

Modèles ARIMA et SARIMA, prédiction et choix de modèle

Angelina Roche

Executive Master Statistique et Big Data

2018–2019

Plan du cours d'aujourd'hui

Retour sur les modèles ARMA

Modèles ARIMA et SARIMA

Tests de non-stationnarité

Plan

Retour sur les modèles ARMA

Modèles ARIMA et SARIMA

Tests de non-stationnarité

Rappel sur les modèles ARMA

- Un processus stationnaire $\{X_t, t \in \mathbb{Z}\}$ obéit à un modèle $\text{ARMA}(p, q)$ s'il vérifie une équation du type

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}, \forall t \in \mathbb{Z}, \quad (1.1)$$

avec $c \in \mathbb{R}$, $(\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$, $(\theta_1, \dots, \theta_q) \in \mathbb{R}^q$ et $\{Z_t, t \in \mathbb{Z}\}$ un bruit blanc.

Rappel sur les modèles ARMA

- Un processus stationnaire $\{X_t, t \in \mathbb{Z}\}$ obéit à un modèle $\text{ARMA}(p, q)$ s'il vérifie une équation du type

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}, \forall t \in \mathbb{Z}, \quad (1.1)$$

avec $c \in \mathbb{R}$, $(\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$, $(\theta_1, \dots, \theta_q) \in \mathbb{R}^q$ et $\{Z_t, t \in \mathbb{Z}\}$ un bruit blanc.

- L'équation (1.1) peut se réécrire à l'aide de l'opérateur retard B (défini par $BX_t = X_{t-1}$) :

$$X_t = c + \varphi_1 BX_t + \varphi_2 B^2 X_t + \dots + \varphi_p B^p X_t + Z_t + \theta_1 BZ_t + \theta_2 B^2 Z_t + \dots + \theta_q B^q Z_t, \forall t,$$

Rappel sur les modèles ARMA

- Un processus stationnaire $\{X_t, t \in \mathbb{Z}\}$ obéit à un modèle $\text{ARMA}(p, q)$ s'il vérifie une équation du type

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}, \forall t \in \mathbb{Z}, \quad (1.1)$$

avec $c \in \mathbb{R}$, $(\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$, $(\theta_1, \dots, \theta_q) \in \mathbb{R}^q$ et $\{Z_t, t \in \mathbb{Z}\}$ un bruit blanc.

- L'équation (1.1) peut se réécrire à l'aide de l'opérateur retard B (défini par $BX_t = X_{t-1}$) :

$$X_t = c + \varphi_1 BX_t + \varphi_2 B^2 X_t + \dots + \varphi_p B^p X_t + Z_t + \theta_1 BZ_t + \theta_2 B^2 Z_t + \dots + \theta_q B^q Z_t, \forall t,$$

ou

$$X - \varphi_1 BX - \varphi_2 B^2 X - \dots - \varphi_p B^p X = c + Z + \theta_1 BZ + \theta_2 B^2 Z + \dots + \theta_q B^q Z.$$

Rappel sur les modèles ARMA

- Un processus stationnaire $\{X_t, t \in \mathbb{Z}\}$ obéit à un modèle ARMA(p, q) s'il vérifie une équation du type

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}, \forall t \in \mathbb{Z}, \quad (1.1)$$

avec $c \in \mathbb{R}$, $(\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$, $(\theta_1, \dots, \theta_q) \in \mathbb{R}^q$ et $\{Z_t, t \in \mathbb{Z}\}$ un bruit blanc.

- L'équation (1.1) peut se réécrire à l'aide de l'opérateur retard B (défini par $BX_t = X_{t-1}$) :

$$X_t = c + \varphi_1 BX_t + \varphi_2 B^2 X_t + \dots + \varphi_p B^p X_t + Z_t + \theta_1 BZ_t + \theta_2 B^2 Z_t + \dots + \theta_q B^q Z_t, \forall t,$$

ou

$$X - \varphi_1 BX - \varphi_2 B^2 X - \dots - \varphi_p B^p X = c + Z + \theta_1 BZ + \theta_2 B^2 Z + \dots + \theta_q B^q Z.$$

ou encore

$$(I - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)X = c + (I + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)Z,$$

avec I l'opérateur identité ($IX_t = X_t$).

Rappels sur les polynômes

- Un polynôme de degré $p \in \mathbb{N}$ est une expression de la forme :

$$a_0 + a_1X^1 + a_2X^2 + \cdots + a_pX^p$$

où X est l'indéterminée et a_0, \dots, a_p sont les coefficients du polynôme.

