1 Les processus linéaires générals

Soit $\{Y_t\}$ une série que est observée et une série de bruit $\{e_t\}$ qui n'est pas observée. Le bruit satisfait les hypothèses classiques.

Nous voudrions écrire $\{Y_t\}$ comme une combinaison linéaire, c'est-à-dire que *nous expliquons la série observée par les perturbations aléatoires passées.*

$$Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots \tag{1}$$

Si c'est une somme infinie, conditions spécifiques doivent être placés sur les coefficients pour que la somme soit convergeant.

1.1 Example

Soit le cas que 1 soit sur la forme

$$Y_t = e_t + \psi e_{t-1} + \psi^2 e_{t-2} + \cdots = \sum_{i=0}^{\infty} \psi^i e_{t-i} \quad \psi^0 = 1 \tag{2}$$

Ici, la condition pour 2 soit convergeant est $|\psi| < 1$

- $\mathbb{E}(Y_t) = 0$
- $Var(Y_t) = \frac{\sigma_e^2}{1-\psi^2}$
- $Cov(Y_t, Y_{t-1}) = \frac{\psi\sigma_e^2}{1-\psi^2}$ et $Cov(Y_t, Y_{t-k}) = \frac{\psi^k\sigma_e^2}{1-\psi^2}$
- $Corr(Y_t, Y_{t-1}) = \psi$ et $Corr(Y_t, Y_{t-k}) = \psi^k$

La processus est stationnaire, l'autocovariance ne dépend pas au temps t, seulement à décalage du temps (lag) k

2 Les processus Moyen Mobile (Moving Average) MA(q)

En cas général

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \tag{3}$$

Nous appelons le grade (rank) du modèle le plus distant indice q

2.1 MA(1)

$$Y_t = e_t - \theta e_{t-1} \tag{4}$$

- $\mathbb{E}(Y_t) = 0$
- $Var(Y_t) = \sigma_e^2(1 + \theta^2)$
- $Cov(Y_t, Y_{t-1}) = -\theta\sigma_e^2$ et $Cov(Y_t, Y_{t-2}) = 0$
- $Corr(Y_t, Y_{t-1}) = \frac{-\theta}{1+\theta^2}$ et $Corr(Y_t, Y_{t-2}) = 0$

3 Les processus autogrégressifs (Autoregressive) AR(p)

Ici, nous expliquons la série ses valeurs précédents et une perturbation aléatoire.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t \tag{5}$$

e_t est un bruit et nous pouvons l'imaginer comme une innovation à la série, il n'est pas explicable par rapport aux valeurs précédents de la série. Ainsi, à chaque instante de temps t on suppose que c'est indépendant de $Y_{t-1}, Y_{t-2}, ..etc.$

3.1 AR(1)

$$Y_t = \phi Y_{t-1} + e_t \tag{6}$$

Nous supposons que la série est stationnaire, et de moyenne zéro

- $Var(Y_t) = \frac{\sigma_e^2}{1-\phi^2}$
- $Cov(Y_t, Y_{t-1}) = \frac{\phi\sigma_e^2}{1-\phi^2}$ et $Cov(Y_t, Y_{t-k}) = \phi^k \frac{\sigma_e^2}{1-\phi^2}$
- $Corr(Y_t, Y_{t-1}) = \phi$ et $Corr(Y_t, Y_{t-k}) = \phi^k$

Remarque
De la variance on pourrais suspecter la condition $|\phi| < 1$ mais pour le justifier rigoureusement on doit d'aller plus loin sur les propriétés d'un modèle AR.

