

Modèles de prévision

Partie 2 - séries temporelles # 2

Arthur Charpentier

charpentier.arthur@uqam.ca

[http ://freakonometrics.blog.free.fr/](http://freakonometrics.blog.free.fr/)



ÉTÉ 2014

Plan du cours

- **Motivation et introduction aux séries temporelles**
- **Méthodes de lissage**
 - Modèles de régression (Buys-Ballot)
 - Lissage(s) exponentiel(s) (Holt-Winters)
- **Notions générales sur les processus stationnaires**
- **Les processus *SARIMA***
 - Les modèles autorégressifs, $AR(p)$, $\Phi(L)X_t = \varepsilon_t$
 - Les modèles moyennes mobiles, $MA(q)$ (*moving average*), $X_t = \Theta(L)\varepsilon_t$
 - Les modèles autorégressifs et moyenne mobiles, $ARMA(p, q)$,
 $\Phi(L)X_t = \Theta(L)\varepsilon_t$
 - Les modèles autorégressifs, $ARIMA(p, d, q)$, $(1 - L)^d\Phi(L)X_t = \Theta(L)\varepsilon_t$
 - Les modèles autorégressifs, $SARIMA(p, d, q)$,
 $(1 - L)^d(1 - L^s)\Phi(L)X_t = \Theta(L)\varepsilon_t$
 - Prédiction avec un *SARIMA*, ${}_T\hat{X}_{T+h}$

Le modèle $AR(p)$ - autorégressif à l'ordre p

On appelle **processus autoregressif d'ordre p** , noté $AR(p)$, un processus stationnaire (X_t) vérifiant une relation du type

$$X_t - \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t \text{ pour tout } t \in \mathbb{Z}, \quad (1)$$

où les ϕ_i sont des réels et (ε_t) est un bruit blanc de variance σ^2 . (1) est équivalent à l'écriture

$$\Phi(L) X_t = \varepsilon_t \text{ où } \Phi(L) = \mathbb{I} - \phi_1 L - \dots - \phi_p L^p$$

Remarque : En toute généralité, supposons $\Phi(L) X_t = \mu + \varepsilon_t$. Il est possible de se ramener à un processus (1) centré par une simple translation : on pose $Y_t = X_t - m$ où $m = \mu / \Phi(1)$. En effet, $\Phi(L) m = \Phi(1) m$.

Le modèle $AR(p)$ - autorégressif à l'ordre p

Si $\Phi(L) = 1 - (\varphi_1 L + \dots + \varphi_p L)$ et que $|z| \leq 1 \Rightarrow \Phi(z) \neq 0$ (les racines de Φ sont de module strictement supérieur à 1), (X_t) admet une représentation $MA(\infty)$ i.e.

$$X_t = \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \quad \text{où } a_0 = 1, a_k \in \mathbb{R}, \sum_{k=0}^{+\infty} |a_k| < +\infty.$$

On sait que $\Phi(L)X_t = \varepsilon_t$, donc $X_t = \Phi(L)^{-1}(\varepsilon_t)$.

Autocorrélations d'un processus $AR(p)$, $h \mapsto \rho(h)$

Le processus (X_t) s'écrit

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

En multipliant par X_t , on obtient

$$\begin{aligned} X_t^2 &= \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t + \dots + \phi_p X_{t-p} X_t + \varepsilon_t X_t \\ &= \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t + \dots + \phi_p X_{t-p} X_t \\ &\quad + \varepsilon_t (\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t) \\ &= \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t + \dots + \phi_p X_{t-p} X_t + \varepsilon_t^2 \\ &\quad + [\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p}] \varepsilon_t, \end{aligned}$$

d'où, en prenant l'espérance

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \dots + \phi_p \gamma(p) + \sigma^2.$$

Autocorrélations d'un processus $AR(p)$, $h \mapsto \rho(h)$

Le dernière terme étant nul

$$\mathbb{E}([\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p}] \varepsilon_t) = 0$$

car ε_t est supposé indépendant du passé de X_t , $\{X_{t-1}, X_{t-2}, \dots, X_{t-p}, \dots\}$. De plus, en multipliant (2) par X_{t-h} , en prenant l'espérance et en divisant par $\gamma(0)$, on obtient

$$\rho(h) - \sum_{i=1}^p \phi_i \rho(h-i) = 0 \text{ pour tout } h > 0.$$

Autocorrélations d'un processus $AR(p)$, $h \mapsto \rho(h)$

Proposition 1. *Soit (X_t) un processus $AR(p)$ d'autocorrélation $\rho(h)$. Alors*

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(p-1) \\ \rho(p) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \ddots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & & \ddots & \rho(p-2) \\ \rho(2) & \rho(1) & 1 & \ddots & & \rho(p-3) \\ \ddots & & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 & \rho(1) \\ \rho(p-1) & \rho(p-2) & \rho(p-3) & & \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{bmatrix}$$

Ce sont les équations de [Yule-Walker](#). De plus les $\rho(h)$ décroissent exponentiellement vers 0.

Autocorrélations d'un processus $AR(p)$, $h \mapsto \rho(h)$

Preuve : $\forall h > 0$, $\rho(h) - \varphi_1\rho(h-1) - \dots - \varphi_p\rho(h-p) = 0$. Le polynôme caractéristique de cette relation de récurrence est :

$$z^p - \varphi_1 z^{p-1} - \dots - \varphi_{p-1} z - \varphi_p = z^p \left(1 - \frac{\varphi_1}{z} - \dots - \frac{\varphi_{p-1}}{z^{p-1}} - \frac{\varphi_p}{z^p} \right) = z^p \Phi\left(\frac{1}{z}\right),$$

avec $\Phi(L)X_t = \varepsilon_t$ et $\Phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$.

Proposition 2. *Pour un processus $AR(p)$ les autocorrélations partielles ψ sont nulles au delà de rang p , $\psi(h) = 0$ pour $h > p$.*

Preuve : Si (X_t) est un processus $AR(p)$ et si $\Phi(L)X_t = \mu + \varepsilon_t$ est sa représentation canonique, en notant $\psi(h)$ le coefficient de X_{t-h} dans $EL(X_t | X_{t-1}, \dots, X_{t-h})$ alors,

$$X_t = \mu + \underbrace{\varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}}_{\in \mathcal{L}(1, X_t, \dots, X_{t-p}) \subset \mathcal{L}(1, X_t, \dots, X_{t-h})} + \varepsilon_t$$

Autocorrélations d'un processus $AR(p)$, $h \mapsto \psi(h)$

... de telle sorte que

$$\begin{aligned} EL(X_t | X_{t-1}, \dots, X_{t-h}) &= \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + EL(\varepsilon_t | X_{t-1}, \dots, X_{t-h}) \\ &= \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + 0 \end{aligned}$$

Aussi, si $h > p$, le coefficient de X_{t-h} est 0. Et si $h = p$, le coefficient de X_{t-p} est $\varphi_p \neq 0$.

Proposition 3. *Pour un processus $AR(p)$ les autocorrélations partielles ψ est non nulle au rang p , $\psi(p) \neq 0$.*

Le modèle $AR(1)$ - autorégressif à l'ordre 1

La forme général des processus de type $AR(1)$ est

$$X_t - \phi X_{t-1} = \varepsilon_t \text{ pour tout } t \in \mathbb{Z},$$

où (ε_t) est un bruit blanc de variance σ^2 .

Remark si $\phi = \pm 1$, le processus (X_t) n'est pas stationnaire. Par exemple, pour $\phi = 1$, $X_t = X_{t-1} + \varepsilon_t$ peut s'écrire

$$X_t - X_{t-h} = \varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_{t-h+1},$$

et donc $\mathbb{E}(X_t - X_{t-h})^2 = h\sigma^2$. Or pour un processus stationnaire, il est possible de montrer que $\mathbb{E}(X_t - X_{t-h})^2 \leq 4V(X_t)$. Puisqu'il est impossible que pour tout h , $h\sigma^2 \leq 4\text{Var}(X_t)$, le processus n'est pas stationnaire.

Ici, si $\phi = 1$, (X_t) est une **marche aléatoire**.

Remark si $|\phi| < 1$ alors on peut inverser le polynôme, et

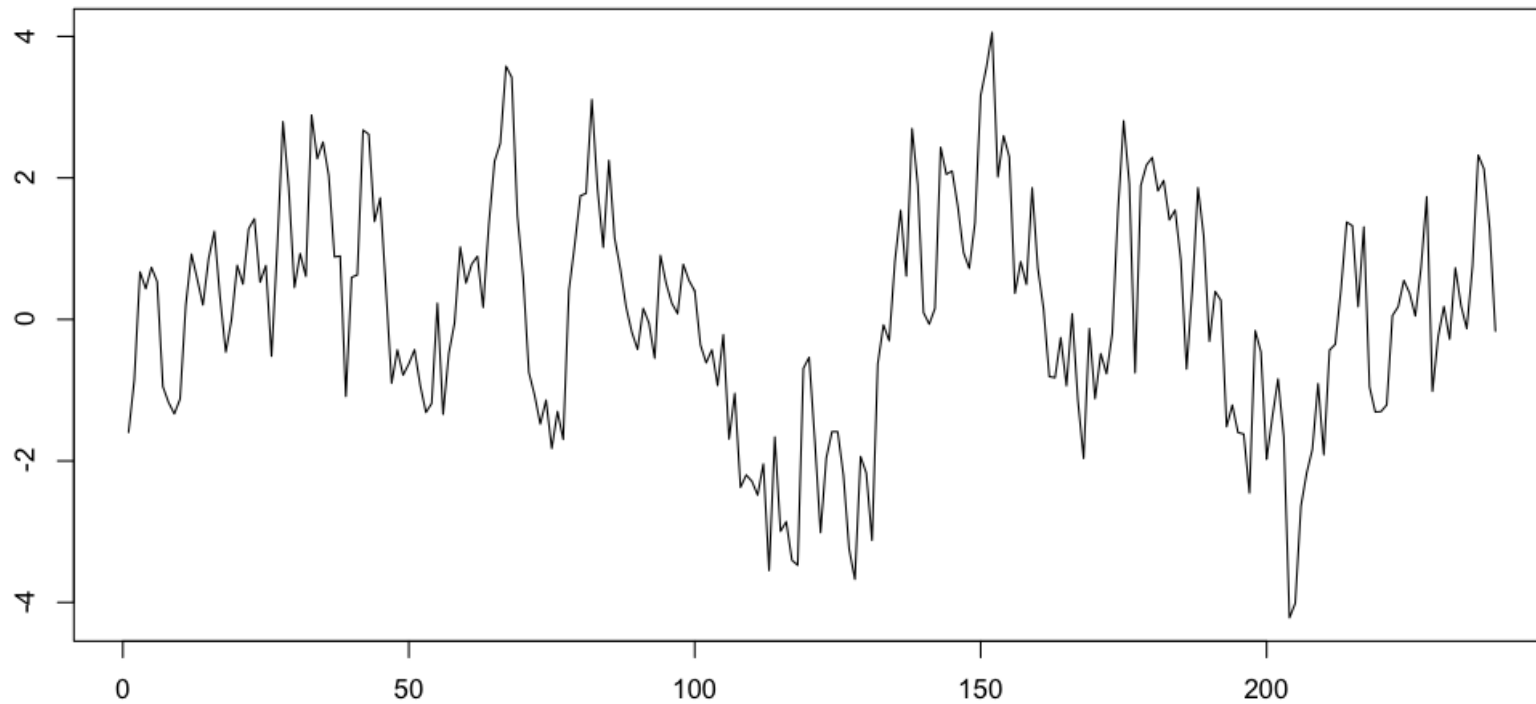
$$X_t = (1 - \phi L)^{-1} \varepsilon_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \text{ (en fonction du passé de } (\varepsilon_t) \text{)}. \quad (3)$$

Proposition 4. *Si (X_t) est stationnaire, la fonction d'autocorrélation est donnée par $\rho(h) = \phi^h$.*

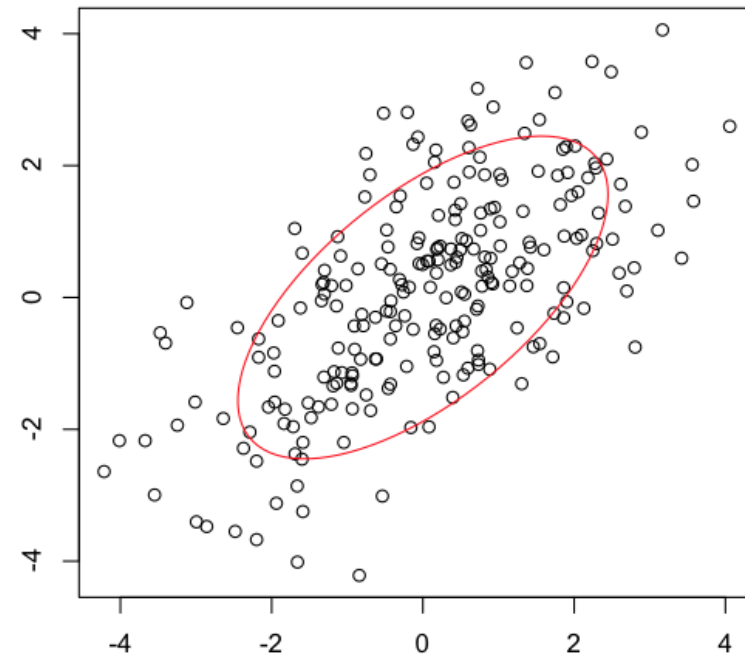
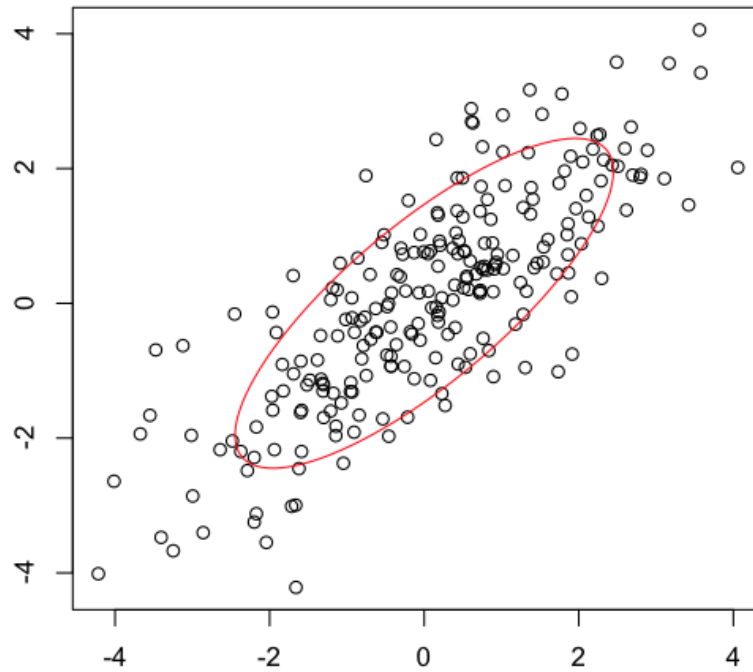
Preuve : $\rho(h) = \phi \rho(h-1)$

```
> X=arima.sim(n = 240, list(ar = 0.8),sd = 1)
> plot(X)
> n=240; h=1
> plot(X[1:(n-h)],X[(1+h):n])
> library(ellipse)
> lines(ellipse(0.8^h), type = 'l',col="red")
```

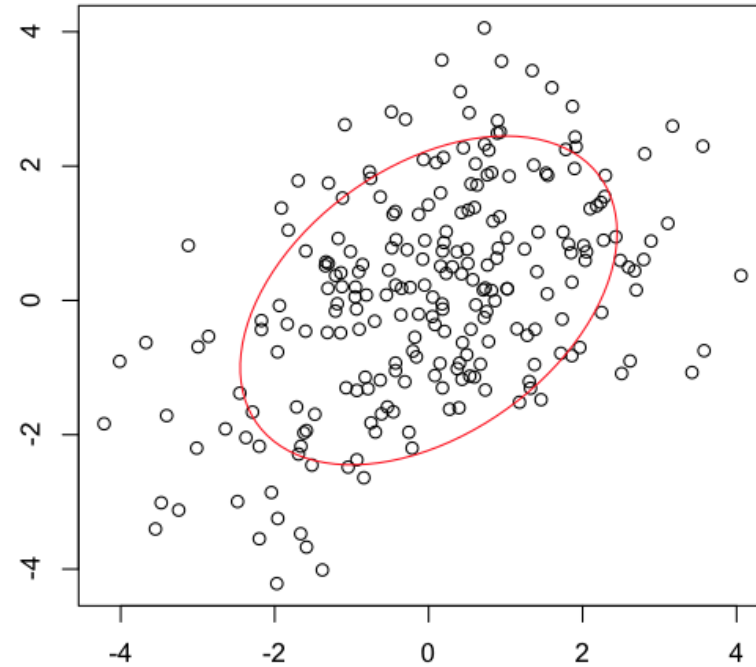
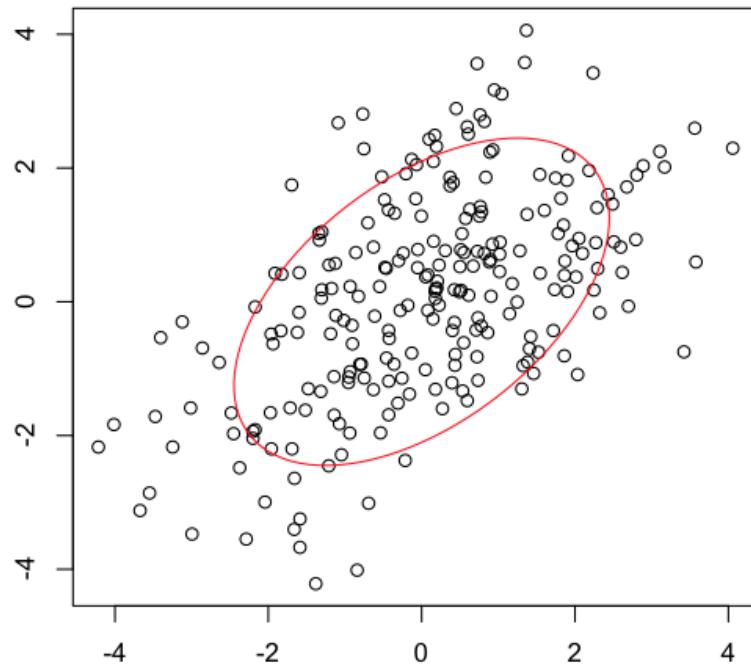
Le modèle $AR(1)$ - autorégressif à l'ordre 1



