Chapitre 2

Modélisation des séries stationnaires

Dans ce chapitre nous présentons comment modéliser une série, qui une fois tendance et saisonnalité supprimées, est stationnaire.

2.1 Les processus autorégressifs AR(p)

L'idée de ce premier processus est que l'observation au temps t s'explique linéairement par les observations précédentes.

Définition. On dit que (X_t) est un processus autorégressif d'ordre p s'il s'écrit

$$X_t = \delta + \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t$$

où ε_t est un bruit blanc et c'est l'innovation contenue dans le processus au temps t. L'observation X_t au temps t est la somme d'un choc aléatoire à l'instant t indépendant de l'historique, et d'une fonction linéaire de son passé, qui peut être vue comme la prédiction de X_t à partir des p dernières observations passées.

Les coefficients φ_j doivent vérifier la condition suivante pour assurer la stationnarité du processus : les solutions du polynôme caractéristique du processus $\Phi(z) = 1 - \varphi_1 z - \cdots \varphi_p z^p$ sont en module > 1. Si $\delta = 0$, le processus est centré.

Proposition

-La moyenne du processus AR(p) est : $E(X_t) = \frac{\delta}{1-\varphi_1-\cdots\varphi_p}$

-La variance est $\gamma_0 = \varphi_1\gamma_1 + \cdots + \varphi_p\gamma_p + \sigma^2$, et l'autocovariance est $\gamma_h = \varphi_1\gamma_{h-1} + \cdots + \varphi_p\gamma_{h-p}$

Le polynôme caractéristique de cette relation de récurrence est $z^p - \varphi_1 z^{p-1} - \cdots - \varphi_p = z^p \Phi(z)$

Les racines λ_i de ce polynôme sont les inverses des racines du polynôme Φ, et sont donc de module strictement inférieur à 1. En supposant que Φ a toute ses racines distinctes, les solutions de l'équation de récurrence précédente sont de la forme :$\gamma_h = a_1 \lambda_1^h + \cdots + a_p \lambda_p^h$.

Puisque $|\lambda_i| < 1$, l'autocovariance décroit exponentiellement avec h. Ces résultats s'étendent directement à l'autocorrélation d'un processus AR(p).

2.1.1 L'autocorrélation partielle

Nous avons vu que l'autocorrélation est un outil important pour la détection de non stationnarité un autre outil important est l'autocorrélaion partiel. La corrélation partielle entre deux variables est la corrélation qui reste en éliminant l'impact de tous les autres variables. Soit un processus stochastque X_t, le coefficient d'autocorrélation partielle entre les deux variables X_1et X_n est

$$r_{X_2,\ldots,X_{n-1}}(X_1, X_n) = corr\left(X_1 - P_{X_2,\ldots,X_{n-1}}(X_1), X_n - P_{X_2,\ldots,X_{n-1}}(X_n)\right)$$

où corr est le coefficient de corrélation classique et P est la projection. La fonction d'autocorrélation partielle r$_h$ d'un processus stationnaire est définie par :

$$\begin{aligned} r_1 &= \rho_1 \\ r_h &= r_{-h} \\ r_h &= r_{X_2,\ldots,X_{n-1}}(X_1, X_{h+1}) \end{aligned}$$

Pour le processus AR(p) l'autocorrélation partielle est nulle à tout ordre strictement supérieur à p, et vaut φ_p à l'ordre p.

2.2 Les processus moyenne mobile MA(q)

Définition. On appelle moyenne mobile (Moving Average) d'ordre q un processus de la forme

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}.$$

avec $\{\varepsilon_t\} \backsim BB\left(0, \sigma^2\right)$.

Proposition

-La moyenne du processus MA(q) est : $E\left(X_t\right) = \mu$

- La variance $\gamma_0 == \left(1 + \theta_1^2 + \cdots + \theta_q^2\right)\sigma^2$, et pour l'autocovariance on a le résultat suivant $\gamma_h = 0$, pour $h > q$.

L'autocorrélation est donc également nulle au dessus de l'ordre q. On retrouve le même comportement que l'autocorrélation partielle d'un AR(p).

L'autocorrélation partielle d'un processus moyenne mobile est compliquée à calculer, et sans grand intérêt. Mais, elle tend vers 0 à vitesse exponentielle lorsque h tend vers l'infini.

Ce comportement symétrique entre les processus MA et AR est du à la propriété suivante :

Proposition. Un processus autorégessif est un processus moyenne mobile d'ordre infini, et réciproquement un processus moyenne mobile est un processus autorégressif d'ordre infini.

