t = E_t \quad \Leftrightarrow \quad X_t = \frac{1}{I - \varphi B} E_t = \left(\sum_{k=0}^{\infty} \varphi^k B^k \right) E_t,$$

on obtient ainsi une moyenne mobile d'ordre q infini dont les coefficients décroissent exponentiellement

$$\text{AR}(1) \sim \text{MA}(\infty).$$

Syntaxe SAS : Pour l'analyse d'une série moyenne-mobile d'ordre 1 par SAS, la syntaxe est :

```
proc ARIMA data=TAB;
  identify var=X;
  estimate q=1;
```

3.5 Modèle autorégressif – Moyenne mobile

3.5.1 Processus ARMA(p, q)

On peut bien évidemment envisager de combiner les deux modèles précédents en introduisant

- une dépendance du processus vis-à-vis de son passé : modèle AR(p),
- un effet retardé des chocs : modèle MA(q).

Un tel modèle, appelé autorégressif - moyenne mobile (ARMA), est caractérisé par le paramètre p de la partie autorégressive et le paramètre q de la partie moyenne mobile. Un processus ARMA(p, q) vérifie l'équation

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + E_t + \theta_1 E_{t-1} + \dots + \theta_q E_{t-q}$$

soit

$$\Phi(B)X_t = \Theta(B)E_t.$$

Le traitement d'un tel processus est plus complexe que celui des 2 précédents. On peut cependant montrer que ses autocorrélations et ses autocorrélations partielles sont des fonctions amorties tendant vers 0 en valeur absolue à vitesses exponentielles.

On peut ainsi dresser un tableau comparatif des corrélogrammes et corrélogrammes partiels des processus

Processus	Autocorrélation ρ_h	Autocorrélation partielle τ_h
AR(p)	amortie	nulle pour $h > p$
MA(q)	nulle pour $h > q$	amortie
ARMA(p, q)	amortie	amortie

L'analyse des corrélogrammes constitue un des outils privilégiés dans l'identification du modèle.

Syntaxe SAS : Pour l'analyse d'une série autorégressive (2) - moyenne-mobile (1) par SAS, la syntaxe est :

```
proc ARIMA data=TAB;
  identify var=X;
  estimate p=2 q=1;
```

3.5.2 Modèles ARIMA et SARIMA

La condition de stationnarité des modèles présentés ici n'est évidemment pas toujours convenable. On peut intégrer certains types de non-stationnarités en élargissant le modèle ARMA. Ces élargissements consistent en une série d'opérations préalables visant à éliminer la tendance ou la saisonnalité.

Modèle ARIMA : On a vu que si un processus X_t admet une tendance polynomiale de degré d , le processus différencié d fois est stationnaire :

$$Y_t = \Delta^d X_t = (I-B)^d X_t.$$

Le modèle ARIMA revient à appliquer un modèle ARMA sur le processus différencié :

$$Y_t = \text{ARMA}(p, q) \quad \Leftrightarrow \quad X_t = \text{ARIMA}(p, d, q).$$

L'équation d'un modèle ARIMA(p, d, q) est donc donnée par

$$\Phi(B)Y_t = \Theta(B)E_t \quad \Leftrightarrow \quad \Phi(B)\Delta^d X_t = \Theta(B)E_t$$

où Φ et Θ sont deux polynômes de degrés respectifs p et q .

Le 'I' de ARIMA signifie 'integrated' comme réciproque de la différenciation.

Evidemment, le degré d n'est généralement pas connu. Pour le déterminer on peut agir par tâtonnements ou avoir recours à des tests de stationnarité : puisqu'un processus ARMA(p, q) est stationnaire, on cherche d tel qu'on puisse accepter l'hypothèse de stationnarité pour le processus $Y_t = \Delta^d X_t$. De façon générale, on se réfère à un principe de parcimonie et cherche la valeur satisfaisante minimale de d . Cette discussion sera menée plus en détail dans le chapitre concernant le choix de modèles.