Le modèle $AR(1)$, $\rho(1)$ et $\rho(2)$

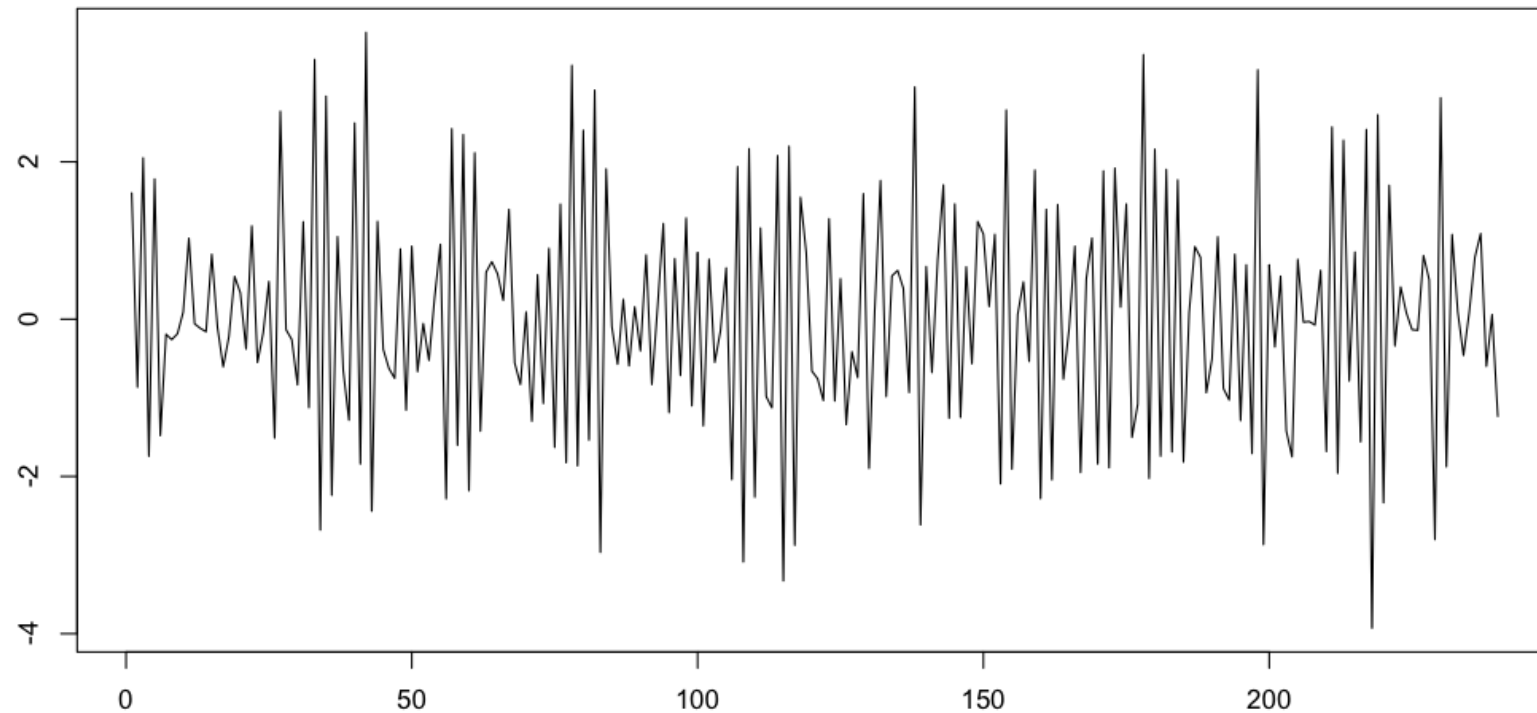


Le modèle $AR(1)$, $\rho(3)$ et $\rho(4)$

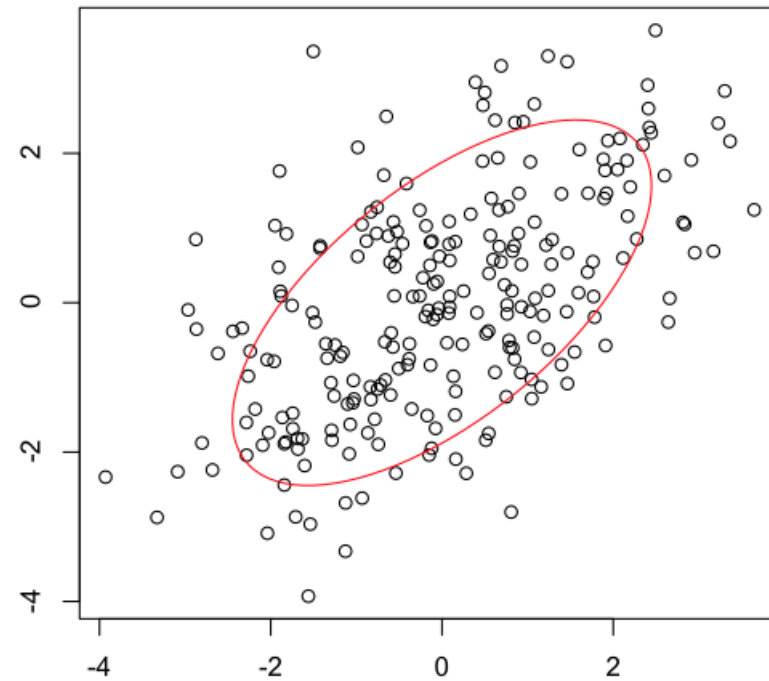
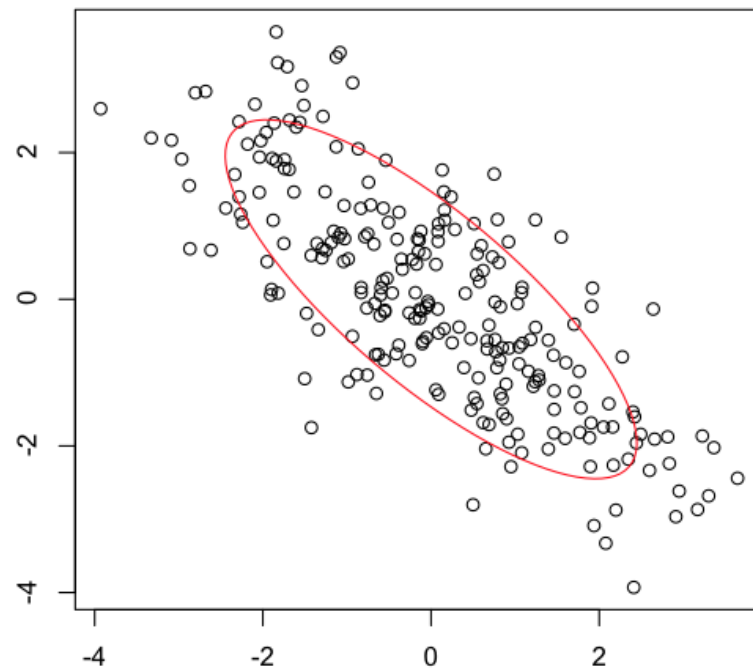


Le modèle $AR(1)$ - autorégressif à l'ordre 1

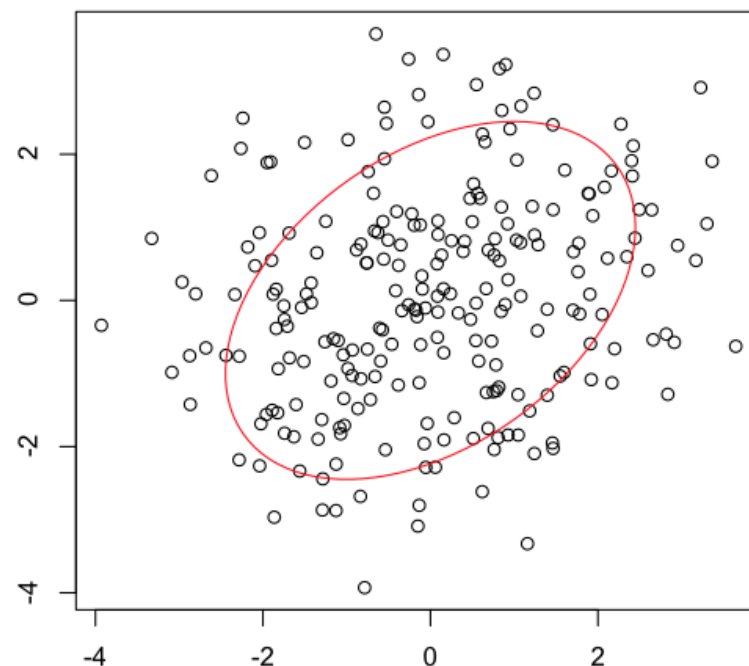
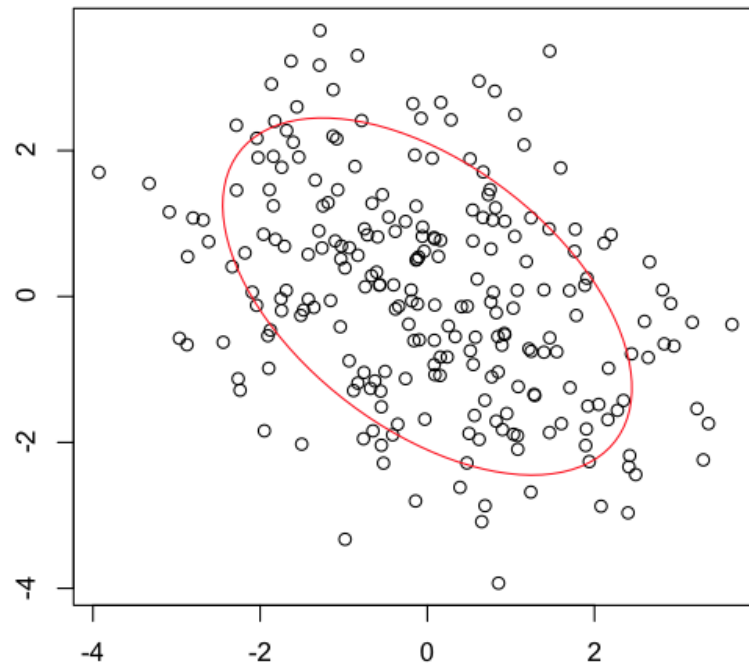
```
> X=arima.sim(n = 240, list(ar = -0.8),sd = 1)
```



Le modèle $AR(1)$, $\rho(1)$ et $\rho(2)$



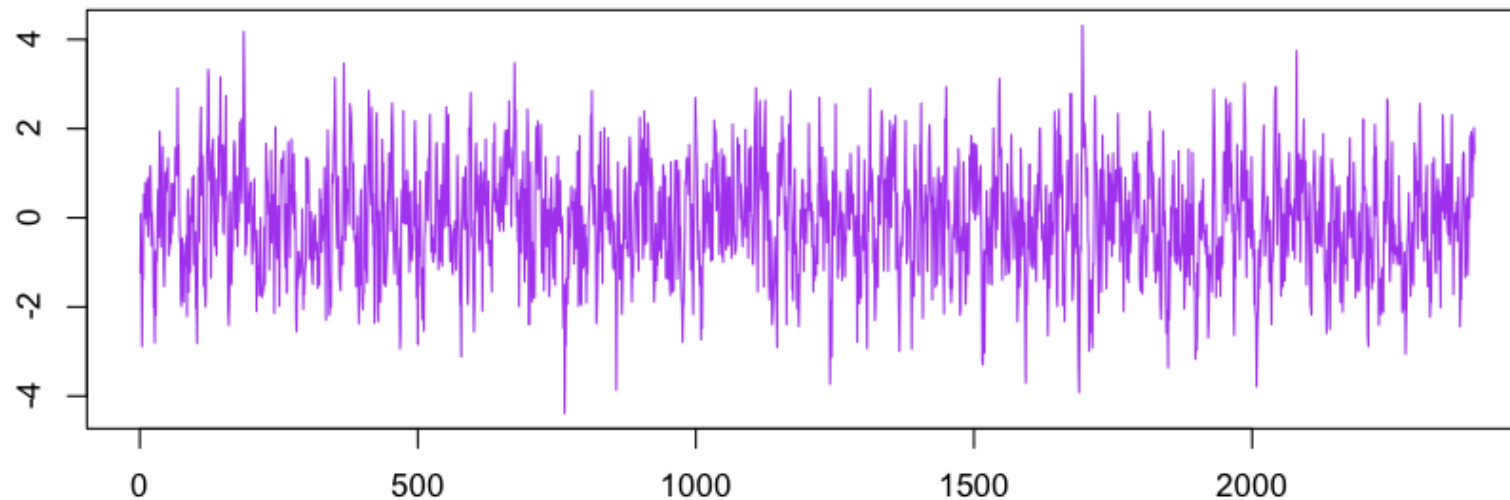
Le modèle $AR(1)$, $\rho(3)$ et $\rho(4)$



Le modèle $AR(1)$ - autorégressif à l'ordre 1

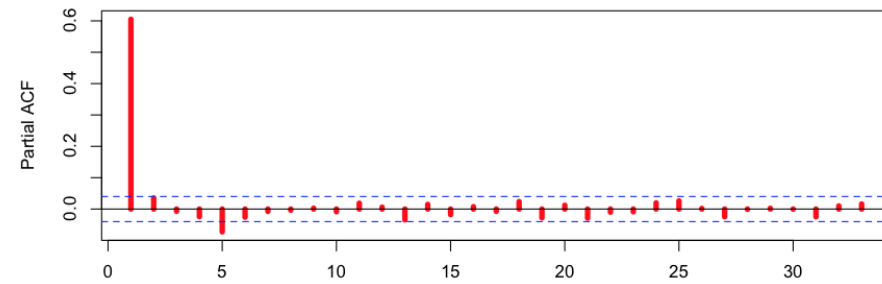
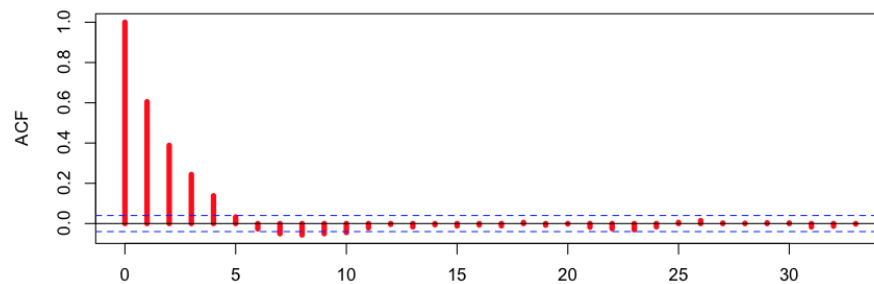
Considérons un processus $AR(1)$ stationnaire avec $\phi_1 = 0.6$.

```
> X=arima.sim(n = 2400, list(ar = 0.6),sd = 1)  
> plot(X)
```



Le modèle $AR(1)$ - autorégressif à l'ordre 1

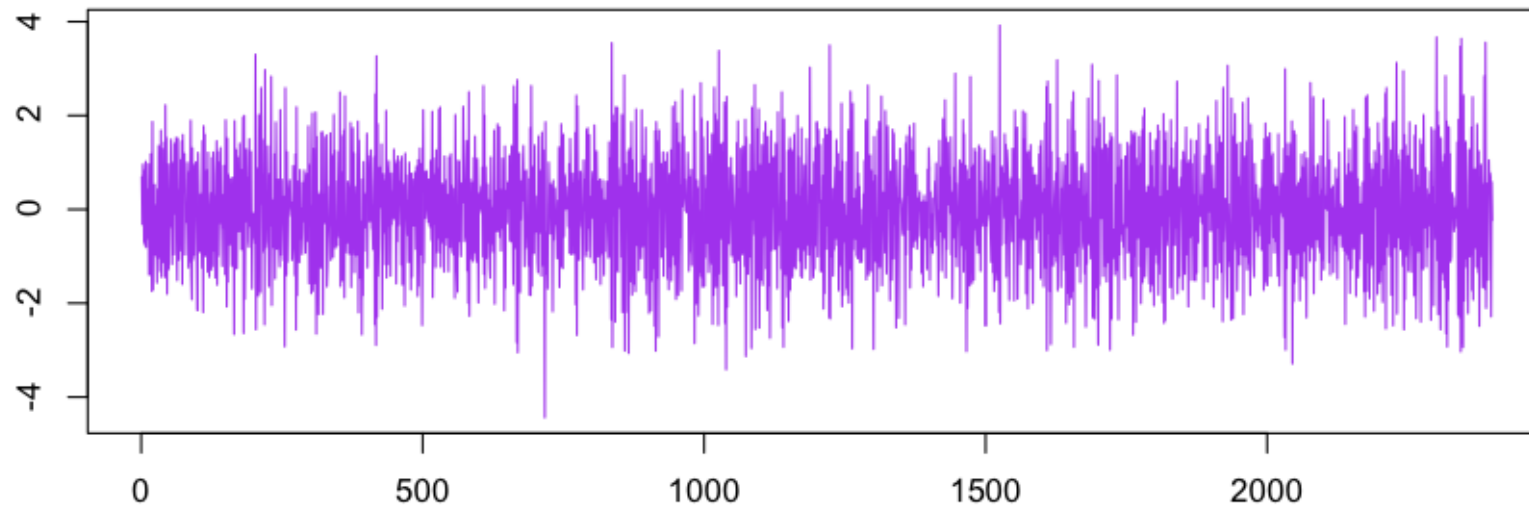
```
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



Le modèle $AR(1)$ - autorégressif à l'ordre 1

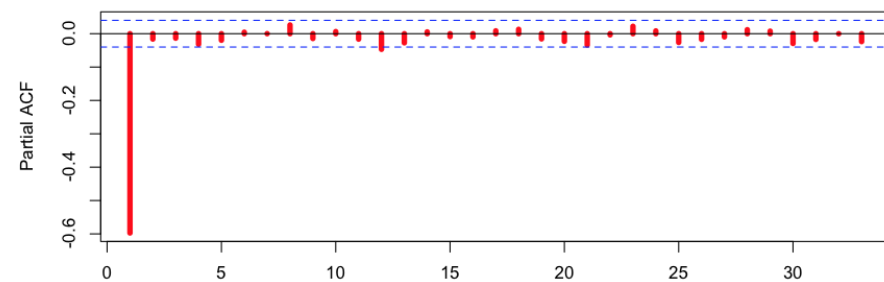
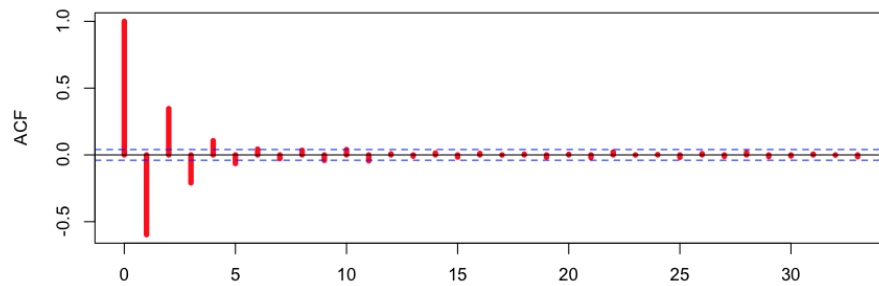
Considérons un processus $AR(1)$ stationnaire avec $\phi_1 = -0.6$.

```
> X=arima.sim(n = 2400, list(ar = -0.6),sd = 1)  
> plot(X)
```



Le modèle $AR(1)$ - autorégressif à l'ordre 1

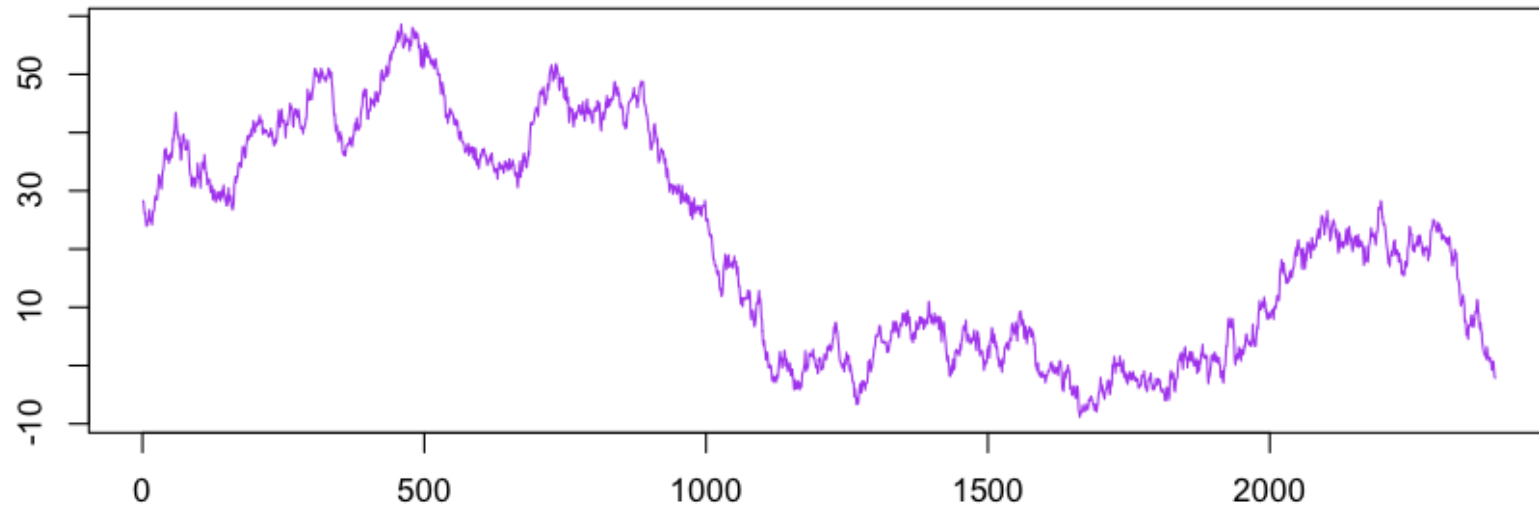
```
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



Le modèle $AR(1)$ - autorégressif à l'ordre 1

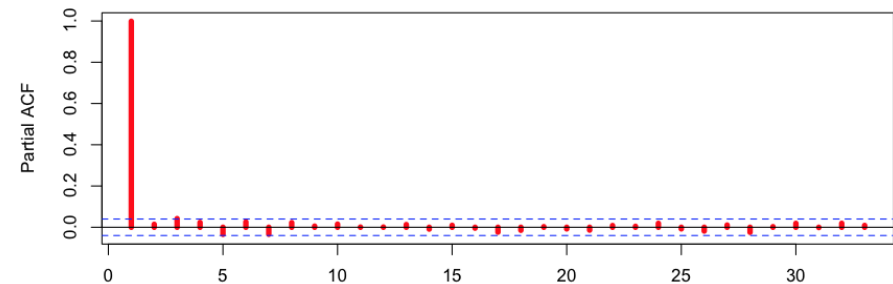
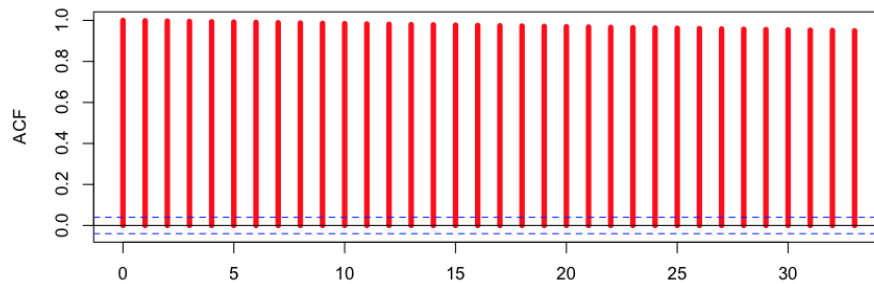
Considérons un processus $AR(1)$ presque plus stationnaire avec $\phi_1 = 0.999$.

```
> X=arima.sim(n = 2400, list(ar = 0.999),sd = 1)  
> plot(X)
```