3.2 L'inversion d'un modèle AR(1) et la condition pour la stationnarité

Nous rendrons le modèle AR(1) sous la forme rétroactif (c'est à dire, nous chercherons à faire la recursion sur les valeurs précédents de la série)

$$Y_t = \phi Y_{t-1} + e_t = e_t + \phi \underbrace{(\phi Y_{t-2} + e_{t-1})}_{Y_{t-1}}$$

Si nous répétons la substitution $k-1$ fois, nous arrivons à

$$Y_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots + \phi^{k-1} e_{t-k+1} + \phi^k e_{t-k} \tag{7}$$

Ce qui resemble à le modèle 2, et en plus, si nous laisserions $k \to \infty$, alors nous récupérons modèle 2 et la condition pour la stationnarité.

3.3 AR(2)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t \tag{8}$$

Je définir un opérateur de décalage du temps (lag operator) L, c.a.d. $L(Y_t) = Y_{t-1}$ et en general $L^k(Y_t) = Y_{t-k}$ alors 8 pourrait être écrit:

$$(1 - \phi_1 L - \phi_2 L^2)Y_t = e_t \tag{9}$$

par 9 on appelle polynôme caractéristique:

$$1 - \phi_1 z - \phi_2 z^2 = 0 \tag{10}$$

Pour que $(1 - \phi_1 L - \phi_2 L^2)$ soit invertible, il faut placer les conditions sur les racines de 10. Après, changer le côté de 9

$$\psi(L) = \frac{1}{1 - \phi_1 L - \phi_2 L^2} \tag{11}$$

$$Y_t = \psi(L)e_t \tag{12}$$

Alors pour inverser le modèle AR à MA on a besoin de developer 11 aux puissants de L par la série géométrique,

$$\frac{1}{1 - \phi_1 L - \phi_2 L^2} = 1 + \psi_1 L + \psi_2 L^2 + \cdots$$

Ainsi de 12

$$Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots \tag{13}$$

Nous arrivons à la forme inversée infini. Pour preciser la forme des coefficients ψ_i à 13 , soit r_1, r_2 les racines du polynôme caractéristique et $R_{1,2} = \frac{1}{r_{1,2}}$, on a la factorisation

$$\frac{1}{1 - \phi_1 z - \phi_2 z^2} = \frac{1}{(1 - R_1 z)(1 - R_2 z)} = \frac{A}{1 - R_1 z} + \frac{B}{1 - R_2 z}$$

$$A = \frac{r_2}{r_2 - r_2} \quad B = \frac{-r_1}{r_2 - r_1}$$

$$\frac{1}{1 - \phi_1 z - \phi_2 z^2} = \sum_{i=0}^{\infty} (AR_1^i + BR_2^i) z^i \tag{14}$$

C'est facile de verifier par 14 que $\psi_0 = 1$

Par cette derivation, il est clair que on a besoin d' étudier les racines du polynôme caractéristique.

3.3.1 Etude des racines du polynôme caractéristique

Pour AR(2) d'être invertible et alors stationnaire il faut que ses racines sont hors du circle unitaire. Nous montrerons quelques conditions pour les coefficients de 8 Nous savons que.

$$r_{1,2} = \frac{\phi_2\sqrt{\phi_1 + 4\phi_2}}{-2\phi_2}$$

Et

$$R_1 = \frac{1}{r_2} = \frac{\phi_1 - \sqrt{\phi_1 + 4\phi_2}}{2} \quad R_2 = \frac{\phi_1 + \sqrt{\phi_1 + 4\phi_2}}{2}$$

Par 14 est claire qu'il faut $|R_i| < 1$ for $i = 1, 2$

Racines reels Pour avoir deux solutions reels il faut $\phi_1^2 + 4\phi_2 \geq 0$

$$-1 < \frac{\phi_1 - \sqrt{\phi_1 + 4\phi_2}}{2} < \frac{\phi_1 + \sqrt{\phi_1 + 4\phi_2}}{2} < 1$$

en faisant les calcules nécessaires on arrive à les conditions

$$\phi_2 - \phi_1 < 1 \quad et \quad \phi_2 + \phi_1 < 1$$

Racines complexes Pour avoir deux solutions complexes il faut $\phi_1^2 + 4\phi_2 < 0$ Ici, les $R_{1,2}$ sont complex conjugates et