Modèle SARIMA : La saisonnalité est un autre facteur de non-stationnarité. On a vu qu'une façon simple d'éliminer une saisonnalité de période ω consiste à appliquer l'opérateur Δ_ω .

$$Z_t = \Delta_\omega X_t = (I-B^\omega)X_t. = X_t - X_{t-\omega}.$$

De façon général, on peut supposer que l'influence des chocs se transmet entre dates distantes d'un nombre entier de périodes selon un processus ARIMA(P, D, Q) :

$$\Phi_\omega(B^\omega)\Delta_\omega^D X_t = \Theta_\omega(B^\omega)U_t$$

et que ces chocs eux-mêmes suivent un modèle ARIMA(p, d, q)

$$\Phi(B)\Delta^d U_t = \Theta(B)E_t$$

où $\{E_t\}$ est un bruit blanc.

Un tel modèle est noté SARIMA(p, d, q) \times (P, D, Q) $_\omega$ et son équation générale est

$$\Phi(B)\Phi_\omega(B^\omega)\Delta^d\Delta_\omega^D X_t = \Theta(B)\Theta_\omega(B)E_t$$

où Φ , Φ_ω , Θ et Θ_ω sont des polynômes de degrés respectifs p, P, q, Q .

Exemple : Un modèle SARIMA(0,1,1) \times (0,1,1)₁₂ a pour objet d'éliminer l'effet annuel par un modèle de la forme

$$\Delta_{12}X_t = (1 - \theta B^{12})U_t$$

puis à poser pour $\{U_t\}$ un modèle ARIMA(0,1,1) :

$$\Delta U_t = (1 - \vartheta B)E_t.$$

On obtient ainsi

$$\Delta\Delta_{12}X_t = (1 - \vartheta B)(1 - \theta B^{12})E_t$$

soit

$$X_t - X_{t-1} - X_{t-12} + X_{t-13} = E_t - \vartheta E_{t-1} - \theta E_{t-12} + \theta\vartheta E_{t-13}.$$

Syntaxe SAS : Le principe des modèles est de se ramener à un modèle ARMA(p, q) en opérant des différenciations sur la série. Dans SAS cette opération se fait en indiquant les différenciations à effectuer dans la procédure ARIMA elle-même. Ainsi le programme

```
proc ARIMA data=TAB;
  identify var=X(1,12);
```

analyse non pas la série X_t mais la série $\Delta\Delta_{12}X_t$. Et le programme

```
proc ARIMA data=TAB;
  identify var=X(1,1,12);
```

analyse la série $\Delta\Delta\Delta_{12}X_t = \Delta^2\Delta_{12}X_t$.

3.6 Identification et estimation des paramètres

3.6.1 Identification du modèle : méthode de Box & Jenkins

Le choix entre les différents modèles présentés ici (AR(p), MA(q), ARMA(p, q), ARIMA(p, d, q), SARIMA, *etc.*) ne peut généralement pas se faire a priori. On est le plus souvent réduit à des tâtonnements par un système d'essais / erreurs.

Une méthodologie générale a été proposée par Box & Jenkins qui peut se résumer dans l'organigramme suivant :

Il s'agit d'une méthodologie "pas à pas" qui implique la remise en cause de chaque modèle envisagé jusqu'à obtenir un modèle acceptable.

Un modèle est acceptable lorsqu'il prend en compte toute la structure de la partie aléatoire et ne laisse qu'un bruit blanc. Par exemple, on considère qu'un modèle ARMA(p, q) est acceptable si on peut accepter l'hypothèse selon laquelle $\{E_t\} = \{X_t - \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \theta_1 E_{t-1} + \dots + \theta_q E_{t-q}\}$ est un bruit blanc

Il faut bien noter qu'il est tout à fait possible d'obtenir plusieurs modèles satisfaisants. On a alors besoin de critères de comparaison de modèle comme le R^2 ou le critère d'Akaike. Cette discussion sera menée plus en détail dans le chapitre concernant le choix de modèles.

3.6.2 Détermination du type et de l'ordre du modèle

On s'intéresse ici à l'identification d'un processus stationnaire : on se limite aux processus ARMA(p, q).

On suppose donc que la saisonnalité et la tendance ont été supprimées. Comme on l'a vu précédemment, cette élimination peut s'opérer au moyen de différentiations de différents types :

- $\Delta^d = (I-B)^d$ pour éliminer une tendance polynomiale de degré d ,
- $\Delta_\omega = (I-B^\omega)$ pour éliminer une saisonnalité de période ω .