Le modèle $AR(1)$ - autorégressif à l'ordre 1

```
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



Le modèle $AR(2)$ - autorégressif à l'ordre 2

Ces processus sont également appelés **modèles de Yule**, dont la forme générale est

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = (1 - \phi_1 L - \phi_2 L^2) X_t = \varepsilon_t \text{ pour tout } t \in \mathbb{Z},$$

où (ε_t) est un bruit blanc de variance σ^2 , et où les racines du polynôme caractéristique $\Phi(z) = 1 - \phi_1 z - \phi_2 z^2$ sont supposées à l'extérieur du disque unité, de telle sorte que le processus soit stationnaire. Cette condition s'écrit

$$\begin{cases} 1 - \phi_1 + \phi_2 > 0 \\ 1 + \phi_1 - \phi_2 > 0 \\ \phi_1^2 + 4\phi_2 > 0, \end{cases}$$

Le modèle $AR(2)$ - autorégressif à l'ordre 2

La fonction d'autocorrélation satisfait l'équation de récurrence

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2) \text{ pour } h \geq 2,$$

et la fonction d'autocorrélation partielle vérifie

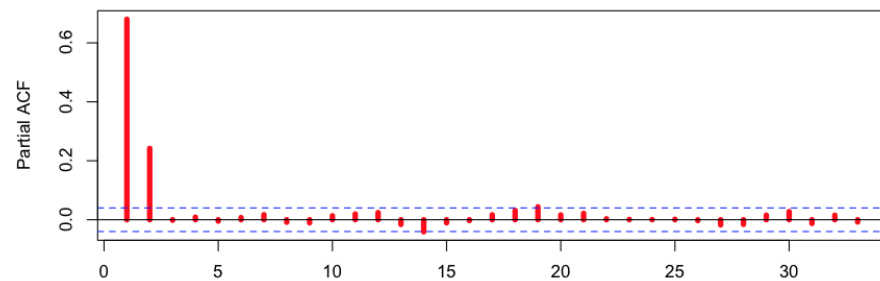
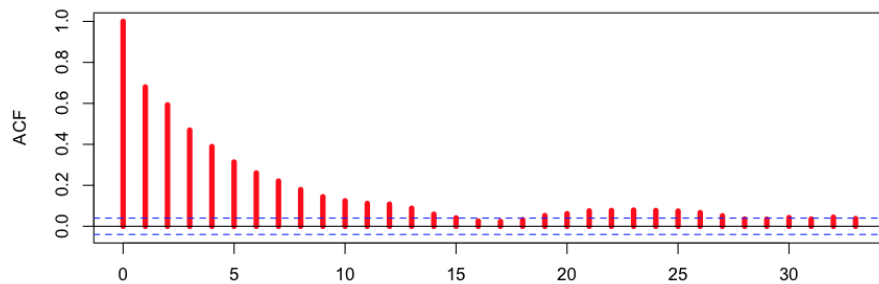
$$\psi(h) = \begin{cases} \rho(1) & \text{pour } h = 1 \\ \left[\rho(2) - \rho(1)^2 \right] / \left[1 - \rho(1)^2 \right] & \text{pour } h = 2 \\ 0 & \text{pour } h \geq 3. \end{cases}$$

À partir des équations de Yule Walker, la fonction d'autocorrélation vérifie la relation de récurrence

$$\begin{cases} \rho(0) = 1 \text{ et } \rho(1) = \phi_1 / (1 - \phi_2), \\ \rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2) \text{ pour } h \geq 2, \end{cases}$$

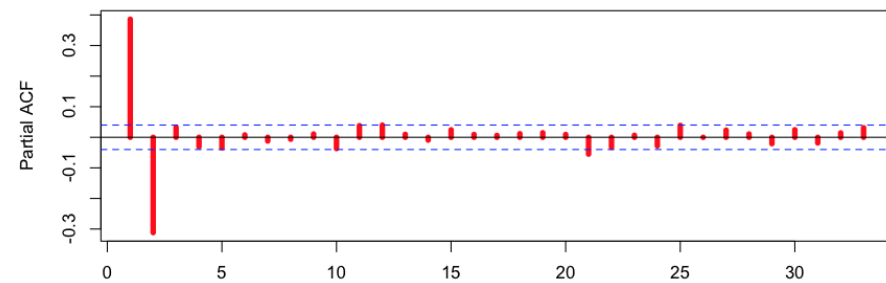
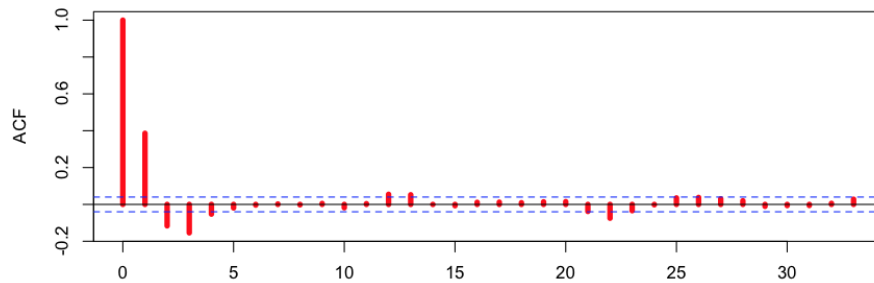
Le modèle $AR(2)$ - autorégressif à l'ordre 2

```
> X=arima.sim(n = 2400, list(ar = c(0.6,0.4)),sd = 1)  
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



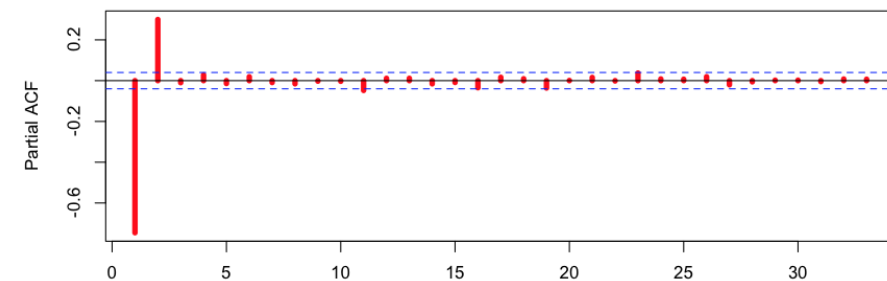
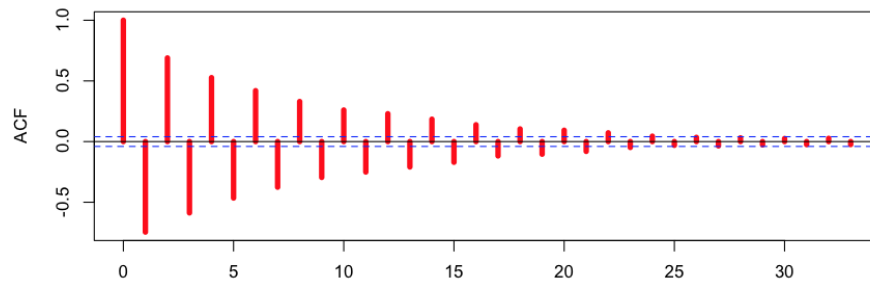
Le modèle $AR(2)$ - autorégressif à l'ordre 2

```
> X=arima.sim(n = 2400, list(ar = c(0.6,-0.4)),sd = 1)  
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



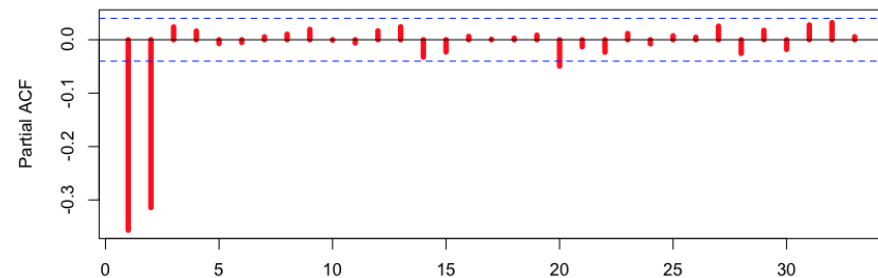
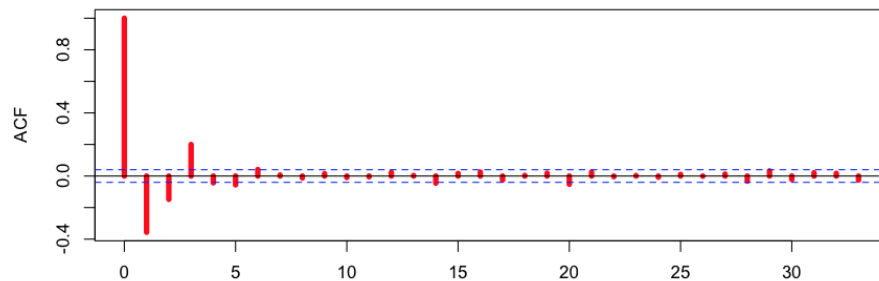
Le modèle $AR(2)$ - autorégressif à l'ordre 2

```
> X=arima.sim(n = 2400, list(ar = c(-0.6,0.4)),sd = 1)  
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



Le modèle $AR(2)$ - autorégressif à l'ordre 2

```
> X=arima.sim(n = 2400, list(ar = c(-0.6,-0.4)),sd = 1)  
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



Le modèle $MA(q)$ - moyenne mobile d'ordre q

On appelle processus **moyenne mobile** (moving average') d'ordre q , noté $MA(q)$, un processus stationnaire (X_t) vérifiant une relation du type

$$X_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \text{ pour tout } t \in \mathbb{Z}, \quad (4)$$

où les θ_i sont des réels et (ε_t) est un bruit blanc de variance σ^2 . (4) est équivalent à l'écriture

$$X_t = \Theta(L) \varepsilon_t \text{ où } \Theta(L) = \mathbb{I} + \theta_1 L + \dots + \theta_q L^q.$$

Remarque : Contrairement aux processus $AR(p)$, les processus $MA(q)$ sont toujours des processus stationnaires (si $q < \infty$ ou si la série des θ_k est absolument convergente si $q = \infty$).

Le modèle $MA(q)$ - moyenne mobile d'ordre q

La fonction d'autocovariance est donnée par

$$\begin{aligned} \gamma(h) &= \mathbb{E}(X_t X_{t-h}) \\ &= \mathbb{E}([\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}] [\varepsilon_{t-h} + \theta_1 \varepsilon_{t-h-1} + \dots + \theta_q \varepsilon_{t-h-q}]) \\ &= \begin{cases} [\theta_h + \theta_{h+1} \theta_1 + \dots + \theta_q \theta_{q-h}] \sigma^2 & \text{si } 1 \leq h \leq q \\ 0 & \text{si } h > q, \end{cases} \end{aligned}$$

avec, pour $h = 0$, la relation

$$\gamma(0) = [1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2] \sigma^2.$$

Cette dernière relation peut se réécrire

$$\gamma(k) = \sigma^2 \sum_{j=0}^q \theta_j \theta_{j+k} \text{ avec la convention } \theta_0 = 1.$$

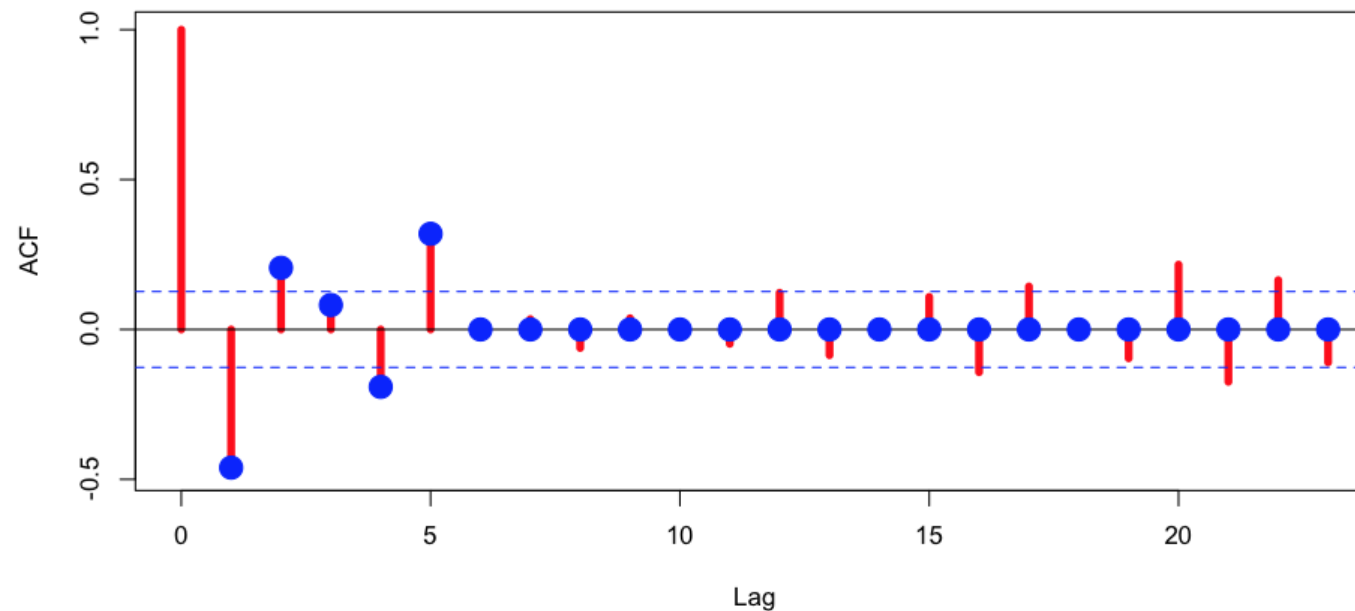
Le modèle $MA(q)$ - moyenne mobile d'ordre q

D'où la fonction d'autocovariance,

$$\rho(h) = \frac{\theta_h + \theta_{h+1}\theta_1 + \dots + \theta_q\theta_{q-h}}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \text{ si } 1 \leq h \leq q,$$

et $\rho(h) = 0$ pour $h > q$.

Proposition 5. *Si (X_t) est un processus $MA(q)$, $\gamma(q) = \sigma^2\theta_q \neq 0$, alors que $\gamma(k) = 0$ pour $k > q$.*



Le modèle $MA(1)$ - moyenne mobile d'ordre 1

La forme générale des processus de type $MA(1)$ est

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}, \text{ pour tout } t \in \mathbb{Z},$$

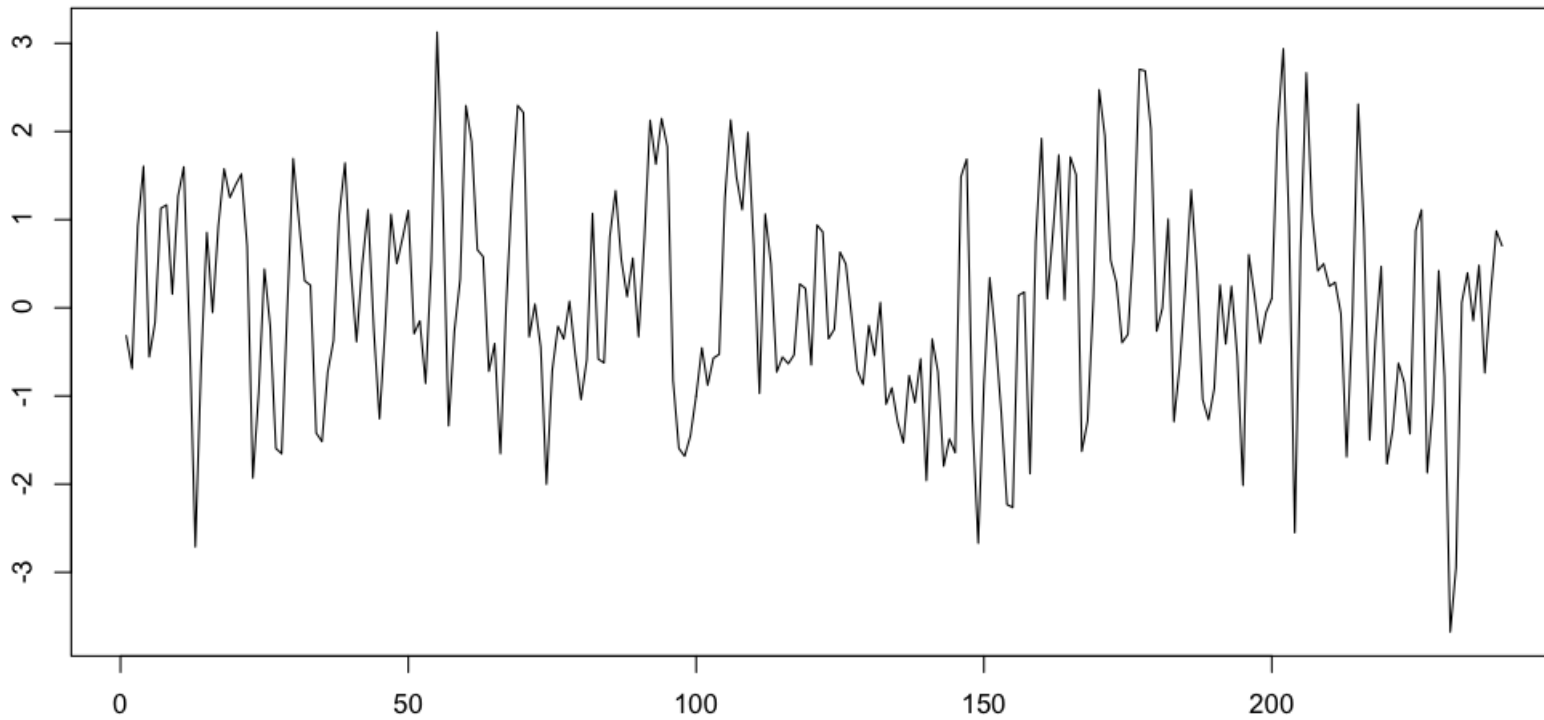
où (ε_t) est un bruit blanc de variance σ^2 . Les autocorrélations sont données par

$$\rho(1) = \frac{\theta}{1 + \theta^2}, \text{ et } \rho(h) = 0, \text{ pour } h \geq 2.$$

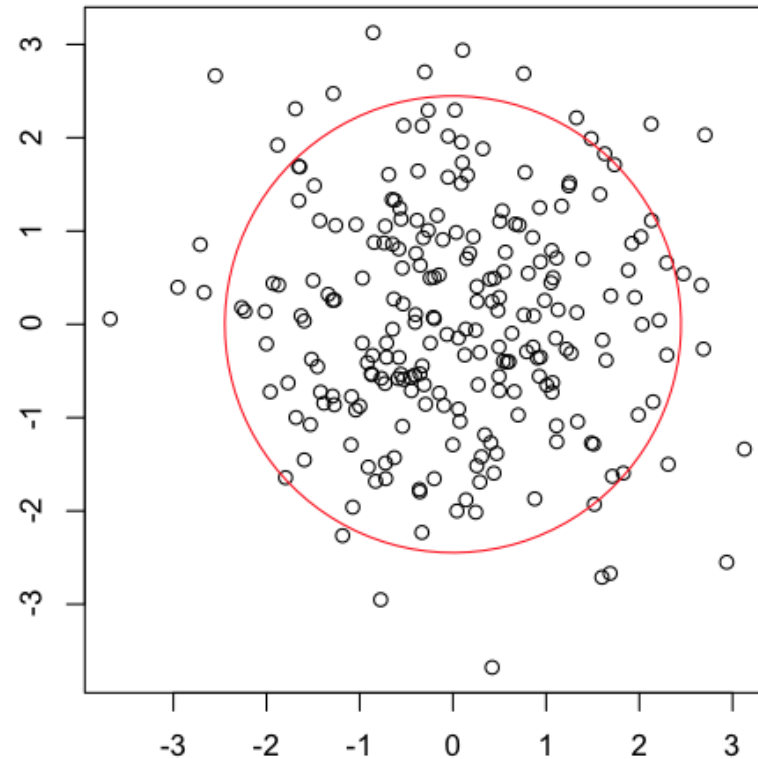
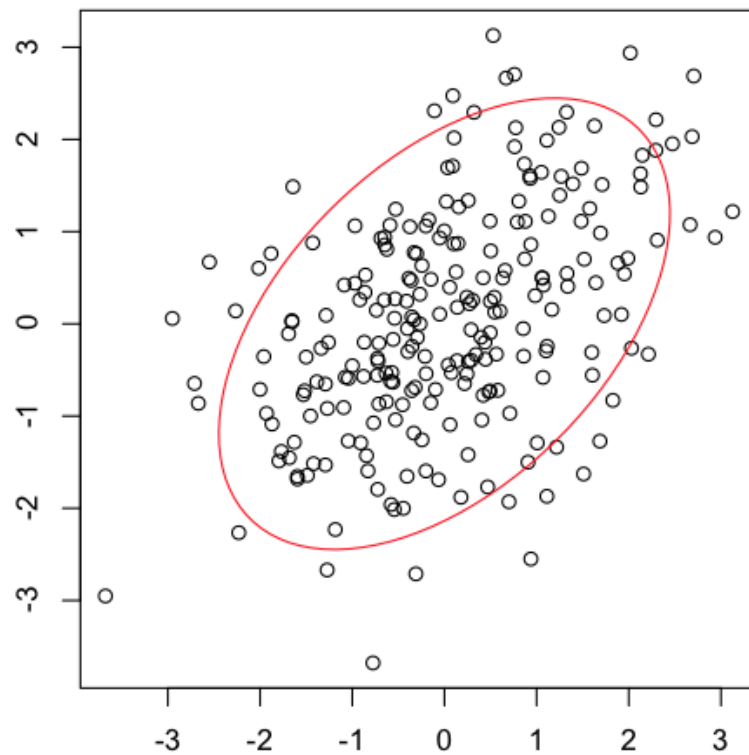
On peut noter que $-1/2 \leq \rho(1) \leq 1/2$: les modèles $MA(1)$ ne peuvent avoir de fortes autocorrélations à l'ordre 1.