$$|R_1| = |R_2| < 1 \Leftrightarrow |R_1|^2 < 1$$

$$|R_1|^2 = \frac{\phi_1^2 + (-\phi_1^2 - 4\phi_2)}{4} = -\phi_2$$

Et alors $|\phi_2| < 1$

4 Les modèles des séries non stationnaires

N'importe quand une série n'a pas moyenne constant, elle nonstationnaire. Pour une série d'être stationnaire il faut que son moyen et sa covariance ne dépend pas au temps t

Example 1 Soit un processus $X_t, \mathbb{E}[X_t] = 0$ et une fonction de moyenne non-constant $\mu_t, \mathbb{E}[\mu_t] = c(t)$

$$Y_t = \mu_t + X_t$$

est une série non-stationnaire. On peut aller plus loin dans cette "catégorie" des modèles non-stationnaires en écrivant l'example à la forme

$$Y_t = a + \delta t + \psi(L)e_t \tag{15}$$

$\psi(L)$ comme à 11 , une partie de la série a une tendance linaire et l'autre est stationnaire.

Devoir 1

Calculer : l'espérance, la variance, covariance et la correlation de la processus

$$Y_t = a + \delta t + e_t$$

Example 2
$$Y_t = 3Y_{t-1} + e_t$$
En transformant la série en forme rétroactif
$$Y_t = e_t + 3e_{t-1} + 3^2 e_{t-2} + \cdots + 3^{k-1}e_1 + 3^k Y_0$$
On remarque que les coefficients ne se tend pas vers zéro.

Example 3 (Marche Aléatoire avec dérive)
$$Y_t = Y_{t-1} + \delta + e_t \tag{16}$$
Ici, $\delta \in \mathbb{R}$ est la dérive, si $\delta = 0$ on reviens à la cas classique d'une marche aléatoire.

Devoir 2
Calculer : l'espérance, la variance, la covariance et la correlation du marche aléatoire avec dérive. Verifier qu'il s'agit d'un processus non-stationnaire.

En général, dans une série non-stationnaire, l'influence des valeurs precedents augmente. On peut appeler cette comportement explosif.

4.1 Les processus ARIMA

Dans cette section nous nous occupons avec les cas ou la source de la non-stationnarité est la violation des conditions du polynôme characteristic comme il était montré par les examples précédents. Pour un processus $AR(1)$ il faut que la racine du polynôme soit entre $1, -1$ strictement. On peut écrire l'example 3 en utilisant l'opérateur de décalage du temps (lag) comme
$$(1 - L)Y_t = \delta + e_t$$
Ici, il est clair qu'il y a une racine unitaire, et alors ce n'est pas possible de retourner le modèle en forme rétroactif convergante. Mais si on prendre les premières difference on voit que
$$\Delta Y_t = Y_t - Y_{t-1} = \delta + e_t \tag{17}$$
qui donne une série stationnaire. Les même argument s'applique aussi à la cas générale de la forme
$$(1 - L)Y_t = \delta + \psi(L)e_t \tag{18}$$
et on appelle cette catégorie des processus non-stationnaires, *processus de racine unitaire (unit root processes)* et le processus qui sort des differences, *processus intégré d'ordre 1*, I(1)

On peut aussi imaginer un processus ARMA qui contient les racines unitaires en multiple fois.
$$(1 - L)^k(1 - R_1L)(1 - R_2L)\cdots(1 - R_{p-k}L)Y_t = \psi(L)e_t \tag{19}$$
ou $R_1, \cdots R_{p-k}$ sont les reciprocals des racines différents, et la multitude de racine unitaire est k. On s'appelle alors 19 un processus $ARIMA(p, d, q)$ avec $d = k$ et q dépendant à la représentation du $\psi(L)$