Exemple : Pour une série de données mensuelles présentant une tendance apparemment linéaire, on combine les deux filtres Δ et Δ_ω pour éliminer la tendance et une saisonnalité de période 12. Cela revient à travailler sur la série

$$Y_t = \Delta\Delta_{12}X_t = \Delta(X_t - X_{t-12}) = X_t - X_{t-1} - X_{t-12} + X_{t-13}.$$

Si la série x_t comprend T valeurs, la série transformée n'en contient plus que $T - 13$ c'est à dire que l'analyse du modèle ARMA(p, q) se fait sur une série raccourcie de plus d'un an.

Une fois cette manipulation effectuée, on tente de reconnaître le type de processus auquel on a affaire en étudiant les corrélogrammes. Leurs formes théoriques étant connues, on choisit alors un modèle autorégressif, moyenne mobile ou ARMA. La lecture des corrélogrammes se fait en recherchant les chutes

- dans le corrélogramme partiel pour un AR(p),
- dans le corrélogramme pour un MA(q),
- *etc.*

3.6.3 Estimation des paramètres et prévision

Méthodes des moments : On a vu dans le paragraphe sur les méthodes empiriques comment estimer les autocorrélations et on a donné dans ce chapitre leurs valeurs théoriques pour les processus classiques. La méthode des moments est une des plus utilisées : pour un modèle donné, on sait que les autocorrélations ρ_h et τ_h dépendent des paramètres $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_d$ selon des équations théoriques connues :

$$(\rho_1, \rho_2, \dots, \tau_1, \tau_2, \dots) = F(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_d).$$

Or on sait facilement estimer les autocorrélations ρ_h et τ_h , il suffit donc d'inverser les équations pour estimer les paramètres :

$$(\widehat{\varphi}_1, \dots, \widehat{\varphi}_p, \widehat{\theta}_1, \dots, \widehat{\theta}_d) = F^{-1}(\widehat{\rho}_1, \widehat{\rho}_2, \dots, \widehat{\tau}_1, \widehat{\tau}_2, \dots).$$

Il n'est pas nécessaire d'explicitier ici la forme des résultats.

Maximum de vraisemblance : L'autre méthode classique d'estimation est le maximum de vraisemblance. Elle nécessite de faire des hypothèses sur la loi du bruit blanc $\{E_t\}$. L'hypothèse la plus fréquemment retenue est qu'il s'agit d'un bruit blanc gaussien de variance σ^2 .

On obtient alors des estimateurs du maximum de vraisemblance des paramètres $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_d$ et de σ .

Cette méthode peut donner lieu à des calculs très lourds aussi bien mathématiquement que numériquement à cause de la forme très "tourmentée" des fonctions de vraisemblance. Pour éviter des temps de calculs colossaux, on a le plus souvent recours à des algorithmes qui utilisent des approximations plus ou moins fortes.

Test sur les paramètres : Sous l'hypothèse que le bruit blanc est gaussien, on peut montrer que les estimateurs $\hat{\varphi}_j$ et $\hat{\theta}_l$ sont approximativement gaussiens. On peut donc effectuer des tests sur les paramètres en utilisant la loi de Student :

$$\frac{\hat{\varphi}_j}{\sqrt{\hat{V}(\hat{\varphi}_j)}} \approx \mathcal{T}.$$

Dans le cadre d'un modèle ARMA(p, q), on peut remarquer que le test sur un paramètre φ_p

$$\mathbf{H}_0 : " \varphi_p = 0 " \quad \text{contre} \quad \mathbf{H}_A : " \varphi_p \neq 0 "$$

est équivalent au test sur les modèles

$$\mathbf{H}_0 : " \text{ARMA}(p, q) " \quad \text{contre} \quad \mathbf{H}_A : " \text{ARMA}(p-1, q) "$$

puisque φ_p est le paramètre du dernier terme de la partie autorégressive du modèle.

Prévision : Une fois les paramètres $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_d$ estimés, il est possible de donner une prédiction de la série à la date t en utilisant l'équation du modèle

$$\begin{aligned} \Phi(B)X_t &= \Theta(B)E_t \\ \text{soit} \quad X_t &= \{\varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \theta_1 E_{t-1} + \dots + \theta_q E_{t-q}\} + E_t \end{aligned}$$

Les X_{t-k} ($1 \leq k \leq p$) sont connus à la date $t-1$ et la partie restante (E_t) est nulle en espérance.