```
> X=arima.sim(n = 240, list(ma = 0.8),sd = 1)
> plot(X)
> n=240;h=1
> plot(X[1:(n-h)],X[(1+h):n])
> library(ellipse)
> lines(ellipse(.8/(1+.8^2)), type = 'l',col="red")
```

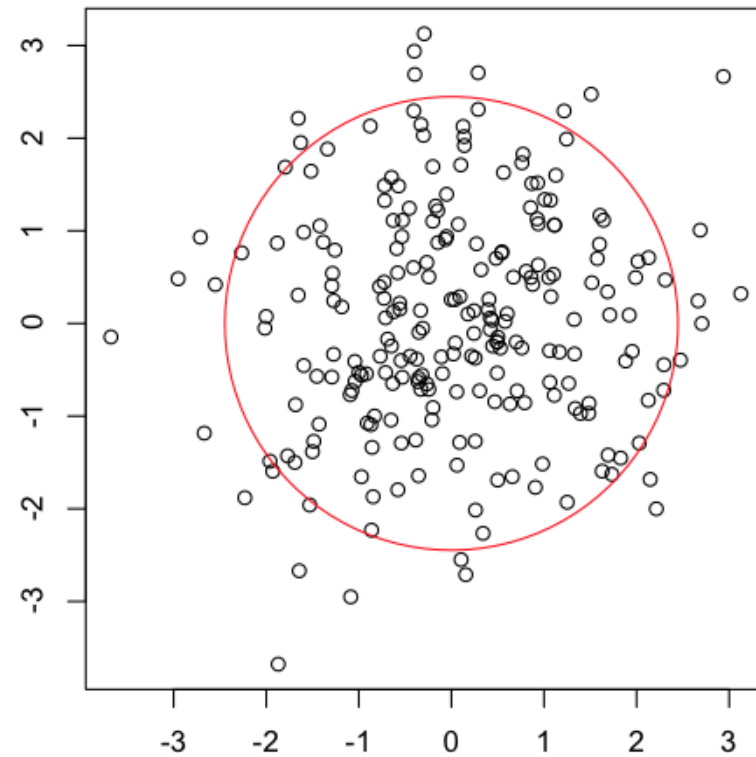
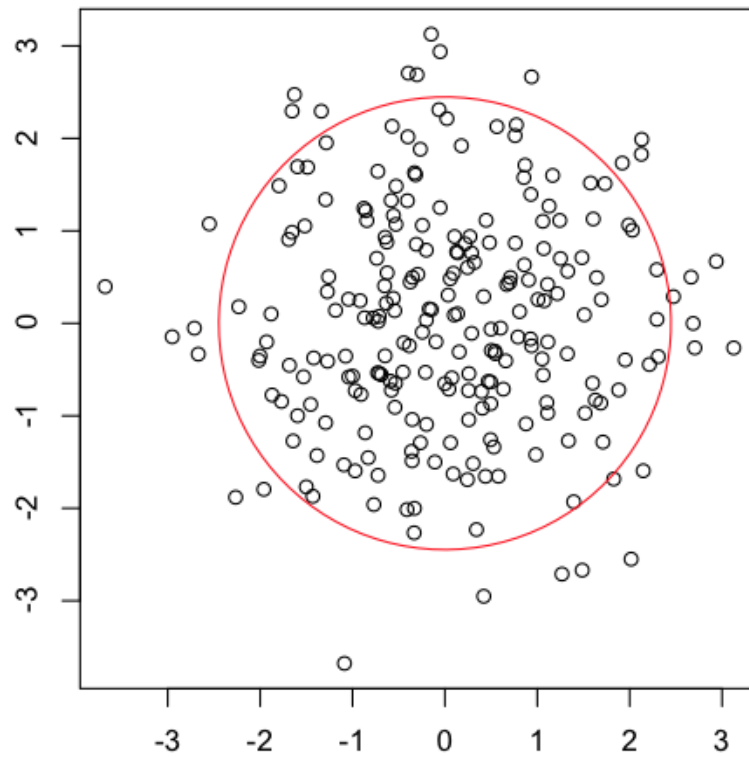
Le modèle $MA(1)$ - moyenne mobile d'ordre 1



Le modèle $MA(1)$ - moyenne mobile d'ordre 1

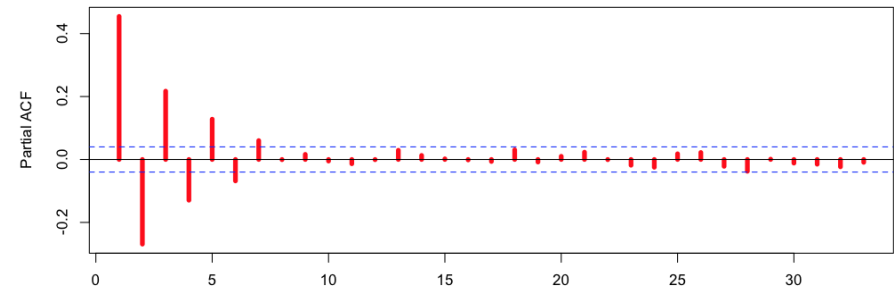
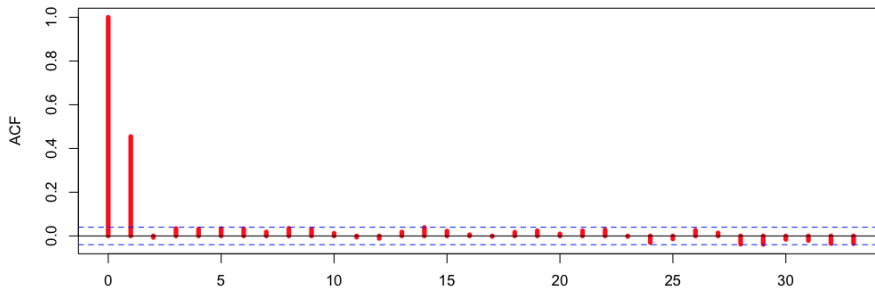


Le modèle $MA(1)$ - moyenne mobile d'ordre 1



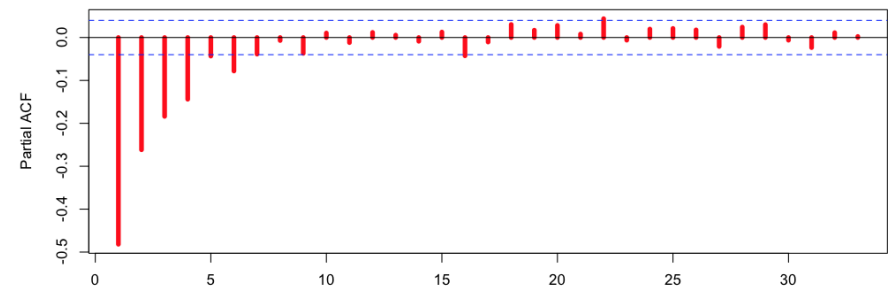
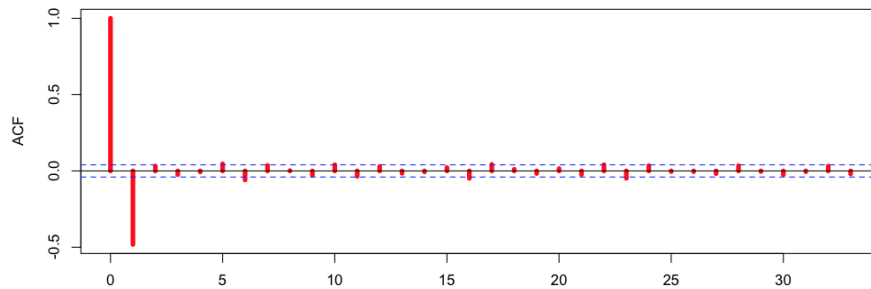
Le modèle $MA(1)$ - moyenne mobile d'ordre 1

```
> X=arima.sim(n = 2400, list(ma = .7),sd = 1)  
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



Le modèle $MA(1)$ - moyenne mobile d'ordre 1

```
> X=arima.sim(n = 2400, list(ma = -0.7),sd = 1)  
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



Le modèle $MA(2)$ - moyenne mobile d'ordre 2

La forme générale de (X_t) suivant un processus $MA(2)$ est

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \text{ pour tout } t \in \mathbb{Z},$$

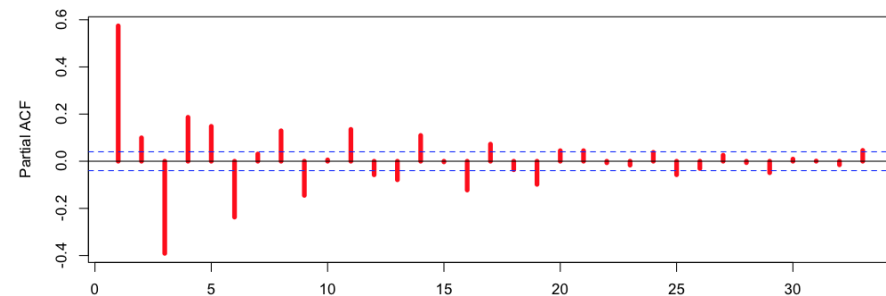
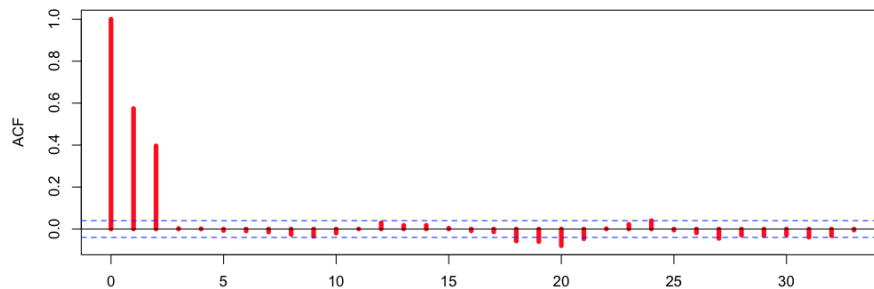
où (ε_t) est un bruit blanc de variance σ^2 .

La fonction d'autocorrélation est donnée par l'expression suivante

$$\rho(h) = \begin{cases} \theta_1 [1 + \theta_2] / [1 + \theta_1^2 + \theta_2^2] & \text{pour } h = 1 \\ \theta_2 / [1 + \theta_1^2 + \theta_2^2] & \text{pour } h = 2 \\ 0 & \text{pour } h \geq 3, \end{cases}$$

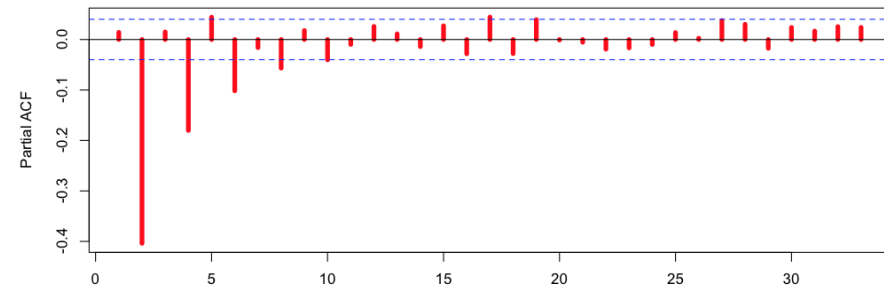
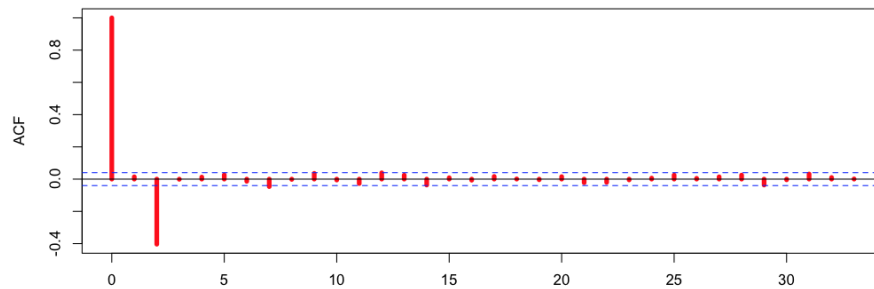
Le modèle $MA(2)$ - moyenne mobile d'ordre 2

```
> X=arima.sim(n = 2400, list(ma = c(0.7,0.9)),sd = 1)  
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



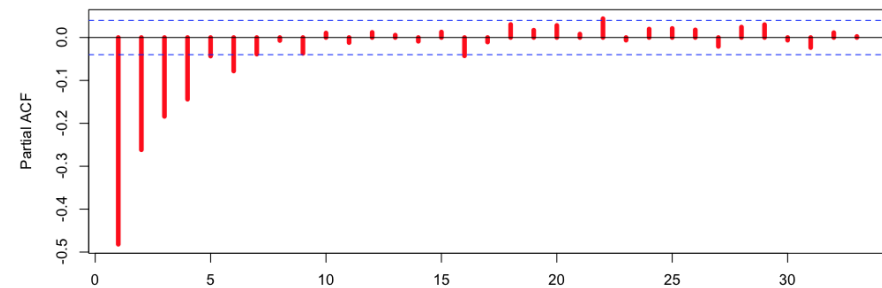
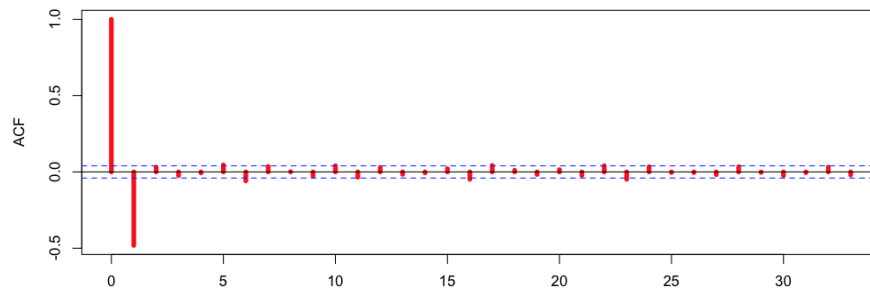
Le modèle $MA(2)$ - moyenne mobile d'ordre 2

```
> X=arima.sim(n = 2400, list(ma = c(0.7,-0.9)),sd = 1)  
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



Le modèle $MA(2)$ - moyenne mobile d'ordre 2

```
> X=arima.sim(n = 2400, list(ma = c(0.7,-0.9)),sd = 1)  
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



Le modèle $ARMA(p, q)$

On appelle processus $ARMA(p, q)$, un processus stationnaire (X_t) vérifiant une relation du type

$$X_t - \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \text{ pour tout } t \in \mathbb{Z}, \quad (5)$$

où les θ_i sont des réels et (ε_t) est un bruit blanc de variance σ^2 . (5) est équivalent à l'écriture

$$\Phi(L) X_t = \Theta(L) \varepsilon_t \text{ où } \begin{cases} \Theta(L) = \mathbb{I} + \theta_1 L + \dots + \theta_q L^q \\ \Phi(L) = \mathbb{I} - \phi_1 L - \dots - \phi_p L^p \end{cases}$$

On supposera de plus que les polyômes Φ et Θ n'ont pas de racines en module strictement supérieures à 1, et n'ont pas de racine commune. On supposera de plus que $\theta_q \neq 0$ et $\phi_p \neq 0$. On dira dans ce cas que cette écriture est la **forme minimale**.

Le modèle $ARMA(p, q)$

Proposition 6. *Soit (X_t) un processus $ARMA(p, q)$, alors les autocovariances $\gamma(h)$ satisfont*

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = 0 \text{ pour } h \geq q+1. \quad (6)$$

Proposition 7. *Soit (X_t) un processus $ARMA(p, q)$, alors les autocorrélations $\gamma(h)$ satisfont*

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = \sigma^2 [\theta_h + h_1 \theta_{h+1} + \dots + h_{q-h} \theta_q] \text{ pour } 0 \leq h \leq q, \quad (7)$$

où les h_i correspondent aux coefficients de la forme $MA(\infty)$ de (X_t) ,

$$X_t = \sum_{j=0}^{+\infty} h_j \varepsilon_{t-j}.$$

Le modèle $ARMA(p, q)$

Remarque : La variance de X_t est donnée par

$$\text{Var}(X_t) = \gamma(0) = \frac{1 + \theta_1^2 + \dots + \theta_q^2 + 2\phi_1\theta_1 + \dots + \phi_h\theta_h}{1 - \phi_1^2 - \dots - \phi_p^2} \sigma^2 \text{ où } h = \min(p, q).$$

Le modèle $ARMA(1, 1)$

Soit (X_t) un processus $ARMA(1, 1)$ défini par

$$X_t - \phi X_{t-1} = \varepsilon_t + \theta \varepsilon_{t-1}, \text{ pour tout } t,$$

où $\phi \neq 0$, $\theta \neq 0$, $|\phi| < 1$ et $|\theta| < 1$. Ce processus peut de mettre sous forme $AR(\infty)$, puisque

$$(1 - \phi L)(1 + \theta L)^{-1} X_t = \Pi(L) X_t = \varepsilon_t,$$

où

$$\Pi(L) = (1 - \phi L) \left[1 - \theta L + \theta^2 L^2 + \dots + (-1)^h \theta^h L^h + \dots \right],$$

aussi

$$\Pi(L) = \sum_{i=0}^{+\infty} \pi_i L^i \text{ où } \begin{cases} \pi_0 = 1 \\ \pi_i = (-1)^i [\phi + \theta] \theta^{i-1} \text{ pour } i \geq 1. \end{cases}$$

Le modèle $ARMA(1, 1)$

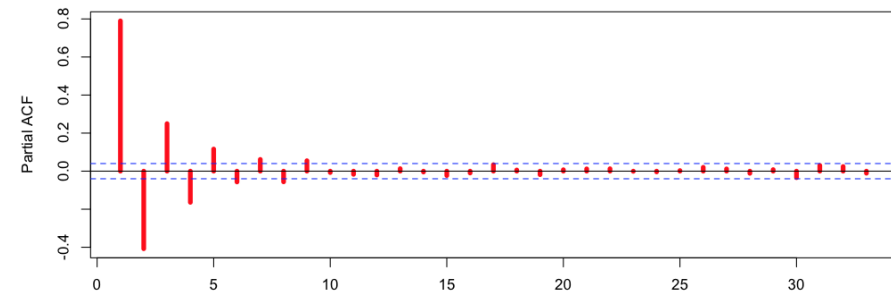
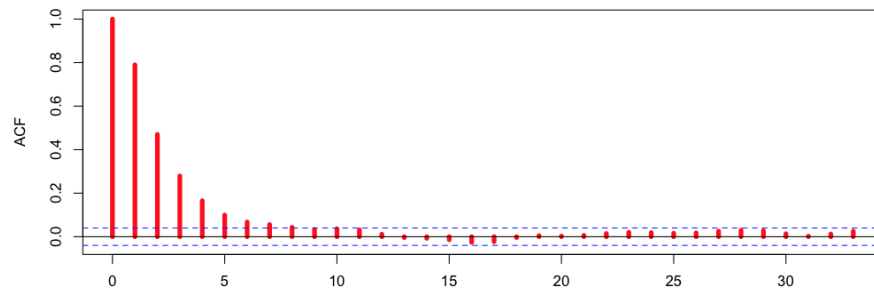
La fonction d'autocorrélation s'écrit

$$\begin{cases} \rho(1) = (1 + \phi\theta) (\phi + \theta) / [1 + \theta^2 + 2\phi\theta] \\ \rho(h) = \phi^h \rho(1) \text{ pour } h \geq 2, \end{cases}$$

et la fonction d'autocorrélations partielles a le même comportement qu'une moyenne mobile, avec comme valeur initiale $a\psi(1) = \rho(1)$.

Le modèle $ARMA(1, 1)$

```
> X=arima.sim(n = 2400, list(ar=0.6, ma = 0.7),sd = 1)  
> plot(acf(X),lwd=5,col="red")  
> plot(pacf(X),lwd=5,col="red")
```



Le modèle $ARIMA(p, d, q)$

Définition l'opérateur Δ par $\Delta X_t = X_t - X_{t-1}$, i.e.

$$\begin{cases} \Delta X_t = X_t - X_{t-1} = (1 - L) X_t \\ \Delta^d X_t = (1 - L)^d X_t \end{cases}$$

Un processus (X_t) est un processus $ARIMA(p, d, q)$ - autorégressif moyenne mobile **intégré** - s'il vérifie une équation du type

$$\Phi(L)(1 - L)^d X_t = \Theta(L)\varepsilon_t \text{ pour tout } t \geq 0$$

où

$$\begin{cases} \Phi(L) = \mathbb{I} - \phi_1 L - \phi_2 L^2 + \dots - \phi_p L^p \text{ où } \phi_p \neq 0 \\ \Theta(L) = \mathbb{I} + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q \text{ où } \theta_q \neq 0 \end{cases}$$

sont des polynômes dont les racines sont de module supérieur à 1, ...

Le modèle $ARIMA(p, d, q)$

... et où les conditions initiales

$$Z_{-1} = \{X_{-1}, \dots, X_{-p}, \varepsilon_{-1}, \dots, \varepsilon_{-q}\}$$

sont non-corrélées avec $\varepsilon_0, \dots, \varepsilon_t, \dots$ et où le processus (ε_t) est un bruit blanc de variance σ^2 .

Remarque : Si les processus $ARMA$ peuvent être définis sur \mathbb{Z} , il n'en est pas de même pour les processus $ARIMA$ qui doivent commencer à une certaine date ($t = 0$ par convention), avec des valeurs initiales (q valeurs pour les ε_t , et $p + d$ pour X_t).