- À chaque polynôme, on peut associer une fonction polynomiale définie par :

$$f(z) = a_0 + a_1z + \dots + a_pz^p.$$

- On appelle **racine** d'un polynôme f dans \mathbb{C} tout élément de $z \in \mathbb{C}$ tel que $f(z) = 0$.

Représentation polynomiale d'un ARMA

Définition alternative

Un processus $\{X_t, t \in \mathbb{Z}\}$ obéit à une équation de type $\text{ARMA}(p,q)$ s'il est stationnaire et vérifie une équation du type

$$\Phi(B)X = c + \Theta(B)Z,$$

où Φ est un polynôme de degré p et Θ un polynôme de degré q .

Remarque :

On supposera par la suite que Φ n'admet pas de racine α de module <1 . Dans le cas contraire :

- ▶ Si Φ admet une racine α de module 1 et
 - ▶ si $\Theta(\alpha) = 0$, on peut simplifier l'écriture car $\Phi(z) = (z - \alpha)\tilde{\Phi}(z)$ et $\Theta(z) = (z - \alpha)\tilde{\Theta}(z)$;
 - ▶ si $\Theta(\alpha) \neq 0$ alors le processus X ne peut pas être stationnaire.
- ▶ Si Φ admet une/des racine(s) de module <1 . On peut réécrire l'équation (représentation isospectrale causale inversible).

Exemple de méthode de prévision d'un processus ARMA (I)

- Si Φ n'admet pas de racine de module <1 , alors il existe $a < 1 < b$, et $(c_j)_{j \geq 0}$ tels que

$$\frac{\Theta(z)}{\Phi(z)} = \sum_{j \geq 0} c_j z^j \text{ pour tout } z \text{ tel que } a < |z| < b.$$

- L'équation ARMA $\Phi(B)X = \Theta(B)Z$ se réécrit

$$X_t = \sum_{j \geq 0} c_j B^j Z_t = \sum_{j \geq 0} c_j Z_{t-j}.$$

Exemple de méthode de prévision d'un processus ARMA (II)

- ▶ À partir de la forme $X_t = \sum_{j \geq 0} c_j Z_{t-j}$ on peut écrire, pour un horizon $h > 0$

$$X_{t+h} = \sum_{j \geq 0} c_j Z_{t+h-j}$$

- ▶ Après calcul, on obtient une mise à jour récursive de la prédiction $X_{t+h|t} = \mathbb{E}[X_{t+h}|X_1, \dots, X_t]$:

$$X_{t+h|t+1} = X_{t+h|t} - c_{h-1}(X_{t+1} - X_{t+1|t}).$$

- ▶ En pratique, après avoir choisi les ordre p et q , on estime les coefficients par maximum de vraisemblance, ce qui nous donne une estimation des paramètres φ et θ . On calcule ensuite une estimation de la suite c_0, \dots, c_{h-1} , et on initialise (par exemple $\hat{X}_{1|0} = \dots = \hat{X}_{h|0} = 0$ puis :

$$\hat{X}_{t+h|t+1} = \hat{X}_{t+h|t} - \hat{c}_{h-1}(X_{t+1} - \hat{X}_{t+1|t}).$$

Intervalle de prévision

- ▶ Si le bruit $(Z_t)_{t \in \mathbb{Z}}$ est un bruit blanc gaussien de variance σ^2 , alors

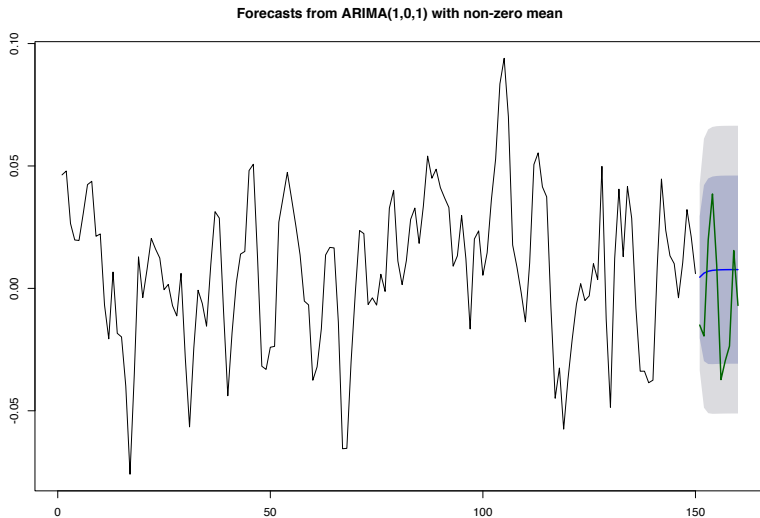
$$\hat{X}_{t+h} - X_{t+h|t} = \sum_{j=0}^{h-1} c_j Z_{t+h-j} \sim \mathcal{N} \left(0, \sigma^2 \sum_{j=0}^{h-1} c_j^2 \right).$$