4.2 IMA(1,1)

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1} \tag{20}$$

Ici, c'est un modèle avec une racine unitaire et un lag au bruit, alors $ARIMA(0,1,1)$ Pour le rendre stationnaire nous prenons les premières differences

$$W_t = \Delta Y_t = e_t - \theta e_{t-1}$$

Maintenant, W_t est un processus $MA(1)$ que nous avons déjà étudié

Équation 20 pourrait être transformé en forme rétroactif

$$Y_t = e_t + (1-\theta)e_{t-1} + \cdots + (1-\theta)e_{t-k+1} + \theta e_{t-k} \tag{21}$$

Devoir
Avec l'aide de 21 montrez

- $\mathbb{E}[Y_t] = 0$
- $Var(Y_t) = [1 + \theta^2 + (1-\theta)^2(t+k)]\sigma_e^2$
- $Cov(Y_t, Y_{t-h}) = 1 + \theta^2 + (1-\theta)^2(t+k-h)$

4.3 ARI(1,1)

$$Y_t = (1-\phi)Y_{t-1} - \phi Y_{t-2} + e_t \tag{22}$$

ou

$$W_t = \Delta Y_t = \phi W_{t-1} + e_t$$

Maintenant, W_t est un processus $AR(1)$ comme aux chapitres précédents. Si nous voulons écrire le polynôme characteristic pour cette processus, nous arrivons à la formule

$$\frac{1}{(1-z)(1-\phi z)} = 1 + \psi_1 z + \psi_2 z^2 + \cdots$$

ou

$$(1 - (1+\phi)z + \phi z^2)(1 + \psi_1 z + \psi_2 z^2 + \cdots) = 1 \tag{23}$$

pour définir les ψ_i nous arrivons au system

$$\begin{aligned} -(1+\phi) + \psi_1 &= 0 \\ \phi - (1+\phi)\psi_1 + \psi_2 &= 0 \end{aligned}$$

et

$$\psi_k = (1+\phi)\psi_{k-1} - \phi\psi_{k-2}, \quad \psi_0 = 1 \tag{24}$$

Devoir Calculer la variance, la covariance et la correlation du processus ARI(1,1) 22

5 Specification du Modèle, la méthodologie Box-Jenkins

Après l'étude théorique des modèles ARMA et ARIMA maintenant nous nous occupons avec les questions pratiques.

- Comment modéliser les donnes comme un processus $ARIMA(p, d, q)$?
- Est-ce que la série est stationnaire?
- Comment choisir p, d, q?

Pour répondre à ces questions à la manière pratique et pragmatique, Box et Jenkins[1] ont développé une methodology en 4 étapes:

1. **Identification** En cet étape on s'occupe de determiner l'ordre du modèle, et le comportement stationnaire (ou non) et saisonnier (ou non). Il y a deux approches pour traiter cette partie

   - L'approche graphique: On plot la série, les ACF, PACF en cas que ACF tend ver zéro très lentement ou pas du tout on soupçonne non-stationnarité. Box et Jenkins conseillent à prendre les premières differences et refaire les graphs en cas qu'ils indiquent stationnarité.Sinon, on reprendre les différences jusqu'à le graphe a l'air stationnaire.

     Pour determiner l'order du modèle p, q il y a quelque règles empiriques à partir de graph de ACF

     (a) Décroissance exponentielle vers zéro : Modèle AR (PACF indique l'ordre p)
     (b) Oscillations amortis décroissants rapidement vers zéro: AR (PACF indique l'ordre p)
     (c) Quelque peu pics, le reste nulle: MA (l'ordre est le nombre des pics)
     (d) Décroissance exponentielle après quelque pas de temps (lags)
     (e) pattern régulière en écarts de temps : saisonnalité

     Un remarque général, ACF nous aide à determiner l'order MA du modèle et PACF l'ordre AR.