Proposition 8. *Soit (X_t) un processus $ARIMA(p, d, q)$ alors le processus $(\Delta^d X_t)$ converge vers un processus $ARMA(p, q)$ stationnaire.*

Le modèle $ARIMA(p, d, q)$

Proposition 9. Soit (X_t) un processus $ARIMA(p, d, q)$ de valeurs initiales Z_{-1} , alors (X_t) peut s'écrire sous la forme suivante, fonction du passé du bruit,

$$X_t = \sum_{j=1}^t h_j \varepsilon_{t-j} + h^*(t) Z_{-1},$$

où les h_j sont les coefficients de la division selon les puissances croissantes de Θ par Φ , et $h^*(t)$ est un vecteur (ligne) de fonctions de t

Proposition 10. Soit (X_t) un processus $ARIMA(p, d, q)$ de valeurs initiales Z_{-1} , alors (X_t) peut s'écrire sous la forme suivante, fonction du passé de X_t

$$X_t = \sum_{j=1}^t \pi_j X_{t-j} + \bar{h}^*(t) Z_{-1} + \varepsilon_t,$$

où les π_j sont les coefficients (pour $j \geq 1$) de la division selon les puissances croissantes de Φ par Θ , et $\bar{h}^*(t)$ est un vecteur (ligne) de fonctions de t quand $t \rightarrow \infty$.

Le(s) modèle(s) *SARIMA*(s, p, d, q)

De façon ‘générale’, soient s_1, \dots, s_n n entiers, alors un processus (X_t) est un processus *SARIMA*(p, d, q) - autorégressif moyenne mobile intégré saisonnier - s’il vérifie une équation du type

$$\Phi(L)(1 - L^{s_1}) \dots (1 - L^{s_n}) X_t = \Theta(L)(1 - L)^d \varepsilon_t \text{ pour tout } t \geq 0$$

où $\Phi(L) = \mathbb{I} - \phi_1 L - \phi_2 L^2 + \dots - \phi_p L^p$ (où $\phi_p \neq 0$) et

$\Theta(L) = \mathbb{I} + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$ (où $\theta_q \neq 0$) sont des polynômes dont les racines sont de module supérieur à 1, et où les conditions initiales

$$Z_{-1} = \{X_{-1}, \dots, X_{-p}, \varepsilon_{-1}, \dots, \varepsilon_{-q}\}$$

sont non-corrélées avec $\varepsilon_0, \dots, \varepsilon_t, \dots$ et où le processus (ε_t) est un bruit blanc de variance σ^2 .

Le(s) modèle(s) *SARIMA*(s, p, d, q)

Il est possible d'avoir de la saisonnalité, sans forcément avoir de racine unité saisonnière.

Sous \mathbb{R} , un modèle $\text{arima}(p, d, q) (ps, ds, qs) [s]$ (sans constante) s'écrit

$$(1 - L)^d (1 - L^s)^{ds} \Phi(L) \Phi_s(L^s) X_t = \Theta(L) \Theta_s(L^s)$$

avec $\deg(\Phi) = p$, $\deg(\Theta) = q$, $\deg(\Phi_s) = ps$, et $\deg(\Theta_s) = qs$.

Par exemple $\text{arima}(1, 0, 2) (1, 0, 0) [4]$ est un modèle

$$(1 - \phi L)(1 - \phi_4 L^4) X_t = (1 + \theta_1 L + \theta_2 L^2) \varepsilon_t$$

Le(s) modèle(s) *SARIMA*(s, p, d, q)

Les deux formes les plus utilisées sont les suivantes,

$$\Phi(L)(1 - L^s)X_t = \Theta(L)\varepsilon_t \text{ pour tout } t \geq 0$$

$$\Phi(L)(1 - L^s)(1 - L)^d X_t = \Theta(L)\varepsilon_t \text{ pour tout } t \geq 0$$

Le(s) modèle(s) *SARIMA*(s, p, d, q)

Soit $S \in \mathbb{N} \setminus \{0\}$ correspondant à la saisonnalité, et considérons le processus

$$X_t = (1 - \alpha L) (1 - \beta L^S) \varepsilon_t = \varepsilon_t - \alpha \varepsilon_{t-1} - \beta \varepsilon_{t-S} + \alpha \beta \varepsilon_{t-S-1}.$$

Les autocorrélations sont données par

$$\rho(1) = \frac{-\alpha(1 + \beta^2)}{(1 + \alpha^2)(1 + \beta^2)} = \frac{-\alpha}{1 + \alpha^2},$$

$$\rho(S-1) = \frac{\alpha\beta}{(1 + \alpha^2)(1 + \beta^2)},$$

$$\rho(S) = \frac{-\beta(1 + \alpha^2)}{(1 + \alpha^2)(1 + \beta^2)} = \frac{-\beta}{1 + \beta^2},$$

$$\rho(S+1) = \frac{\alpha\beta}{(1 + \alpha^2)(1 + \beta^2)},$$

et $\rho(h) = 0$ ailleurs. On peut noter que $\rho(S-1) = \rho(S+1) = \rho(1) \times \rho(S)$.

Le(s) modèle(s) *SARIMA*(s, p, d, q)

Soit $S \in \mathbb{N} \setminus \{0\}$ correspondant à la saisonnalité, et considérons le processus

$$(1 - \phi L^S) X_t = (1 - \alpha L) (1 - \beta L^S) \varepsilon_t \text{ ou } X_t - \phi X_{t-S} = \varepsilon_t - \alpha \varepsilon_{t-1} - \beta \varepsilon_{t-S} + \alpha \beta \varepsilon_{t-S-1}.$$

Les autocorrélations sont données par

$$\rho(1) = \frac{-\alpha(1 + \beta^2)}{(1 + \alpha^2)(1 + \beta^2)} = \frac{-\alpha}{1 + \alpha^2},$$

$$\rho(S-1) = \frac{\alpha \left[\beta - \phi - \phi(\beta - \phi)^2 / (1 - \phi^2) \right]}{(1 + \alpha^2) \left[1 + (\beta - \phi)^2 / (1 - \phi^2) \right]},$$

$$\rho(S) = \frac{-(1 + \alpha^2)}{\alpha} \rho_{S-1},$$

avec $\rho(h) = 0$ pour $2 \leq h \leq S-2$, puis $\rho(S+1) = \rho(S-1)$ et $\rho(h) = \phi \rho(h-S)$ pour $h \geq S+2$. En particulier $\rho(kS) = \phi^{k-1} \rho(S)$.

Le théorème de Wold

Theorem 11. *Tout processus (X_t) , centré, et stationnaire au second ordre, peut être représenté sous une forme proche de la forme MA*

$$X_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + \eta_t,$$

où

- (ε_t) est l'innovation, au sens où $\varepsilon_t = X_t - EL(X_t | X_{t-1}, X_{t-2}, \dots)$,
- $EL(\varepsilon_t | X_{t-1}, X_{t-2}, \dots) = 0$, $\mathbb{E}(\varepsilon_t X_{t-j}) = 0$, $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}(\varepsilon_t^2) = \sigma^2$ (indépendant de t) et $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$ pour $t \neq s$,
- toutes les racines de $\Theta(L)$ sont à l'extérieur du cercle unité : le polynôme Θ est inversible,
- $\sum_{j=0}^{\infty} \theta_j^2 < \infty$ et $\theta_0 = 1$,
- les coefficients θ_j et le processus (ε_t) sont uniques,
- (η_t) vérifie $\eta_t = EL(\eta_t | X_{t-1}, X_{t-2}, \dots)$.

Estimation d'un SARIMA : Box & Jenkins

La méthodologie pour estimer un processus SARIMA est la suivante

- identification de l'ordre d : poser $Y_t = (1 - L)^d X_t$
- identification de l'ordre S : poser $Z_t = (1 - L^S)Y_t$
- identification de l'ordre p, q tels que $\Phi_p(L)Z_t = \Theta_q(L)\varepsilon_t$
- estimer ϕ_1, \dots, ϕ_p et $\theta_1, \dots, \theta_q$
- construire la série $(\hat{\varepsilon}_t)$, en déduire un estimateur de σ^2
- vérifier que $(\hat{\varepsilon}_t)$ est un bruit blanc

Identification de l'ordre d d'intégration

Le **test de Dickey & Fuller simple**, H_0 : le processus suit une marche aléatoire contre l'hypothèse alternative H_1 : le processus suit un modèle $AR(1)$ (stationnaire).

- $Y_t = \rho Y_{t-1} + \varepsilon_t$: on teste $H_0 : \rho = 1$ (*marche aléatoire sans dérive*)
- $Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$: on teste $H_0 : \alpha = 0$ et $\rho = 1$ (*marche aléatoire sans dérive*)
- $Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$: on teste $H_0 : \alpha \neq 0$ et $\rho = 1$ (*marche aléatoire avec dérive*)
- $Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t$: on teste $H_0 : \alpha = 0, \beta = 0$ et $\rho = 1$ (*marche aléatoire sans dérive*)

Identification de l'ordre d d'intégration

Le **test de Dickey & Fuller augmenté**, H_0 : le processus suit une marche aléatoire contre l'hypothèse alternative H_1 : le processus suit un modèle $AR(p)$ (stationnaire).

- $\Phi(L) Y_t = \varepsilon_t$: on teste $H_0 : \Phi(1) = 0$
- $\Phi(L) Y_t = \alpha + \varepsilon_t$: on teste $H_0 : \alpha = 0$ et $\Phi(1) = 0$
- $\Phi(L) Y_t = \alpha + \varepsilon_t$: on teste $H_0 : \alpha \neq 0$ et $\Phi(1) = 0$
- $\Phi(L) Y_t = \alpha + \beta t + \varepsilon_t$: on teste $H_0 : \alpha = 0, \beta = 0$ et $\Phi(1) = 0$

Ces 4 cas peuvent être réécrits en introduisant les notations suivantes,

$$\Phi(L) = \Phi(1) + (1-L)\Phi^*(L) = \Phi(1) - \left[\sum_{i=0}^{p-1} \alpha_i L^i \right] (1-L)$$

où $\alpha_0 = \Phi(1) - 1$ et $\alpha_i = \alpha_{i-1} - \phi_i = \phi_{i+1} + \dots + \phi_p$, pour $i = 1, \dots, p$.

Identification de l'ordre d d'intégration

En posant $\rho = 1 - \Phi(1)$, on peut réécrire les 4 cas en

(1) $Y_t = \rho Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$: on teste $H_0 : \rho = 1$

(2) $Y_t = \alpha + \rho Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$: on teste $H_0 : \alpha = 0$ et $\rho = 1$

(3) $Y_t = \alpha + \rho Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$: on teste $H_0 : \alpha \neq 0$ et $\rho = 1$

(4) $Y_t = \alpha + \beta t + \rho Y_{t-1} + \sum \alpha_i \Delta y_{t-i} + \varepsilon_t$: on teste $H_0 : \alpha = 0, \beta = 0$ et $\rho = 1$

> `library(urca)`

> `summary(ur.df(y=,lag=1,type="trend"))`

Il est aussi possible de laisser le logiciel choisir le nombre optimal de retard à considérer (à l'aide du BIC, e.g.)

> `library(urca)`

> `summary(ur.df(y=,lag=6,selectlags="BIC",type="trend"))`

Les tests de racine unité

Dans le test de Dickey-Fuller (simple), on considère un modèle autorégressif (à l'ordre 1),

$$X_t = \alpha + \beta t + \phi X_{t-1} + \varepsilon_t$$

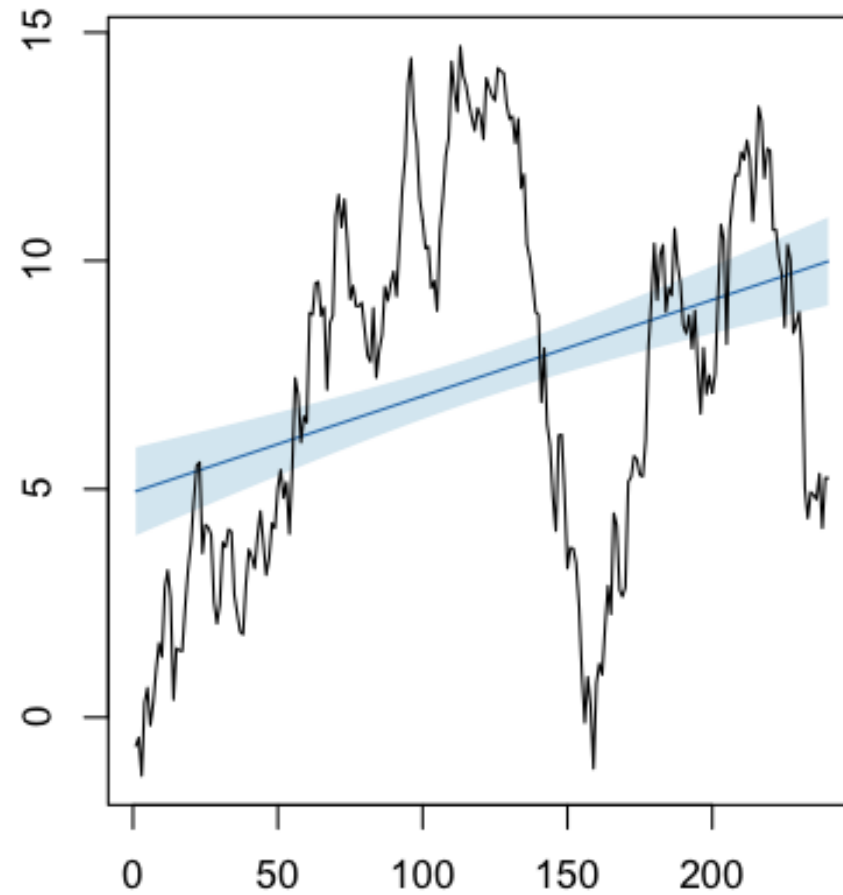
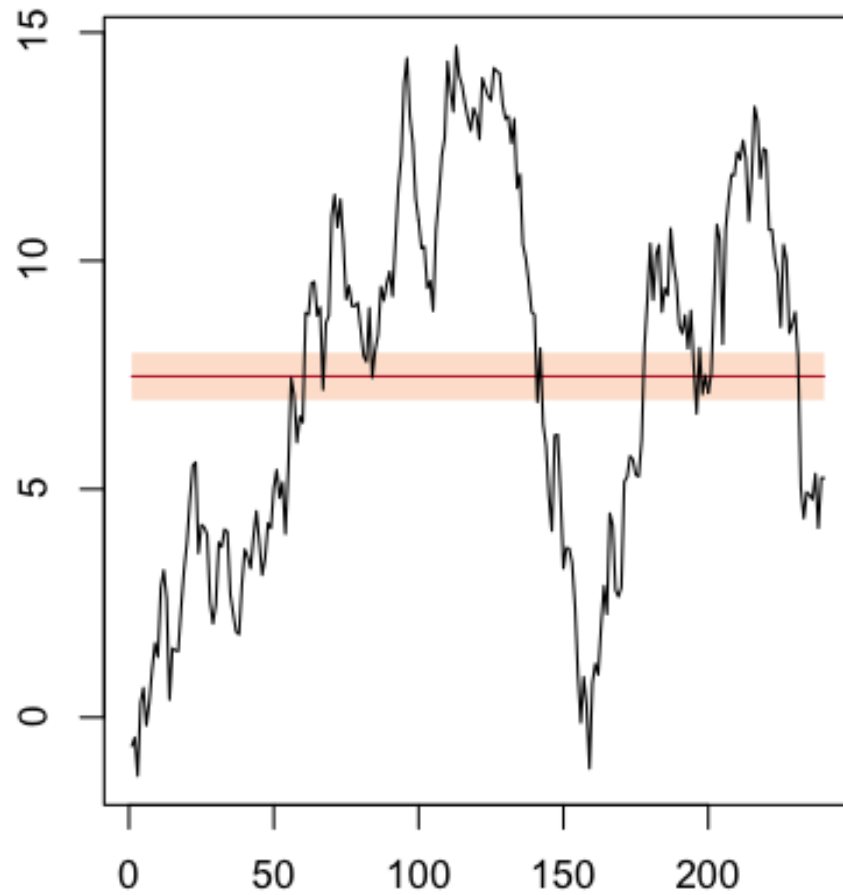
qui peut aussi s'écrire

$$X_t - X_{t-1} = \Delta X_t = \alpha + \beta t + \phi X_{t-1} + \varepsilon_t, \text{ où } \rho = \phi - 1$$

Une racine unité se traduit par $\phi = 1$ ou encore $\rho = 0$, que l'on peut tester par un test de Student.

Considérons la série simulée suivante

```
> set.seed(1)
> E=rnorm(240)
> X=cumsum(E)
> plot(X,type="l")
```



```
> lags=0  
> z=diff(X)  
> n=length(z)  
> z.diff=embed(z, lags+1)[,1]
```

```
> z.lag.1=X[(lags+1):n]
> #z.diff.lag = x[, 2:lags]
> summary(lm(z.diff~0+z.lag.1 ))
```

Call:

```
lm(formula = z.diff ~ 0 + z.lag.1)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
z.lag.1	-0.005609	0.007319	-0.766	0.444

Residual standard error: 0.963 on 238 degrees of freedom

Multiple R-squared: 0.002461, Adjusted R-squared: -0.00173

F-statistic: 0.5873 on 1 and 238 DF, p-value: 0.4442

Les tests de racine unité

```
> summary(lm(z.diff~0+z.lag.1 ))$coefficients[1,3]
```

```
[1] -0.7663308
```

```
> qnorm(c(.01,.05,.1)/2)
```

```
[1] -2.575829 -1.959964 -1.644854
```

```
> library(urca)
```

```
> df=ur.df(X,type="none",lags=0)
```

```
> summary(df)
```

```
Test regression none
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-2.84466	-0.55723	-0.00494	0.63816	2.54352

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
z.lag.1	-0.005609	0.007319	-0.766	0.444

Residual standard error: 0.963 on 238 degrees of freedom

Multiple R-squared: 0.002461, Adjusted R-squared: -0.00173

F-statistic: 0.5873 on 1 and 238 DF, p-value: 0.4442

Value of test-statistic is: -0.7663

Critical values for test statistics:

1pct 5pct 10pct

tau1 -2.58 -1.95 -1.62

Les tests de racine unité

Dans le test de Dickey-Fuller augmenté, on rajoute des retards dans la régression,

```
> lags=1
> z=diff(X)
> n=length(z)
> z.diff=embed(z, lags+1)[,1]
> z.lag.1=X[(lags+1):n]
> k=lags+1
> z.diff.lag = embed(z, lags+1)[, 2:k]
> summary(lm(z.diff~0+z.lag.1+z.diff.lag ))
```

Call:

```
lm(formula = z.diff ~ 0 + z.lag.1 + z.diff.lag)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
z.lag.1	-0.005394	0.007361	-0.733	0.464
z.diff.lag	-0.028972	0.065113	-0.445	0.657

Residual standard error: 0.9666 on 236 degrees of freedom

```
Multiple R-squared: 0.003292,      Adjusted R-squared: -0.005155  
F-statistic: 0.3898 on 2 and 236 DF,  p-value: 0.6777
```

```
> summary(lm(z.diff~0+z.lag.1+z.diff.lag ))$coefficients[1,3]  
[1] -0.7328138
```

```
> summary(df)
```

```
#####  
# Augmented Dickey-Fuller Test Unit Root Test #  
#####
```

```
Test regression none
```

```
Value of test-statistic is: -0.7328
```

```
Critical values for test statistics:
```

```
      1pct  5pct 10pct  
tau1 -2.58 -1.95 -1.62
```

Et on peut rajouter autant de retards qu'on le souhaite. Mais on peut aussi tester un modèle avec constante

```
> summary(lm(z.diff~1+z.lag.1+z.diff.lag ))
```

Call:

```
lm(formula = z.diff ~ 1 + z.lag.1 + z.diff.lag)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.29175	0.13153	2.218	0.0275 *
z.lag.1	-0.03559	0.01545	-2.304	0.0221 *
z.diff.lag	-0.01976	0.06471	-0.305	0.7603

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9586 on 235 degrees of freedom

Multiple R-squared: 0.02313, Adjusted R-squared: 0.01482

F-statistic: 2.782 on 2 and 235 DF, p-value: 0.06393

voire une tendance linéaire

```
> temps=(lags+1):n
```

```
> summary(lm(z.diff~1+temps+z.lag.1+z.diff.lag ))
```

Call:

```
lm(formula = z.diff ~ 1 + temps + z.lag.1 + z.diff.lag)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.3227245	0.1502083	2.149	0.0327	*
temps	-0.0004194	0.0009767	-0.429	0.6680	
z.lag.1	-0.0329780	0.0166319	-1.983	0.0486	*
z.diff.lag	-0.0230547	0.0652767	-0.353	0.7243	

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.9603 on 234 degrees of freedom

Multiple R-squared: 0.0239, Adjusted R-squared: 0.01139

F-statistic: 1.91 on 3 and 234 DF, p-value: 0.1287

Dans le premier cas, les grandeurs d'intérêt sont

```
> summary(lm(z.diff~1+z.lag.1+z.diff.lag ))$coefficients[2,3]
[1] -2.303948
```

et pour le second

```
> summary(lm(z.diff~1+temps+z.lag.1+z.diff.lag ))$coefficients[3,3]
[1] -1.98282
```

Mais on peut aussi regarder si les spécifications de tendance ont du sens. Pour le modèle avec constante, on peut faire une analyse de la variance

```
> anova(lm(z.diff ~ z.lag.1 + 1 + z.diff.lag),lm(z.diff ~ 0 + z.diff.lag))
Analysis of Variance Table
```

```
Model 1: z.diff ~ z.lag.1 + 1 + z.diff.lag
```

```
Model 2: z.diff ~ 0 + z.diff.lag
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	235	215.97				
2	237	220.99	-2	-5.0231	2.7329	0.0671 .