On prend donc comme prévision de X_t la quantité

$$\hat{X}_t = \hat{\varphi}_1 X_{t-1} + \dots + \hat{\varphi}_p X_{t-p} + \hat{\theta}_1 \hat{E}_{t-1} + \dots + \hat{\theta}_q \hat{E}_{t-q},$$

où les \hat{E}_t sont définis par

$$\hat{E}_t = X_t - \hat{X}_t,$$

les \hat{E}_{t-k} ($1 \leq k \leq q$) étant calculables à la date t .

Evidemment, la prévision ne peut être envisagée qu'après une étape de validation du modèle.

3.6.4 Validation

Test du bruit blanc : Dans l'analyse des séries chronologiques par processus, le bruit blanc joue un rôle particulier puisque c'est un processus sans aucune structure. Quand, pour un processus $\{X_t\}$, on a éliminé toute tendance, toute saisonnalité et toute dépendance vis-à-vis du passé, il reste un processus $\{E_t\}$ complètement imprévisible sur lequel il n'y a plus grand chose à dire.

Le test préliminaire concerne justement ce processus. Quand on étudie une série x_t la première hypothèse testée est

$$\mathbf{H}_0 : \text{"}\{X_t\} \text{ est un bruit blanc"}$$

Si on accepte cette hypothèse, l'analyse de la série est virtuellement achevée : la série étudiée n'a aucune structure.

On peut citer

- le test "Portmanteau" ("fourre-tout" en anglais) fondé sur la somme des carré des autocorrélations :

$$Q_1(\{x_t\}) = T \sum_{h=1}^H \hat{\rho}_h^2(x)$$

où T est le nombre d'observations, $\hat{\rho}_h(x)$ l'autocorrélation d'ordre h estimée sur la série $\{x_t, 1 \leq t \leq T\}$ et H une quantité suffisamment grande.

- le test fondé sur la statistique de Ljung et Box :

$$Q_2(\{x_t\}) = T(T+2) \sum_{h=1}^H \frac{1}{T-h} \hat{\rho}_h^2(x)$$

Sous l'hypothèse d'un bruit blanc gaussien, ces deux statistiques suivent des lois du χ^2 à $H - p - q$ degrés de libertés.

Validation : Les tests de bruit blanc permettent également de "valider" le modèle. Si le processus est bien un ARMA(p, q), le processus résiduel $\{\hat{E}_t\} = \{X_t - \hat{X}_t\}$ doit être un bruit blanc. On accepte (ou rejette) cette hypothèse en effectuant un des tests fondés sur une statistique $Q(\{\hat{e}_t\})$. Pour cela, il faut donc estimer les autocorrélations du processus $\{\hat{E}_t\}$.

On ne peut envisager d'effectuer des prédictions qu'une fois qu'on a accepté l'hypothèse

$$\mathbf{H}_0 : \text{"}\{E_t\} = \{X_t - \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \theta_1 E_{t-1} + \dots + \theta_q E_{t-q}\} \text{ est un bruit blanc"}$$

Chapitre 4

Choix de modèles

4.1 Analyse spectrale : Recherche des périodicités

On a vu dans les paragraphes précédents qu'il pouvait y avoir des composantes saisonnières ou des composantes cycliques dans la partie déterministe d'un processus aléatoire, et comment on peut les modéliser en introduisant des termes en sinus et cosinus, à condition de connaître les périodes de ces composantes. On peut utiliser le corrélogramme pour détecter les périodicités éventuelles. On peut aussi faire une analyse spectrale. L'idée de l'analyse spectrale est empruntée au domaine de la physique et consiste à utiliser l'hypothèse que la série chronologique est composée de sinus et de cosinus de différentes fréquences (analogie avec le son par exemple). Les deux approches seront comparées, chacune d'elles ayant des avantages et des inconvénients.

4.1.1 Rappels sur les fonctions déterministes

Fonctions périodiques

Une fonction $f(t)$ est périodique si elle satisfait l'égalité :

$$f(t) = f(t + kp) \quad \forall t$$

k est un entier positif ou négatif. p est la période. C'est le plus petit nombre pour lequel l'égalité précédente est satisfaite.