   - Des test statistique , voire la procaine séction

   Pour la saisonnalité, nous allons apprendre plus au chapitre 5

2. Estimation du modèle et sélection. Voire section 4.6 du polycopié et feuille d'exercises 5.

[1] G. Box, G. Jenkins, *Time Series Analysis: forecasting and control*, 1970

3. Diagnostique du modèle. Ici, nous nous intéressons principalement au comportement des résidus du modèle. Quels sont le propriétés statistiques des résidus? Hypothèse de normalité, bruit blanc etc. Box-Pierce ou Ljung-Box sont les utiles principales ici. Deux remarques importants
   (a) En testant la correlation des résidus, il n'est possible que de dévoiler un modèle sousparametrisé. Il y a rien à dire pour un modèle surparametrisé.
   (b) Autocorrelation des résidus est possible de faire les autres test statistiques donner les faux résultats.

5.1 Tests statistiques pour la stationnarité

Comme nous avons vu si le polynôme d'une série contient une racine unitaire, la série n'est pas stationnaire. Des premières étapes de la méthodologie Box-Jenkins on peut soupçonner la non-stationnarité par les graphes mais ici notre but c'est de developer des tests statistiques pour tester rigoureusement les séries. Souvent il y a deux cas de non-stationnarité et c'est pas toujours facile à distinguer, trend-stationnaire (c'est-à-dire la série n'est pas stationnaire par rapporte de DEFINITION mais elle est stationnaire autour une tendance deterministic) et unit-root non-stationnaire.

5.1.1 Augumented Dickey-Fuller test de racine unitaire

Dickey-Fuller test

$$Y_t = aY_{t-1} + e_t \tag{25}$$

Avec les hypothèse:

$$\begin{aligned} H_0 &: a = 1 \\ H_1 &: a < 1 \end{aligned}$$

sous l'hypothèse H_0 nous avons un marche aléatoire. DF-test est le t-test de H_0 et on construit la statistique pour l'estimateur $\hat{a}$ de a

$$\hat{\tau} = \frac{\hat{a} - 1}{se(\hat{a})}$$

$\hat{\tau}$ ne suis pas Gaussian et la distribution analytic pourrait être rétablie par Hamilton p489 et table B6 *Remarques*

1. e_t doit être bruit blanc
2. La distribution de $\hat{\tau}$ reste la même aussi pour un modèle ARMA

Augmented Dickey-Fuller test

$$\begin{aligned} Y_t &= aY_{t-1} + X_t \\ X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_k X_{t-k} + e_t \end{aligned} \tag{26}$$

Avec les hypothèse:

$$H_0 : a = 1$$
$$H_1 : a < 1$$

Sous H_0 nous avons ΔY_t d'être un processus stationnaire $AR(k)$

$$Y_t - Y_{t-1} = X_t$$

et Y_t un $AR(k+1)$ avec polynôme caractéristique

$$(1 - az)(1 - \phi_1 z - \cdots \phi_k z^k) = 0$$

Augmented Dickey-Fuller avec contant et dérive

$$\begin{aligned} Y_t &= \mu + \delta t + aY_{t-1} + X_t \\ X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_k X_{t-k} + e_t \end{aligned} \tag{27}$$

Avec les hypothèse:

$$H_0 : a = 1$$
$$H_1 : a < 1$$

Sous H_0

$$\Delta Y_t = X_t + \delta$$

Remarques

1. C'est ADF avec dérive et constant que R utilise comme method standard pour la command `adf.test()` dans la librairie `tseries`.
2. k paramètre en R est précisément l'ordre du modèle AR pour X_t
3. Si on met $k = 0$ en R elle comptera le test DF
4. La distribution pour la statistique $\hat{\tau}$ de chaque modèle change mais il sont disponibles analytiquement à Hamilton p.502 et p.528-529

Il y a encore les tests Philips-Perron et KPSS qui pourront être utilisés en complement de ADF pour verifier la stationnarité ou non d'une série.