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> anova(lm(z.diff ~ z.lag.1 + 1 + z.diff.lag),
+ lm(z.diff ~ 0 + z.diff.lag))$F[2]
[1] 2.732912
```

et dans le cas où on suppose qu'il y a une tendance

```
> anova(lm(z.diff ~ z.lag.1 + 1 + temps + z.diff.lag),lm(z.diff ~ 0 +z.diff.lag))
Analysis of Variance Table
```

```
Model 1: z.diff ~ z.lag.1 + 1 + temps + z.diff.lag
```

```
Model 2: z.diff ~ 0 + z.diff.lag
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	234	215.79				
2	237	220.99	-3	-5.1932	1.8771	0.1341

```
> anova(lm(z.diff ~ z.lag.1 + 1 + temps+ z.diff.lag),lm(z.diff ~ 0 + z.diff.lag))$F[2]
[1] 1.877091
```

```
> anova(lm(z.diff ~ z.lag.1 + 1 + temps+ z.diff.lag),lm(z.diff ~ 1+z.diff.lag))
Analysis of Variance Table
```

```
Model 1: z.diff ~ z.lag.1 + 1 + temps + z.diff.lag
```

```
Model 2: z.diff ~ 1 + z.diff.lag
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	234	215.79				
2	236	220.84	-2	-5.0483	2.7371	0.06683 .

```
---
```



```
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
> anova(lm(z.diff ~ z.lag.1 + 1 + temps+ z.diff.lag),lm(z.diff ~ 1+ z.diff.lag))$F[2]
[1] 2.737086
```

La commande pour faire ces tests est ici

```
> df=ur.df(X,type="drift",lags=1)
> summary(df)

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression drift

Value of test-statistic is: -2.3039 2.7329

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81
```

```
> df=ur.df(X,type="trend",lags=1)
> summary(df)

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Value of test-statistic is: -1.9828 1.8771 2.7371

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.99 -3.43 -3.13
phi2  6.22  4.75  4.07
phi3  8.43  6.49  5.47
```

Autres tests de racines unités, KPSS et PP

Pour le test **KPSS** (Kwiatkowski - Phillips - Schmidt - Shin), il faut spécifier s'il l'on suppose que la srie est de moyenne constante, ou si une tendance doit être prise en compte. Si l'on suppose une constante non-nulle, mais pas de tendance, on utilise

```
> summary(ur.kpss(X,type="mu"))
```

```
#####  
# KPSS Unit Root Test #  
#####
```

```
Test is of type: mu with 4 lags.
```

```
Value of test-statistic is: 0.972
```

```
Critical value for a significance level of:
```

```
10pct  5pct  2.5pct  1pct
```

```
critical values 0.347 0.463  0.574 0.739
```

et si on souhaite prendre en compte une tendance,

```
> summary(ur.kpss(X,type="tau"))
```

```
#####  
# KPSS Unit Root Test #  
#####
```

```
Test is of type: tau with 4 lags.
```

```
Value of test-statistic is: 0.5057
```

```
Critical value for a significance level of:
```

```
10pct  5pct  2.5pct  1pct
```

```
critical values 0.119 0.146  0.176 0.216
```

L'hypothèse nulle correspond à l'absence de racine unité : plus la statistique de test est grande, plus on s'éloigne de la stationnarité (hypothèse nulle).

Autres tests de racines unités, KPSS et PP

Pour le test **PP** de Philippes-Perron, on a un test de type Dickey-Fuller,

```
> PP.test(X)
```

```
Phillips-Perron Unit Root Test
```

```
data: X
```

```
Dickey-Fuller = -2.0116, Trunc lag parameter = 4, p-value = 0.571
```

Modélisation de séries saisonnières

- Box & Jenkins (1970) ont introduit les processus *ARIMA* **saisonniers**, *SARIMA*, avec double saisonnalité, qui peuvent s'approcher de la décomposition proposée par Harvey (1989), ou Hylleberg, Engle, Granger & Yoo (1990)
- un autre classe de modèle propose que les paramètres autoégressifs dépendent de la saisonnalité, ce qui donne les modèles *ARIMA* **périodique**, introduite par Franses & Paap (1996)

Petite digression, les modèles VAR

On appelle **processus vectoriel autoregressif d'ordre p** , noté $VAR(p)$, un processus stationnaire $(\mathbf{Y}_t) = (Y_{1,t}, \dots, Y_{d,t})$ vérifiant une relation du type

$$\mathbf{Y}_t - \sum_{i=1}^p \Phi_i \mathbf{Y}_{t-i} = \boldsymbol{\varepsilon}_t \text{ pour tout } t \in \mathbb{Z}, \quad (8)$$

où les Φ_i sont des matrices (réelles) $d \times d$ et $(\boldsymbol{\varepsilon}_t)$ est un bruit blanc de matrice de variance Σ , avec $\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t-k}) = \mathbf{0}$.

Par exemple, un $VAR(1)$, en dimension 2, est de la forme

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ \phi_{2,1} & \phi_{2,2} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

Petite digression, les modèles VAR

Remarque Tout processus $VAR(p)$ peut se mettre sous une forme $VAR(1)$.

Considérons un processus $VAR(2)$ en dimension 2

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{1,1} & b_{2,1} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

On a un $VAR(1)$ (contraint) en dimension 4

$$\begin{bmatrix} \begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} \\ \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} & \begin{pmatrix} b_{1,1} & b_{2,1} \\ b_{2,1} & b_{2,2} \end{pmatrix} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{bmatrix} \begin{bmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} \\ \begin{pmatrix} Y_{1,t-2} \\ Y_{2,t-2} \end{pmatrix} \end{bmatrix} + \begin{bmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{bmatrix}$$

Processus autorégressif périodique, $PAR(1)$

Considerons un processus $PAR(1)$ sur des **données trimestrielles**

$$Y_t = \phi_{1,s} Y_{t-1} + \varepsilon_t$$

que l'on pourrait écrire, en posant $\mathbf{Y}_T = (Y_{1,T}, Y_{2,T}, Y_{3,T}, Y_{4,T})$, et en remplaçant le temps t par un couple (s, T) (trimestre $s = 1, 2, 3, 4$ et année $T = 1, 2, \dots, n$).

$$\mathbf{\Phi}_0 \mathbf{Y}_T = \mathbf{\Phi}_1 \mathbf{Y}_{T-1} + \varepsilon_T$$

avec

$$\mathbf{\Phi}_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\phi_{1,2} & 1 & 0 & 0 \\ 0 & -\phi_{1,3} & 1 & 0 \\ 0 & 0 & -\phi_{1,4} & 1 \end{pmatrix} \quad \text{et} \quad \mathbf{\Phi}_1 = \begin{pmatrix} 0 & 0 & 0 & \phi_{1,1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Processus autorégressif périodique, $PAR(1)$

On a une représentation avec des vecteurs de trimestres. L'équation caractéristique s'écrit

$$|\Phi_0 - \Phi_1 z| = (1 - \phi_{1,1}\phi_{1,2}\phi_{1,3}\phi_{1,4}z) = 0$$

i.e. on a une **racine unité** si $\phi_{1,1}\phi_{1,2}\phi_{1,3}\phi_{1,4} = 1$.

Processus autorégressif périodique, $PAR(2)$

Considérons un processus $PAR(2)$ sur des **données trimestrielles**

$$Y_t = \phi_{1,s}Y_{t-1} + \phi_{2,s}Y_{t-2} + \varepsilon_t$$

que l'on pourrait écrire, en posant $\mathbf{Y}_T = (Y_{1,T}, Y_{2,T}, Y_{3,T}, Y_{4,T})$, et en remplaçant le temps t par un couple (s, T) (trimestre $s = 1, 2, 3, 4$ et année $T = 1, 2, \dots, n$).

$$\mathbf{\Phi}_0 \mathbf{Y}_T = \mathbf{\Phi}_1 \mathbf{Y}_{T-1} + \varepsilon_T$$

avec

$$\mathbf{\Phi}_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\phi_{1,2} & 1 & 0 & 0 \\ -\phi_{2,3} & -\phi_{1,3} & 1 & 0 \\ 0 & -\phi_{2,4} & -\phi_{1,4} & 1 \end{pmatrix} \quad \text{et} \quad \mathbf{\Phi}_1 = \begin{pmatrix} 0 & 0 & \phi_{2,1} & \phi_{1,1} \\ 0 & 0 & 0 & \phi_{2,2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Processus autorégressif périodique, $PAR(p)$

On a une représentation avec des vecteurs de trimestres. Notons que pour un modèle $PAR(p)$,

$$\Phi_0 \mathbf{Y}_T = \Phi_1 \mathbf{Y}_{T-1} + \cdots + \Phi_p \mathbf{Y}_{T-p} + \varepsilon_T$$

i.e.

$$\mathbf{Y}_T = \Phi_0^{-1} \Phi_1 \mathbf{Y}_{T-1} + \cdots + \Phi_0^{-1} \Phi_p \mathbf{Y}_{T-p} + \Phi_0^{-1} \varepsilon_T$$

Dans l'exemple du $PAR(2)$,

$$\Phi_0^{-1} =$$

19.8 L'équation caractéristique s'écrit

$$|\Phi_0 - \Phi_1 z - \cdots - \Phi_p z^p| = 0$$

Dans le cas d'un processus $PAR(2)$,

$$|\Phi_0 - \Phi_1 z| = \begin{vmatrix} 1 & 0 & -\phi_{2,1}z & -\phi_{1,1}z \\ -\phi_{1,2} & 1 & 0 & -\phi_{2,2}z \\ -\phi_{2,3} & -\phi_{1,3} & 1 & 0 \\ 0 & -\phi_{2,4} & -\phi_{4,1} & 1 \end{vmatrix} = 0$$

Pour faire de la prévision avec un processus $PAR(1)$,

$${}_n\hat{Y}_{n+1} = \mathbb{E}(Y_{n+1}) = \mathbb{E}(\Phi_0^{-1}\Phi_1 Y_n + \Phi_0^{-1}\epsilon_{n+1}) = \Phi_0^{-1}\Phi_1 Y_n$$

L'erreur de prévision est ici

$${}_n\hat{Y}_{n+1} - Y_{n+1}$$

De la saisonnalité aux racines unités saisonnières

Consiérons le cas de donnés **trimestrielles**.

- modèle à saisonnalité déterministe

On suppose que $X_t = Z_t + \sum_{k=0}^{4-1} \gamma_k \mathbf{1}(t = k \text{ mod. } 4) + \varepsilon_t$, i.e.

$$X_t = Z_t + \gamma_1 \mathbf{1}(t \in T_1) + \gamma_2 \mathbf{1}(t \in T_2) + \gamma_3 \mathbf{1}(t \in T_3) + \gamma_4 \mathbf{1}(t \in T_4) + \varepsilon_t$$

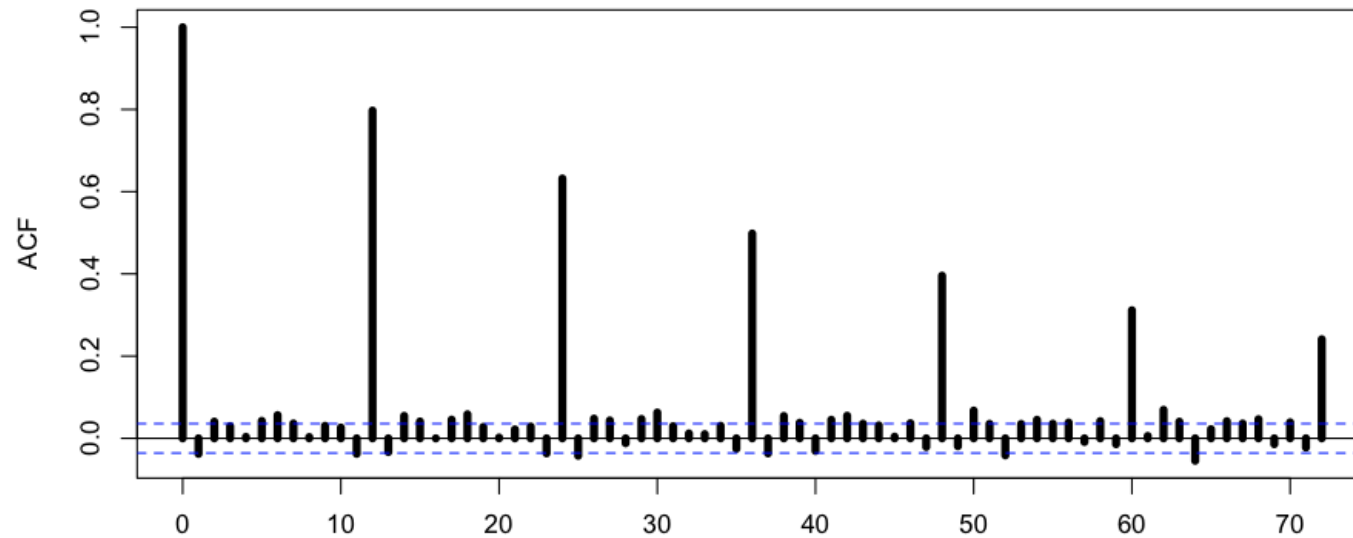
- modèle AR saisonnier stationnaire

On suppose que $X_t = \phi_4 X_{t-4} + \varepsilon_t$, ou plus généralement

$$X_t = \phi_1 X_{t-4} + \phi_2 X_{t-2 \times 4} + \cdots + \phi_p X_{t-p \times 4} + \varepsilon_t$$

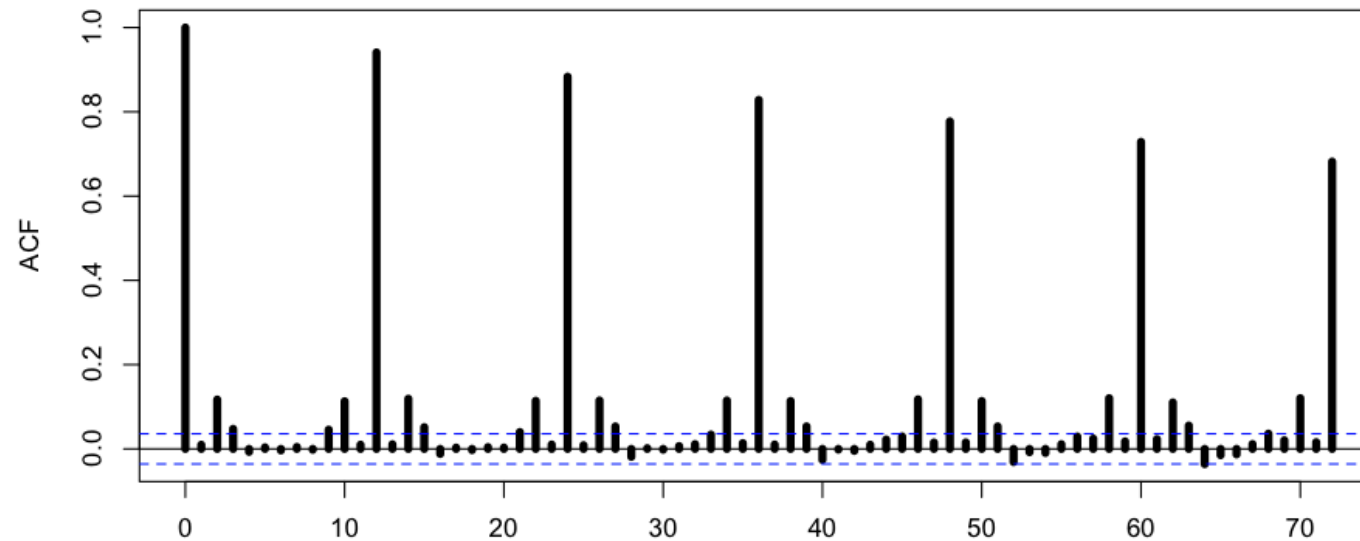
De la saisonnalité aux racines unités saisonnières

```
> X=rep(rnorm(T))  
> for(t in 13:T){X[t]=0.80*X[t-12]+E[t]}  
> acf(X[-(1:200)],lag=72,lwd=5)
```



De la saisonnalité aux racines unités saisonnières

```
> X=rep(rnorm(T))  
> for(t in 13:T){X[t]=0.95*X[t-12]+E[t]}  
> acf(X[-(1:200)],lag=72,lwd=5)
```

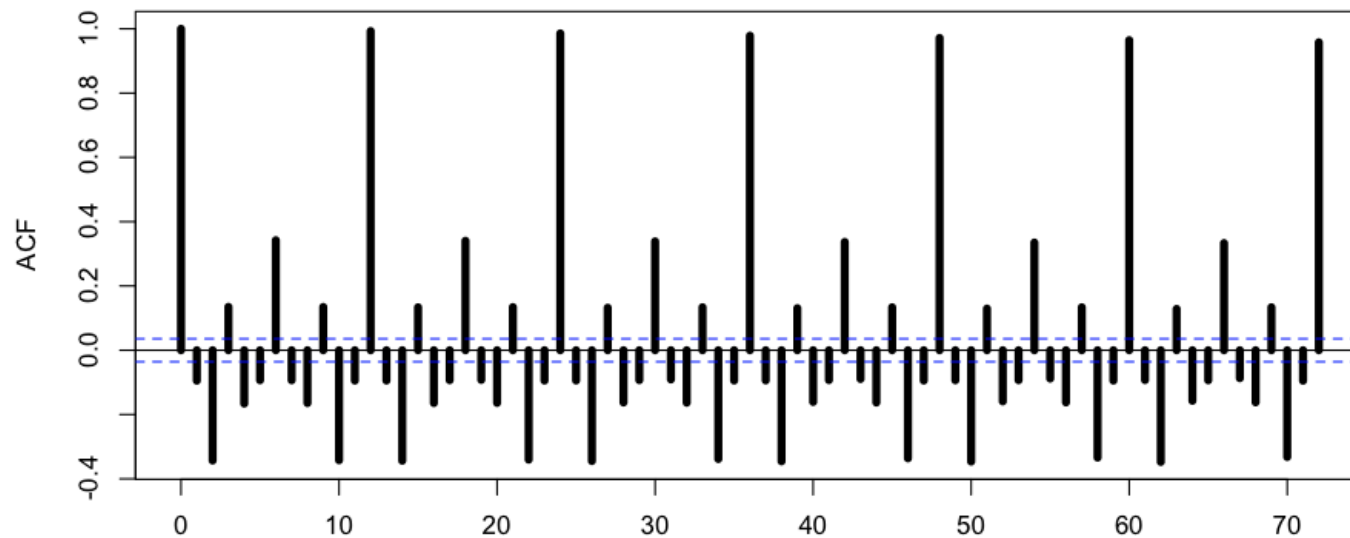


De la saisonnalité aux racines unités saisonnières

- modèle AR saisonnier non-stationnaire

On suppose que $X_t = X_{t-4} + \varepsilon_t$, i.e. $\phi_4 = 1$.