Les fonctions périodiques les plus courantes sont les fonctions sinus et cosinus. Par exemple :

$$A \sin \omega t \text{ et } B \cos \omega t$$

sont deux fonctions périodiques, chacune de période

$$p = \frac{2\pi}{\omega}.$$

La quantité $\omega = \frac{2\pi}{p}$ est appelée fréquence angulaire, la quantité $\frac{1}{p}$ est appelée fréquence (c'est la fréquence angulaire divisée par 2π), et les constantes A et B sont appelées amplitudes.

Ces deux fonctions jouent un rôle majeur dans l'étude des fonctions périodiques, car Joseph Fourier a montré que toute fonction périodique bien conformée de période p peut s'exprimer comme une somme (qui peut être infinie) de sinus et de cosinus et que l'on appelle "série de Fourier" :

$$f(t) = \frac{1}{2}a_0 + \sum_{r=1}^{\infty} \left[a_r \sin \frac{2\pi r}{p}t + b_r \cos \frac{2\pi r}{p}t \right] = \sum_{r=0}^{\infty} A_r e^{i\omega_r t} = \sum_{r=0}^{\infty} A_r e^{2\pi i f_r t}$$

avec

$$A_r = \begin{cases} \frac{1}{2}(a_r - ib_r) & r > 0 \\ \frac{1}{2}a_0 & r = 0 \\ \frac{1}{2}(a_{|r|} + ib_{|r|}) & r < 0 \end{cases}, \quad \omega_r = \frac{2\pi r}{p} \text{ et } f_r = \frac{r}{p}.$$

Par analogie, ω_r est appelée la fréquence angulaire, et $|A_r| = \frac{1}{2}[(a_r^2 + b_r^2)]^{\frac{1}{2}}$ est appelée amplitude. On peut interpréter chacun des termes de cette série. Le premier terme correspondant à $r = 0$ est un terme constant, le deuxième ($r = 1$) représente les oscillations de période de base p , le troisième ($r = 2$) représente les oscillations de période $\frac{p}{2}$, le quatrième, celles de période $\frac{p}{3}$, etc...

On peut aussi introduire la notion d'énergie totale sur l'intervalle de temps $(-\frac{p}{2}, \frac{p}{2})$:

$$\int_{-\frac{p}{2}}^{\frac{p}{2}} f^2(t) dt = \|f(t)\|^2 = \frac{p}{2} \left[\frac{1}{2}a_0^2 + \sum_{r=1}^{\infty} (a_r^2 + b_r^2) \right] = p \sum_{r=0}^{\infty} c_r^2$$

avec

$$c_r = \left[\frac{1}{2}(a_r^2 + b_r^2) \right]^{\frac{1}{2}}$$

On appelle l'égalité qui donne l'énergie totale "**relation de Parseval**". c_r^2 est la contribution de la composante de période $\frac{p}{r}$ à l'énergie totale.

On associe à cette notion d'énergie totale, la notion de puissance totale :

$$\text{Puissance totale} = \frac{\text{énergie totale}}{p} = \sum_{r=0}^{\infty} c_r^2$$

Spectre discret

On peut tracer le graphe des c_r^2 en fonction de la période, on obtient alors un diagramme en bâtons, que l'on appelle spectre discret.

Extension aux fonctions non périodiques

Une fonction non périodique peut être considérée comme une fonction périodique de période infinie. On peut reprendre la série précédente en faisant tendre p vers l'infini. La somme se transforme alors en intégrale sous certaines conditions mathématiques :

$$f(t) = \int_0^\infty [a(\omega) \sin \omega t + b(\omega) \cos \omega t] d\omega = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty G(\omega) e^{i\omega t} d\omega$$

L'égalité précédente n'est valable que si la fonction f est absolument intégrable.

La fonction $G(\omega)$ est appelée "**transformée de Fourier**" de $f(t)$. C'est sur cette fonction qu'est fondée l'analyse spectrale. On peut montrer que si la fonction $f(t)$ a de bonnes propriétés mathématiques, on peut exprimer sa transformée de Fourier en fonction de $f(t)$ par la relation :

$$G(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty f(t) e^{-i\omega t} dt$$