2.3 Les processus mixtes ARMA(p,q)

Définition. Un processus autorégressif moyenne mobile d'ordres p et q est de le forme :

$$X_t = \delta + \sum_{j=1}^{p} \varphi_j X_{t-j} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

avec $\{\varepsilon_t\} \backsim BB\left(0, \sigma^2\right)$.

La stationnarité d'un ARMA(p,q) est assurée lorsque toutes les racines du polynôme $\Phi(z)$ sont de module strictement supérieur à 1. L'autocovariance d'un processus ARMA(p,q) évolue selon la même récurrence qu'un AR(p). Ainsi, l'autocovariance (et l'autocorrélation) d'un processus ARMA(p,q) vont tendre exponentiellement vers 0 lorsque h tend vers l'infini, à partir de l'ordre q + 1.

2.4 Estimation, choix de modèle et prévision

Les principaux modèles de séries temporelles ont été définis. A partir d'une série observée, il faut maintenant choisir un modèle, éventuellement plusieurs, estimer ses paramètres et enfin faire des prévisions pour les réalisations futures. Dans le cas où l'on hésite entre plusieurs modèles, des critères de choix de modèles seront utilisés pour sélectionner le meilleur d'entre eux.

2.4.1 Estimation

Dans le cas général, l'estimation des paramètres des modèles ARMA (AR, MA) est faite par maximum de vraisemblance. L'expression de la vraisemblance étant généralement trop complexe pour que l'on puisse obtenir un maximum explicite, des algorithmes numériques sont utilisés.

2.4.2 Choix de modèle

L'étude des autocorrélations et autocorrélations partielles peut conduire à la proposition de modèle. Une fois quelques modèles choisis, et leur paramètres estimés, des critères vont être utilisés pour choisir le modèle qui effectue le meilleur compromis entre :

– ajustement à la série de données,

– complexité du modèle.

Les critères de choix de modèles le plus courant est :

– le critère AIC (Akaïke Information Criterion), qui sera généralement préféré si l'objectif de l'étude est de faire de la prévision, et qui est défini par :

$$AIC = -2\log L(\theta) + 2\nu$$

où L(.) est la vraisemblance du modèle, θ représente les paramètres du modèle et ν le nombre de ces paramètres. Le modèle ayant la plus petite valeur du critère doit être choisi.

2.4.3 Prévision

L'objectif est de prévoir la valeur que va prendre la variable aléatoire X_{n+h}, h étant appelé l'horizon de la prévision, ayant observé la réalisation des variables aléatoires $X_1, ..., X_n$.

Dans le cadre d'une modélisation ARMA, la prévision empirique $\widehat{X}_n(h)$ est la projection de X_{n+h} sur l'espace vectoriel engendré par les variables $X_1, ..., X_n$.

$$\widehat{X}_n(h) = E(X_{n+h}/X_1, ..., X_n)$$

Intervalle de confiance sur la prévision En supposant que le bruit d'innovation est gaussien, les variables aléatoires X_t sont elles aussi gaussiennes, ainsi que l'erreur de prédiction. Donc, il sera possible de construire des intervalles de confiances sur la prédiction, ce que permettent les fonctions du logiciel R.

2.5 Processus non stationnaire : ARIMA et SARIMA

Les processus $ARMA$ exige la stationnarité qui est rarement vérifiée pour les séries économiques, par contre $\Delta X_t = X_t - X_{t-1}$ différence $1^{ère}$, ou à des ordres plus élevé : $\Delta^d X$: différence d'ordre d, devient stationnaire.

Définition. Un processus (X_t) est un processus ARIMA(p, d, q) si le processus $Y_t = \Delta^d X_t$ est un processus ARMA(p, q) .

Soit l'opérateur Δ_T défini par $\Delta_T X_t = X_t - X_{t-T}$.

Définition. Un processus (X_t) est un processus $SARIMA(p, d, q)_T$ si le processus $\mathrm{Y}_t = \Delta_T \Delta^d \mathrm{X}_t$ est un processus ARMA(p, q).

Les processus $SARIMA(p, d, q)_T$ sont bien adaptés aux séries temporelles présentant une période de longueur T et une tendance.

Fonctions R

-La fonction arima.sim(n,list(ar=c($\varphi_1, ..., \varphi_p$),ma=c($\theta_1, ..., \theta_q$)) permet de simuler un processus ARMA.

-Pour simuler un modèle $ARIMA(p, d, q)$ il faut ajouter le composant order=c(p,d,q).

-La fonction arima(serie,order=c(p,d,q)) permet d'estimer les paramètres.