```
> X=rep(rnorm(T))  
> for(t in 13:T){X[t]=1.00*X[t-12]+E[t]}  
> acf(X[-(1:200)],lag=72,lwd=5)
```



De la saisonnalité aux racines unités saisonnières

Pour des données **trimestrielles**, le polynôme $\Phi(L) = (1 - L^4)$ admet 4 racines (dans \mathbb{C}), $+1$, $+i$, -1 et $-i$, i.e.

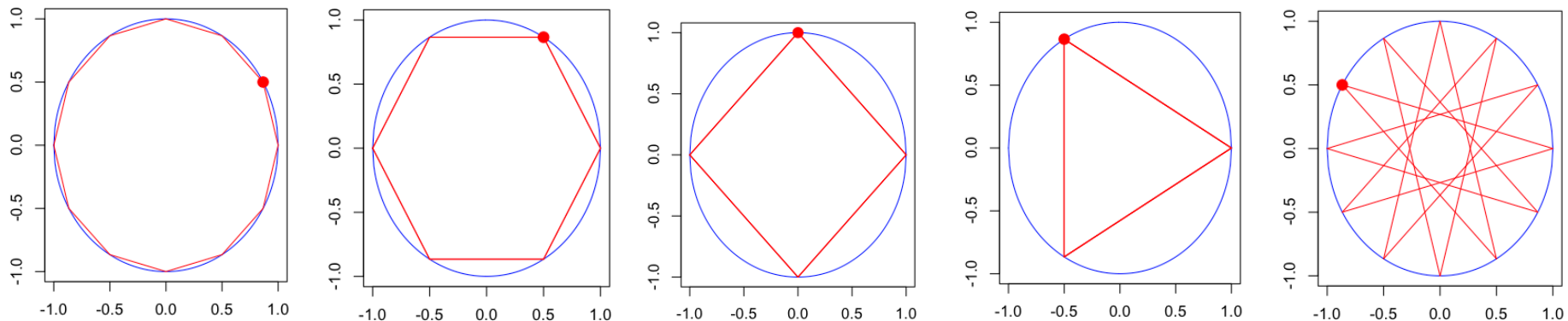
$$(1 - L^4) = (1 - L)(1 + L)(1 + L^2) = (1 - L)(1 + L)(1 - iL)(1 + iL)$$

Pour des données **mensuelle**, le polynôme $\Phi(L) = (1 - L^{12})$ admet 12 racines

$$\pm 1, \pm i, \pm \frac{1 \pm \sqrt{3}i}{2} \text{ et } \pm \frac{\sqrt{3} \pm i}{2}$$

ou encore $e^{i\frac{2k\pi}{12}}$, avec $k \in 1, 2, \dots, 12$.

Remarque La racine k va indiquer le nombre de cycles par an.

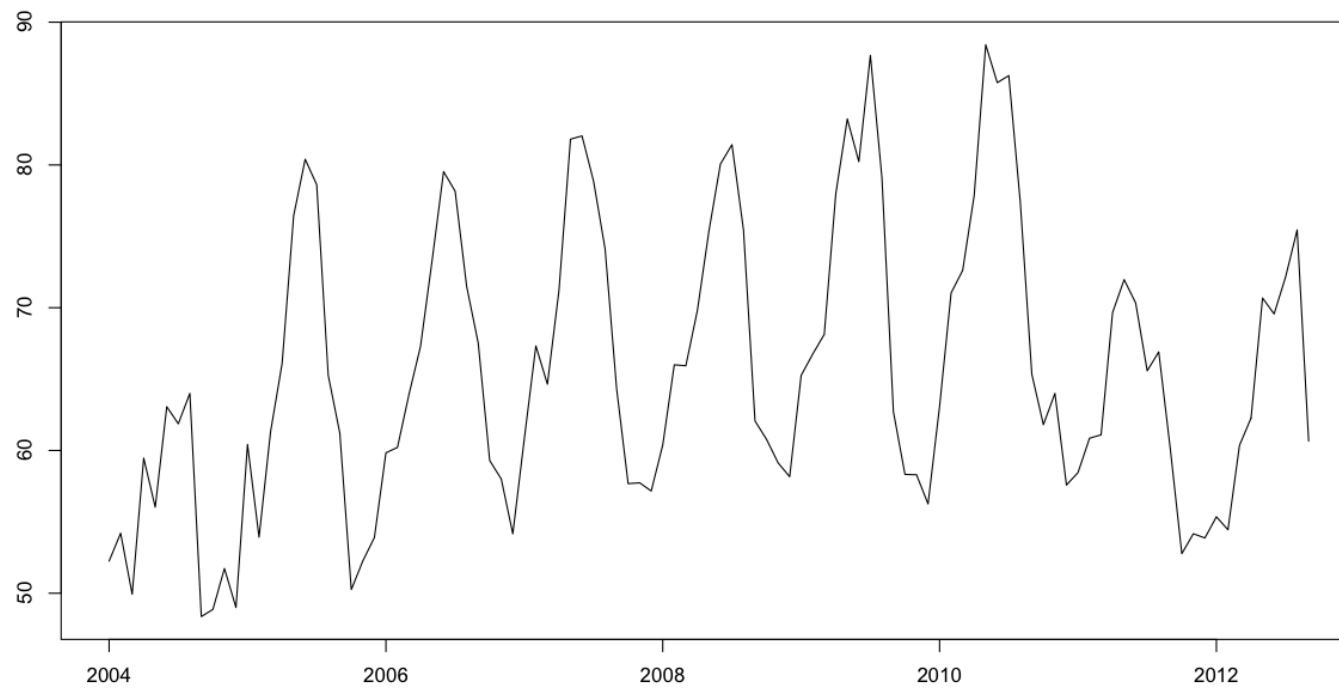


De la saisonnalité aux racines unités saisonnières

- Si la racine est d'argument $\frac{\pi}{6}$, on a 1 cycle par an (sur 12 périodes)
- Si la racine est d'argument $\frac{\pi}{3}$, on a 2 cycle par an (sur 6 périodes)
- Si la racine est d'argument $\frac{\pi}{2}$, on a 3 cycle par an (sur 4 périodes)
- Si la racine est d'argument $\frac{2\pi}{3}$, on a 4 cycle par an (sur 3 périodes)
- Si la racine est d'argument $\frac{5\pi}{6}$, on a 5 cycle par an

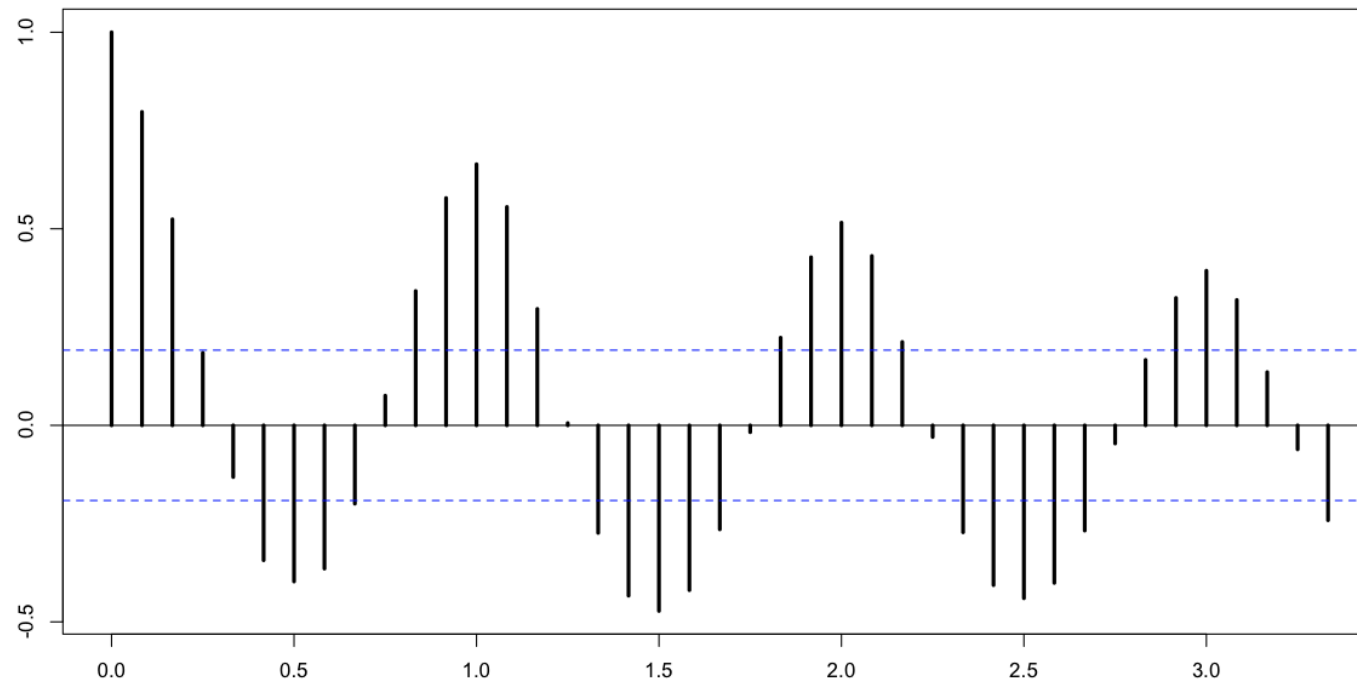
Les tests de racine unité saisonnière

```
> source("http://freakonometrics.blog.free.fr/public/data/sourcets.R")  
> plot(epilation)
```



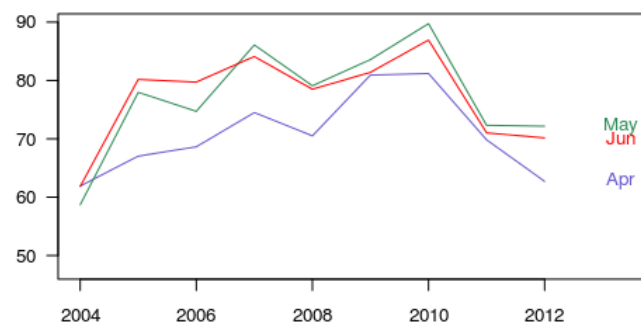
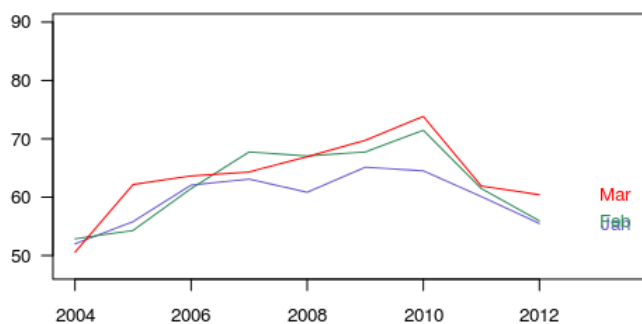
Les tests de racine unité saisonnière

```
> acf(epilation)
```



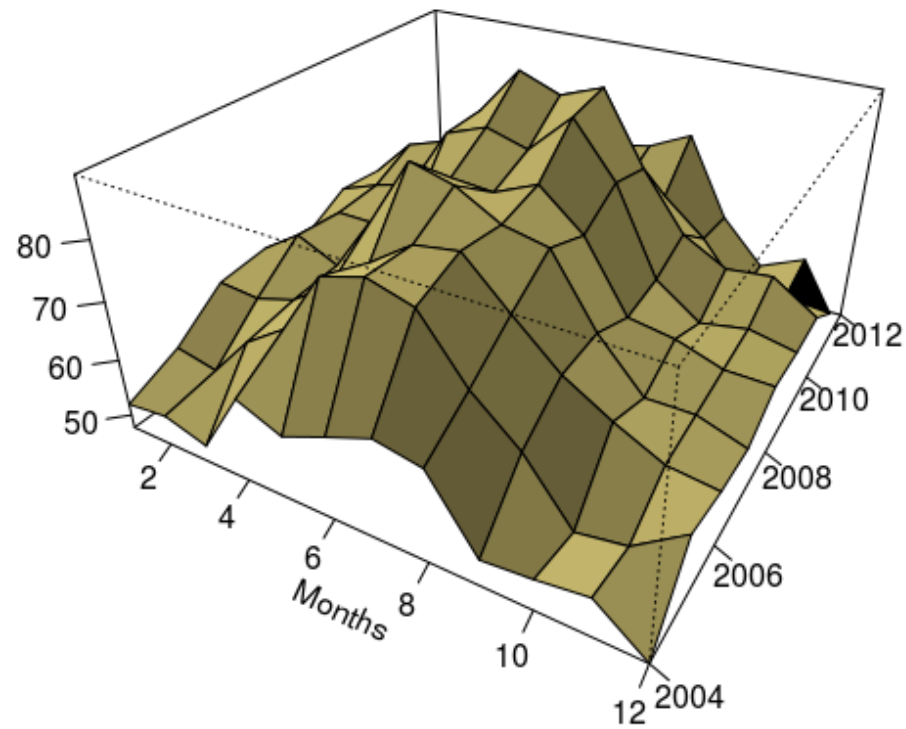
Les tests de racine unité saisonnière

```
> library(urroot)
> bbplot(epilation)
```



Les tests de racine unité saisonnière

```
> bb3D(epilation)
```



Les tests de racine unité saisonnière

Le test **CH** (de Canova & Hansen) suppose un modèle stationnaire, avec un cycle déterministe (hypothèse H_0 , modèle stationnaire)

$$X_t = \alpha X_{t-1} + \sum_{k=0}^{s-1} \gamma_k \mathbf{1}(t = k \bmod. s) + \varepsilon_t$$

Supposant que $|\alpha| < 1$, on va tester la stabilité des coefficients γ_k avec un test de type KPSS, i.e.

$$X_t = \alpha X_{t-1} + \sum_{k=0}^{s-1} \gamma_{k,t} \mathbf{1}(t = k \bmod. s) + \varepsilon_t$$

avec $\gamma_{k,t} = \gamma_{k,t-1} + u_t$. On va tester si $\text{Var}(u_t) = 0$, par un test de rapport de vraisemblance.

Les tests de racine unité saisonnière

```
> library(uroot)
> CH.test(epilation)
```

Null hypothesis: Stationarity.

Alternative hypothesis: Unit root.

Frequency of the tested cycles: $\pi/6$, $\pi/3$, $\pi/2$, $2\pi/3$, $5\pi/6$, π ,

L-statistic: 1.567

Lag truncation parameter: 12

Critical values:

0.10 0.05 0.025 0.01

2.49 2.75 2.99 3.27

Les tests de racine unité saisonnière

Le test **HEGY** suppose un modèle

$$\Delta_s X_t = (1 - L^s)X_t = \sum_{k=0}^{s-1} \pi_k W_{k,t-1} + \varepsilon_t$$

où (pour des données **trimestrielles**)

$$\left\{ \begin{array}{l} W_{1,t} = X_t + X_{t-1} + X_{t-2} + X_{t-3} = (1 + L + L^2 + L^3)X_t = (1 + L)(1 + L^2)X_t \\ W_{2,t} = X_t - X_{t-1} + X_{t-2} - X_{t-3} = (1 - L + L^2 - L^3)X_t = (1 - L)(1 + L^2)X_t \\ W_{3,t} = X_t - X_{t-2} \text{ et } W_{4,t} = -X_{t-1} + X_{t-3} \end{array} \right.$$

et on va tester (test de Fisher) si des π_k sont nuls.

Les tests de racine unité saisonnière

Si $\pi_k = 0 \forall k = 1, 2, 3, 4$, alors (X_t) est une marche aléatoire saisonnière

Si $\pi_1 = 0$, alors Φ tel que $\Phi(L)X_t = \varepsilon_t$ admet $+1$ pour racine

$$\Phi(L) = (1 - L^4) + \pi_2(1 - L)(1 + L^2)L + \pi_3(1 - L)(1 + L)L^2 + \pi_4(1 - L)(1 + L)L$$

(cf. test de Dickey-Fuller)

Si $\pi_2 = 0$, alors Φ tel que $\Phi(L)X_t = \varepsilon_t$ admet -1 pour racine

$$\Phi(L) = (1 - L^4) - \pi_1(1 + L)(1 + L^2)L + \pi_3(1 - L)(1 + L)L^2 + \pi_4(1 - L)(1 + L)L$$

(cycle semi-annuel).

Si $\pi_3 = 0$ et $\pi_4 = 0$, alors Φ tel que $\Phi(L)X_t = \varepsilon_t$ admet $\pm i$ pour racines

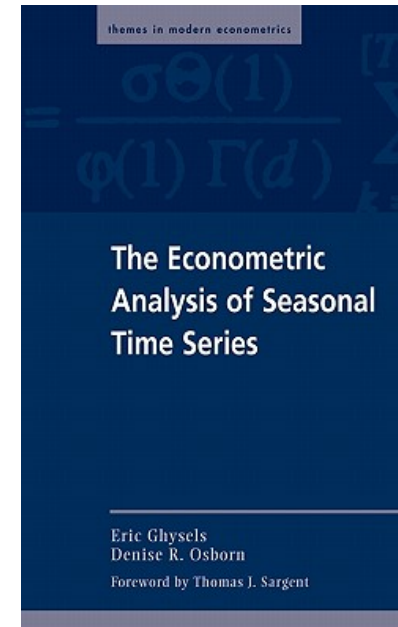
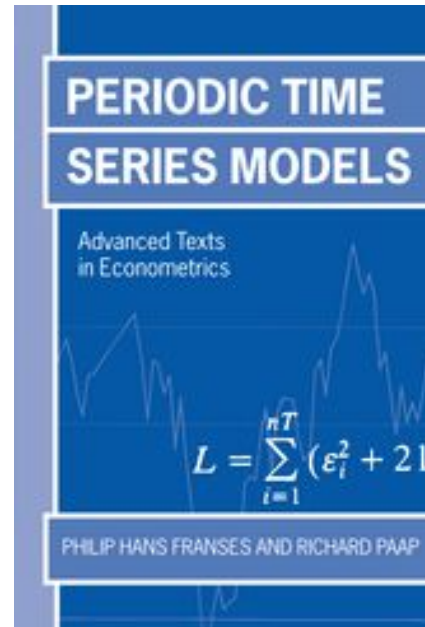
$$\Phi(L) = (1 - L^4) - \pi_1(1 + L)(1 + L^2)L + \pi_2(1 - L)(1 + L^2)$$

Si $\pi_k = 0 \forall k = 2, 3, 4$, alors Φ tel que $\Phi(L)X_t = \varepsilon_t$ admet $-1, \pm i$ pour racines

Les tests de racine unité saisonnière

Remarque dans la version la plus générale, on peut rajouter une tendance, des cycles déterministes, et des retards (cf. Dickey Fuller augmenté)

$$\Delta^{12} X_t = \alpha + \beta t + \sum_{k=0}^{s-1} \gamma_k \mathbf{1}(t = k \bmod s) + \sum_{k=0}^{s-1} \pi_k W_{k,t-1} + \sum_{k=1}^p \phi_k \Delta^{12} X_{t-k} + \varepsilon_t$$



Les tests de racine unité saisonnière

```
> library(uroot)
> HEGY.test(epilation, itsd=c(1,1,c(1:11)))
```

Null hypothesis: Unit root.

Alternative hypothesis: Stationarity.

	Stat.	p-value
tpi_1	-1.734	0.100
tpi_2	-3.505	0.010
Fpi_3:4	4.787	0.010
Fpi_5:6	3.082	0.010
Fpi_7:8	7.759	0.076
Fpi_9:10	9.395	0.100
Fpi_11:12	3.807	0.010
Fpi_2:12	11.641	NA
Fpi_1:12	11.196	NA

Lag orders: 10 12

Number of available observations: 83

Identification des ordres p et q d'un $ARMA(p, q)$

La règle la plus simple à utiliser est celle basée sur les propriétés des fonctions d'autocorrélations pour les processus AR et MA

- pour un processus $AR(p)$
 - $\psi(p) \neq 0$ et $\psi(h) = 0$ pour $h > p$
 - on cherche le plus grand p au delà duquel $\hat{\psi}(h)$ est non-significatif
- pour un processus $MA(q)$
 - $\rho(q) \neq 0$ et $\rho(h) = 0$ pour $h > q$
 - on cherche le plus grand q au delà duquel $\hat{\rho}(h)$ est non-significatif

Identification des ordres p et q d'un $ARMA(p, q)$

```
> autoroute=read.table(  
+"http://freakonometrics.blog.free.fr/public/data/autoroute.csv",  
+ header=TRUE,sep=";")  
> a7=autoroute$a007  
> A7=ts(a7,start = c(1989, 9), frequency = 12)
```

Il est possible d'utiliser la [fonction d'autocorrélation étendue](#) (EACF), introduite par Tsay & Tiao (1984),

Identification des ordres p et q d'un $ARMA(p, q)$

```
> EACF=eacf(A7,13,13)
```

```
AR/MA
```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	o	o	x	x	x	x	x	o	o	x	x	x	o
1	x	x	o	o	x	x	x	o	o	o	x	x	x	x
2	o	x	o	o	o	o	x	o	o	o	o	x	x	o
3	o	x	o	x	o	o	o	o	o	o	o	x	x	x
4	x	x	x	x	o	o	o	o	o	o	o	x	o	o
5	x	x	o	o	o	x	o	o	o	o	o	x	o	x
6	x	x	o	o	o	x	o	o	o	o	o	x	o	o
7	o	x	o	o	o	x	o	o	o	o	o	x	o	o
8	o	x	o	o	o	x	o	o	o	o	o	x	o	o
9	x	x	o	o	x	o	x	o	o	o	o	x	o	o
10	x	x	o	o	x	o	x	o	o	o	o	x	o	o
11	x	x	o	o	x	x	x	o	o	o	x	x	o	o
12	x	x	o	o	o	o	o	o	o	o	o	o	o	o
13	x	x	o	o	o	o	o	o	o	o	o	o	o	o

Identification des ordres p et q d'un $ARMA(p, q)$

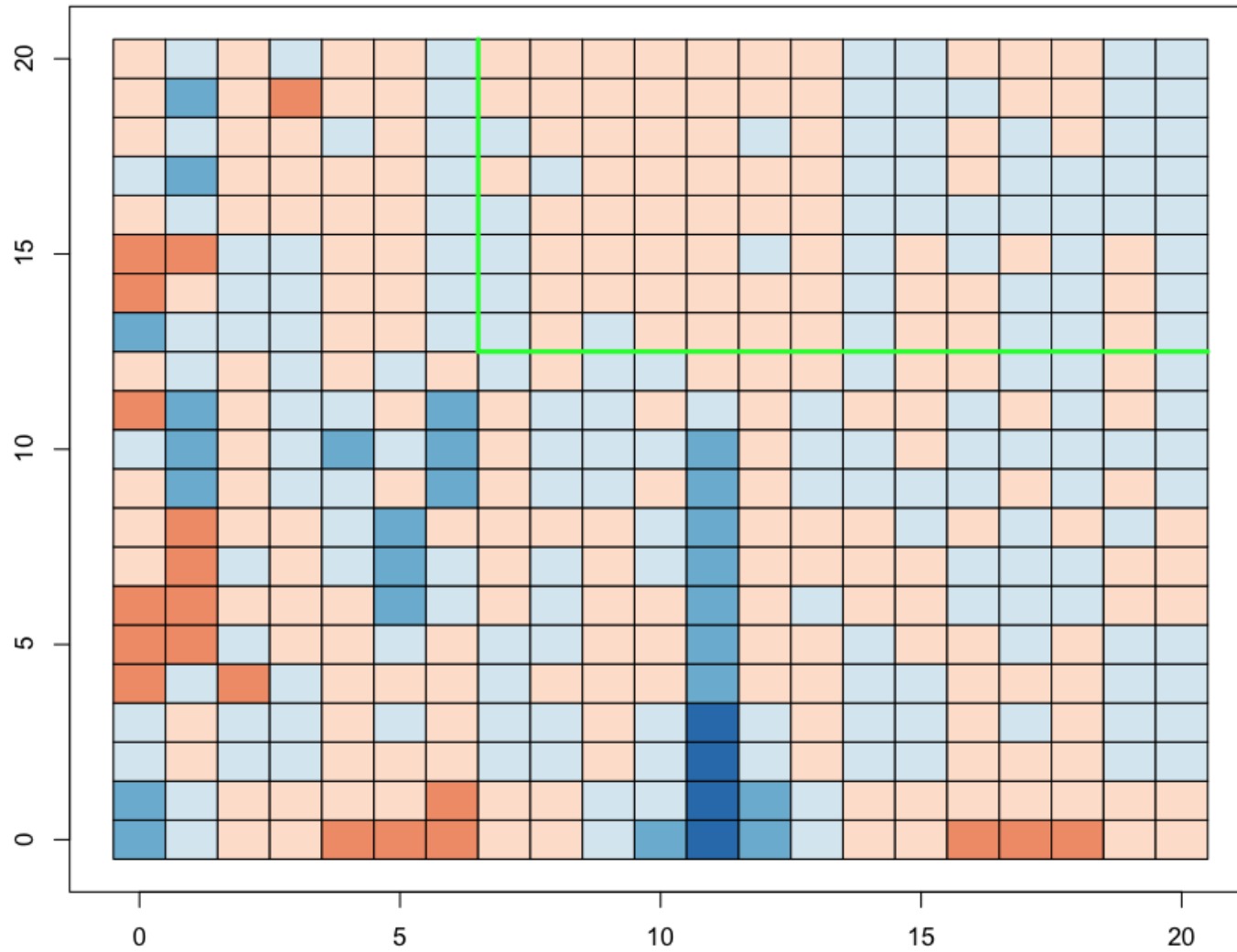
```
> EACF$eacf
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	0.64	0.21	-0.02	-0.24	-0.51	-0.61	-0.52	-0.28	-0.08	0.15
[2,]	0.48	0.27	-0.02	-0.06	-0.25	-0.28	-0.33	-0.21	-0.08	0.02
[3,]	0.10	-0.22	0.03	0.08	-0.15	-0.01	-0.25	0.13	0.00	-0.11
[4,]	0.13	-0.27	0.04	0.28	-0.18	0.03	-0.08	0.11	0.02	-0.21
[5,]	-0.50	0.31	-0.34	0.31	-0.19	-0.09	-0.09	0.08	0.00	-0.02
[6,]	-0.42	-0.48	0.16	-0.09	-0.07	0.29	-0.09	0.01	0.01	-0.03
[7,]	-0.47	-0.43	-0.04	-0.10	0.00	0.36	0.01	-0.03	0.03	-0.07
[8,]	-0.21	-0.40	0.09	-0.06	0.02	0.38	0.16	-0.02	0.02	-0.20
[9,]	-0.14	-0.50	-0.10	-0.06	0.11	0.38	-0.01	-0.02	-0.08	-0.20
[10,]	-0.29	0.48	0.00	0.18	0.27	-0.12	0.41	0.00	0.16	0.15
[11,]	0.24	0.48	-0.04	0.02	0.46	0.10	0.37	-0.10	0.05	0.16
[12,]	-0.59	0.49	-0.18	0.05	0.31	-0.32	0.34	-0.16	0.07	0.16
[13,]	-0.31	0.31	-0.12	0.16	-0.11	0.04	-0.04	0.12	-0.09	0.00
[14,]	0.47	0.26	0.11	0.13	-0.01	-0.01	0.00	0.07	-0.03	0.02