- ▶ On peut donc dériver un intervalle de prévision pour X_{t+h} :

$$PI(\alpha) = \left[\hat{X}_{t+h|t} \pm q_{1-\alpha/2} \hat{\sigma} \sqrt{\sum_{j=0}^{h-1} \hat{c}_j^2} \right],$$

avec $q_{1-\alpha/2}$ quantile d'ordre $1 - \alpha/2$ de la $\mathcal{N}(0, 1)$.

Illustration ARMA(1,1)



Plan

Retour sur les modèles ARMA

Modèles ARIMA et SARIMA

Tests de non-stationnarité

Définition des modèles ARIMA

Définition

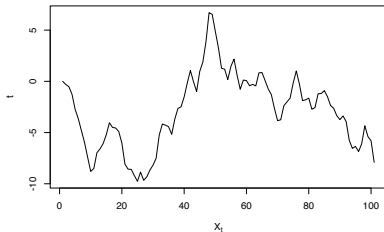
Un processus $(X_t)_{t \in \mathbb{Z}}$ est un processus ARIMA(p, d, q) si $\Delta^d X$ est un processus ARMA(p, q).

Comme $\Delta = 1 - B$, un processus ARIMA(p, d, q) s'écrit :

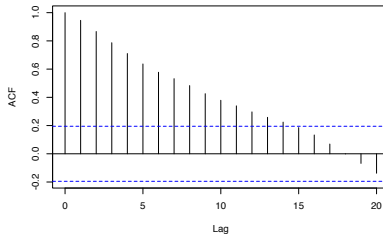
$$(1 - B)^d \Phi(B)X = \Theta(B)Z.$$

Illustration ARIMA(1,1,0)

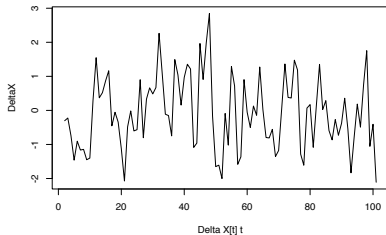
Chronogramme de X



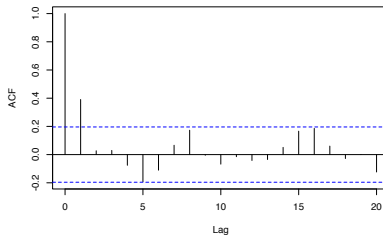
ACF de X



Chronogramme de Delta X



PACF de Delta X



Plan

Retour sur les modèles ARMA

Modèles ARIMA et SARIMA

Tests de non-stationnarité

Test de racine unitaire (Dickey-Fuller)

Cas I : la série ne montre pas de tendance i.e. il est raisonnable de supposer que $\mathbb{E}[X_t]$ constant.

On se place dans le modèle :

$$X_t = \beta_1 + \varphi X_{t-1} + Z_t \text{ avec } (Z_t)_{t \in \mathbb{Z}} \text{ BB gaussien.}$$

On souhaite différencier le cas

$$\mathcal{H}_0 : \varphi = 1 \text{ (non stationnaire)}$$

du cas

$$\mathcal{H}_1 : \varphi < 1 \text{ (stationnaire en général).}$$

Le test est similaire au test de Student $\beta = 1$ contre $\beta < 1$ pour le modèle linéaire $Y_i = \alpha + \beta X_i + \varepsilon_i$.

Test de racine unitaire (Dickey-Fuller)

Cas I : la série ne montre pas de tendance i.e. il est raisonnable de supposer que $\mathbb{E}[X_t]$ constant.

On se place dans le modèle :

$$X_t = \beta_1 + \varphi X_{t-1} + Z_t \text{ avec } (Z_t)_{t \in \mathbb{Z}} \text{ BB gaussien.}$$

On souhaite différencier le cas

$$\mathcal{H}_0 : \varphi = 1 \text{ (non stationnaire)}$$

du cas

$$\mathcal{H}_1 : \varphi < 1 \text{ (stationnaire en général).}$$

Le test est similaire au test de Student $\beta = 1$ contre $\beta < 1$ pour le modèle linéaire $Y_i = \alpha + \beta X_i + \varepsilon_i$.

```
library('urca')
```

```
test1 <- ur.df(X, type="drift", lag=0)
```

Test de racine unitaire (Dickey-Fuller)

Cas II : la série **montre une tendance** linéaire.

On se place dans le modèle :

$$X_t = \beta_1 + \beta_2 t + \varphi X_{t-1} + Z_t \text{ avec } (Z_t)_{t \in \mathbb{Z}} \text{ BB gaussien.}$$

On souhaite différencier le cas

$$\mathcal{H}_0 : \varphi = 1, \beta_2 = 0 \text{ (tendance liée au } \beta_1 \neq 0)$$

du cas

$$\mathcal{H}_1 : \varphi < 1 \text{ (tendance liée au } \beta_2 \neq 0).$$

Test de racine unitaire (Dickey-Fuller)

Cas II : la série **montre une tendance** linéaire.

On se place dans le modèle :

$$X_t = \beta_1 + \beta_2 t + \varphi X_{t-1} + Z_t \text{ avec } (Z_t)_{t \in \mathbb{Z}} \text{ BB gaussien.}$$

On souhaite différencier le cas

$$\mathcal{H}_0 : \varphi = 1, \beta_2 = 0 \text{ (tendance liée au } \beta_1 \neq 0)$$

du cas

$$\mathcal{H}_1 : \varphi < 1 \text{ (tendance liée au } \beta_2 \neq 0).$$

```
test1 <- ur.df(X, type="trend", lag=0)
```

Test de Dickey-Fuller augmenté

Cas I : la série ne montre pas de tendance i.e. il est raisonnable de supposer que $\mathbb{E}[X_t]$ constant.

On se place dans le modèle :

$$\Delta X_t = \beta_1 + \pi X_{t-1} + \zeta_1 \Delta X_{t-1} + \dots + \zeta_{p_1} \Delta X_{t-p+1} + Z_t.$$

On souhaite différencier le cas

$$\mathcal{H}_0 : \beta_1 = 0, \pi = 0 \text{ (non stationnaire)}$$

du cas

$$\mathcal{H}_1 : \pi < 1 \text{ (stationnaire en général).}$$

```
test3 <- ur.df(X, type="drift", lag=p)
```

Test de Dickey-Fuller augmenté

Cas II : la série montre une tendance

On se place dans le modèle :

$$\Delta X_t = \beta_1 + \beta_2 t + \pi X_{t-1} + \zeta_1 \Delta X_{t-1} + \dots + \zeta_{p_1} \Delta X_{t-p_1+1} + Z_t.$$

On souhaite différencier le cas

$$\mathcal{H}_0 : \beta_2 = 0, \pi = 0 \text{ (non stationnaire)}$$

du cas

$$\mathcal{H}_1 : \pi < 1 \text{ (stationnaire à une tendance déterministe près).}$$

```
test4 <- ur.df(X, type="trend", lag=p)
```

Test KPSS

On se place dans le modèle :

$$X_t = R_t + \beta_1 + \beta_2 t + U_t,$$

où

- ▶ $R_t = R_{t-1} + Z_t$, avec $(Z_t)_{t \in \mathbb{Z}}$ bruit blanc gaussien de variance σ_Z^2 ,
- ▶ $(U_t)_{t \in \mathbb{Z}}$ stationnaire.

On teste

$$\mathcal{H}_0 : \sigma_Z = 0 \quad ((R_t + U_t)_{t \in \mathbb{Z}} \text{ stationnaire.})$$

contre

$$\mathcal{H}_1 : \sigma_Z > 0 \quad ((R_t + U_t)_{t \in \mathbb{Z}} \text{ non stationnaire.}).$$

```
testKPSS1 <- ur.kpss(X, type='tau') (cas  $\beta_2 \neq 0$ )
```

```
testKPSS2 <- ur.kpss(X, type='mu') (cas  $\beta_2 = 0$ )
```


Autres modèles (un peu moins) classiques

- ▶ Modèles GARCH : variance conditionnelle $\text{Var}(X_t|X_{t-1})$ non constante.
- ▶ Processus à mémoire longue ($\gamma_X(h) = \alpha h^\beta$ au lieu de $\gamma_X(h) \approx \alpha \gamma^h$ pour les $\text{ARMA}(p,q)$) :
 - ▶ processus auto-similaires (par exemple processus de Lévy)
 - ▶ processus FARIMA ($\Delta^d X$ est un processus $\text{ARMA}(p,q)$ avec $d \in [-1/2, 1/2]$).
- ▶ Séries temporelles multivariées.