Identification des ordres p et q d'un $ARMA(p, q)$

On cherche un coin à partir duquel les autocorrélations s'annulent

```
> library(RColorBrewer)
> CL=brewer.pal(6, "RdBu")
> ceacf=matrix(as.numeric(cut(EACF$eacf,
+ ((-3):3)/3,labels=1:6)),nrow(EACF$eacf),
+ ncol(EACF$eacf))
> for(i in 1:ncol(EACF$eacf)){
+ for(j in 1:nrow(EACF$eacf)){
+ polygon(c(i-1,i-1,i,i)-.5,c(j-1,j,j,j-1)-.5,
+ col=CL[ceacf[j,i]])
+ }}
```



Estimation des $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$

Considérons un processus autorégressif, e.g. un $AR(2)$ dont on souhaite estimer les paramètres,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

où (ε_t) est un bruit blanc de variance σ^2 .

```
> phi1=.5; phi2=-.4; sigma=1.5
> set.seed(1)
> n=240
> WN=rnorm(n,sd=sigma)
> Z=rep(NA,n)
> Z[1:2]=rnorm(2,0,1)
> for(t in 3:n){Z[t]=phi1*Z[t-1]+phi2*Z[t-2]+WN[t]}
```

Estimation des $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$

- utilisation les moindres carrés

On peut réécrire la dynamique du processus sous forme matricielle, car

$$\begin{bmatrix} Y_3 \\ Y_4 \\ \vdots \\ Y_t \end{bmatrix} = \begin{bmatrix} Y_2 & Y_1 \\ Y_3 & Y_2 \\ \vdots & \vdots \\ Y_{t-1} & Y_{t-2} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_3 \\ \varepsilon_4 \\ \vdots \\ \varepsilon_t \end{bmatrix}$$

L'estimation par moindres carrés (classiques) donne ici

```
> base=data.frame(Y=Z[3:n],X1=Z[2:(n-1)],X2=Z[1:(n-2)])
> regression=lm(Y~0+X1+X2,data=base)
> summary(regression)
```

Call:

```
lm(formula = Y ~ 0 + X1 + X2, data = base)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.3491	-0.8890	-0.0762	0.9601	3.6105

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
X1	0.45107	0.05924	7.615	6.34e-13 ***
X2	-0.41454	0.05924	-6.998	2.67e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.449 on 236 degrees of freedom

Multiple R-squared: 0.2561, Adjusted R-squared: 0.2497

F-statistic: 40.61 on 2 and 236 DF, p-value: 6.949e-16

> regression\$coefficients

X1	X2
0.4510703	-0.4145365

> summary(regression)\$sigma

[1] 1.449276

Estimation des $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$

- utilisation les équations de Yule-Walker

On a vu que les fonctions d'autocorrélations et d'autocovariance étaient liées par un système d'équations,

$$\begin{cases} \gamma(1) = \phi_1 \gamma(0) + \phi_2 \gamma(1) \\ \gamma(2) = \phi_1 \gamma(1) + \phi_2 \gamma(0) \end{cases} \quad \text{i.e.} \quad \begin{bmatrix} \gamma(1) \\ \gamma(2) \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(-1) \\ \gamma(1) & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}$$

ou encore,

$$\begin{bmatrix} 1 & \rho(-1) \\ \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix} \quad \text{i.e.} \quad \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho(-1) \\ \rho(1) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix}$$

Estimation des $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$

```
> rho1=cor(Z[1:(n-1)],Z[2:n])
> rho2=cor(Z[1:(n-2)],Z[3:n])
> A=matrix(c(1,rho1,rho1,1),2,2)
> b=matrix(c(rho1,rho2),2,1)
> (PHI=solve(A,b))
[,1]
[1,] 0.4517579
[2,] -0.4155920
```

On peut ensuite extraire le bruit, et calculer son écart-type afin d'estimer σ^2

```
> estWN=base$Y-(PHI[1]*base$X1+PHI[2]*base$X2)
> sd(estWN)
[1] 1.445706
```


Estimation des $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$

- utilisation le **maximum de vraisemblance** (conditionnel)

On va supposer ici que le bruit blanc est **Gaussien**, $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, de telle sorte que

$$X_t | X_{t-1}, X_{t-2} \sim \mathcal{N}(\phi_1 X_{t-1} + \phi_2 X_{t-2}, \sigma^2)$$

La log-vraisemblance conditionnelle est alors

```
> CondLogLik=function(A,TS){
+ phi1=A[1]; phi2=A[2]
+ sigma=A[3] ; L=0
+ for(t in 3:length(TS)){
+ L=L+dnorm(TS[t],mean=phi1*TS[t-1]+
+ phi2*TS[t-2],sd=sigma,log=TRUE)}
+ return(-L)}
```

Estimation des $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$

et on peut alors optimiser cette fonction

```
> LogL=function(A) CondLogLik(A,TS=Z)
> optim(c(0,0,1),LogL)
$par
[1] 0.4509685 -0.4144938 1.4430930

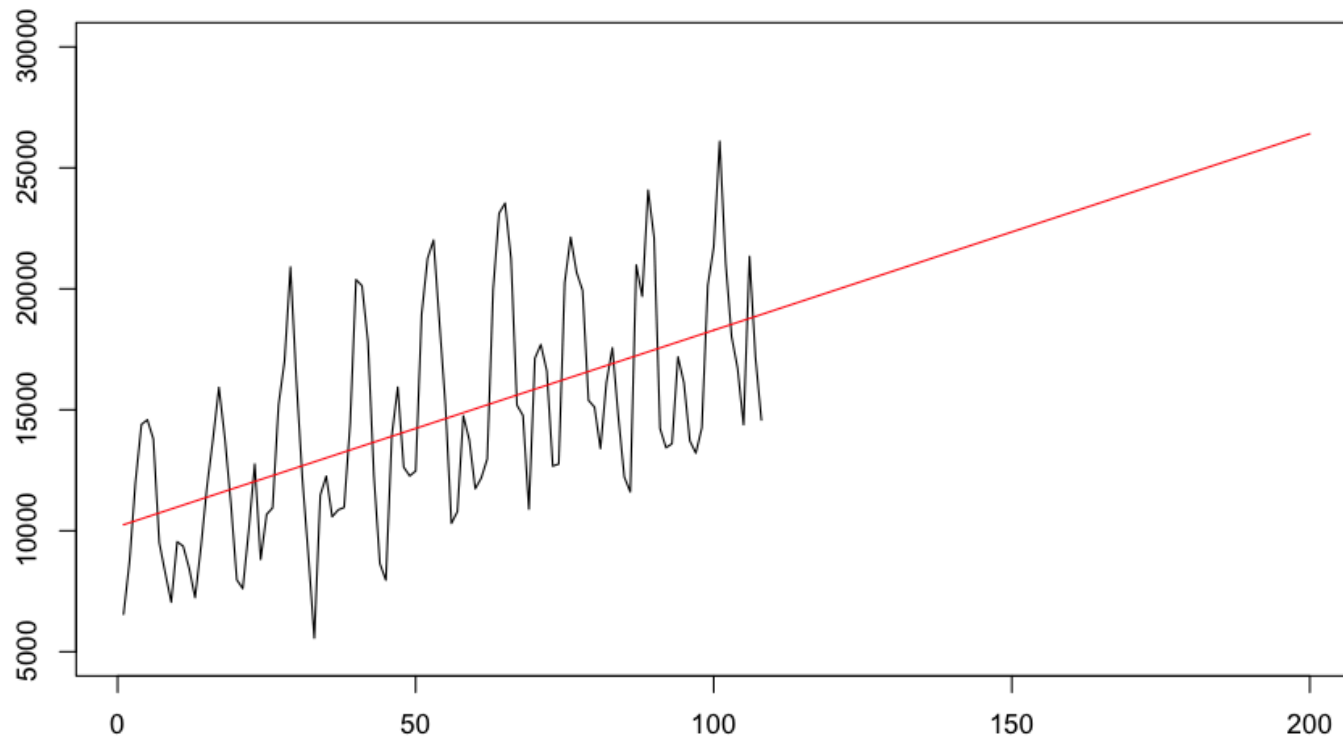
$value
[1] 425.0164
```

Estimation d'un modèle *ARIMA*

On considère la série de vente de voitures au Québec,

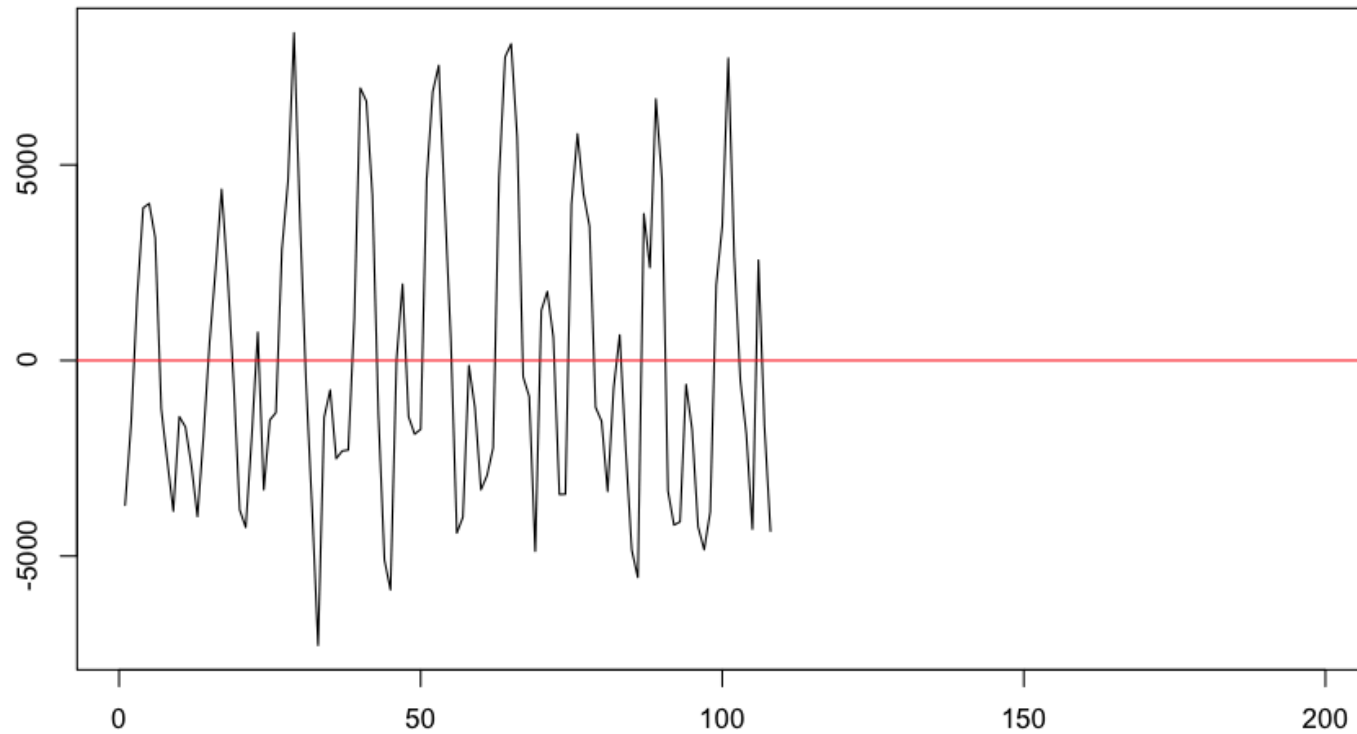
```
> source("http://freakonometrics.blog.free.fr/public/data/sourcets.R")
> X=as.numeric(quebec)
> temps=1:length(X)
> base=data.frame(temps,X)
> reg=lm(X~temps)
> T=1:200
> Xp=predict(reg,newdata=data.frame(temps=T))
> plot(temps,X,xlim=c(1,200),ylim=c(5000,30000),type="l")
> lines(T,Xp,col="red")
```

Estimation d'un modèle *ARIMA*



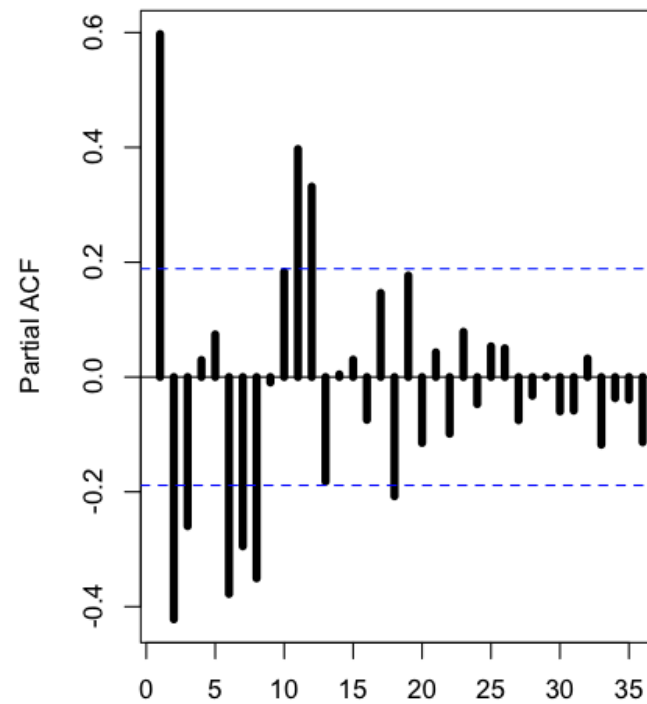
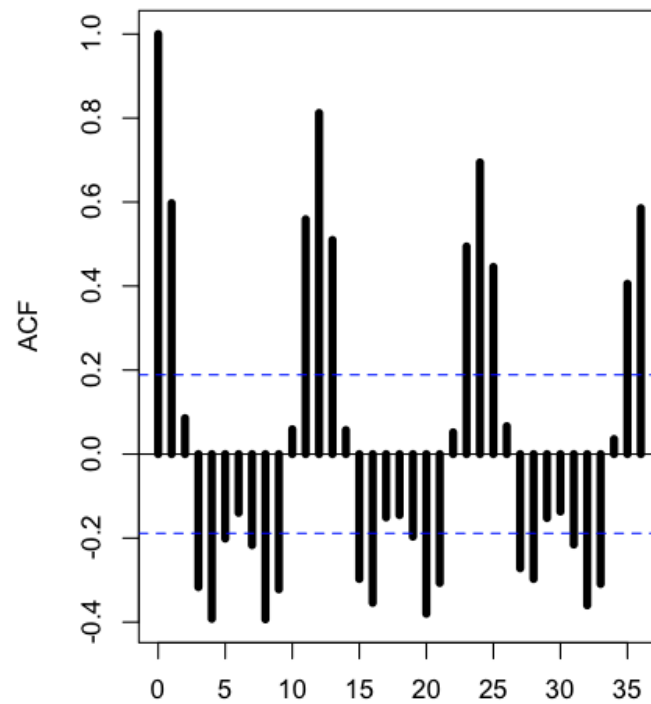
Estimation d'un modèle *ARIMA*

```
> Y=X-predict(reg)  
> plot(temps,Y,type="l",xlim=c(1,200))
```



Estimation d'un modèle *ARIMA*

- > `acf(Y)`
- > `pacf(Y)`



Estimation d'un modèle *ARIMA*

```
> fit.ar12=arima(Y,order=c(12,0,0),include.mean=FALSE)
```

```
> fit.ar12
```

```
Series: Y
```

```
ARIMA(12,0,0) with non-zero mean
```

```
Coefficients:
```

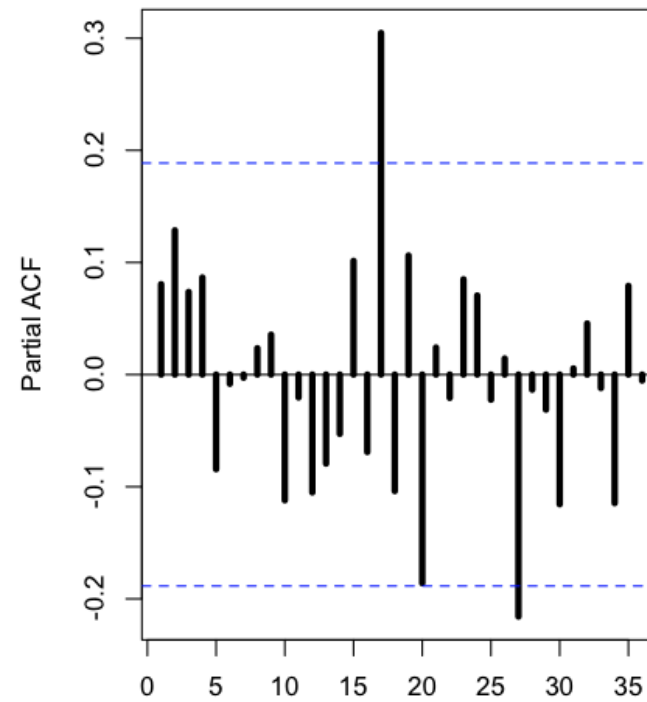
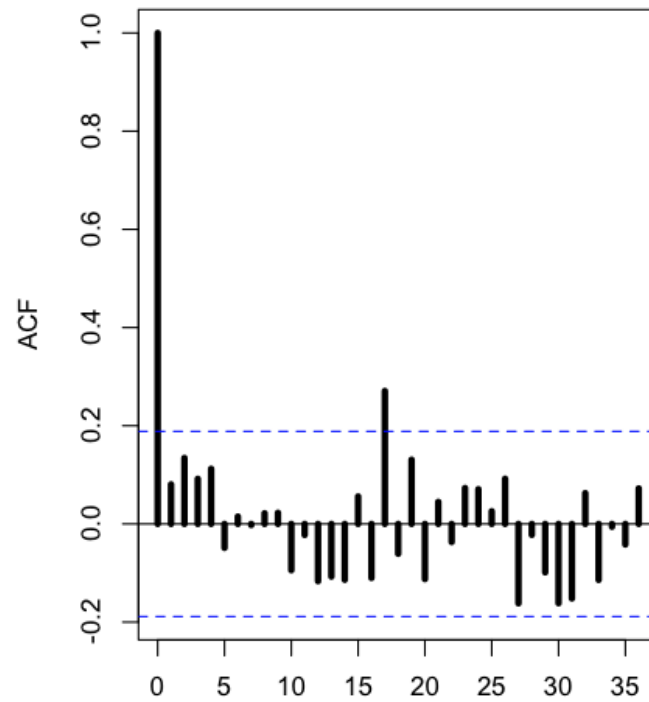
	ar1	ar2	ar3	ar4	ar5	ar6	ar7
	0.1975	0.0832	-0.1062	-0.1212	0.1437	-0.1051	0.0319
s.e.	0.0838	0.0809	0.0826	0.0843	0.0850	0.0833	0.0854
	ar9	ar10	ar11	ar12	intercept		
	-0.0332	-0.0616	0.2635	0.4913	-148.3180		
s.e.	0.0853	0.0840	0.0840	0.0841	384.5095		

```
sigma^2 estimated as 2177974: log likelihood=-946.75
```

```
AIC=1921.51 AICc=1926.03 BIC=1959.06
```

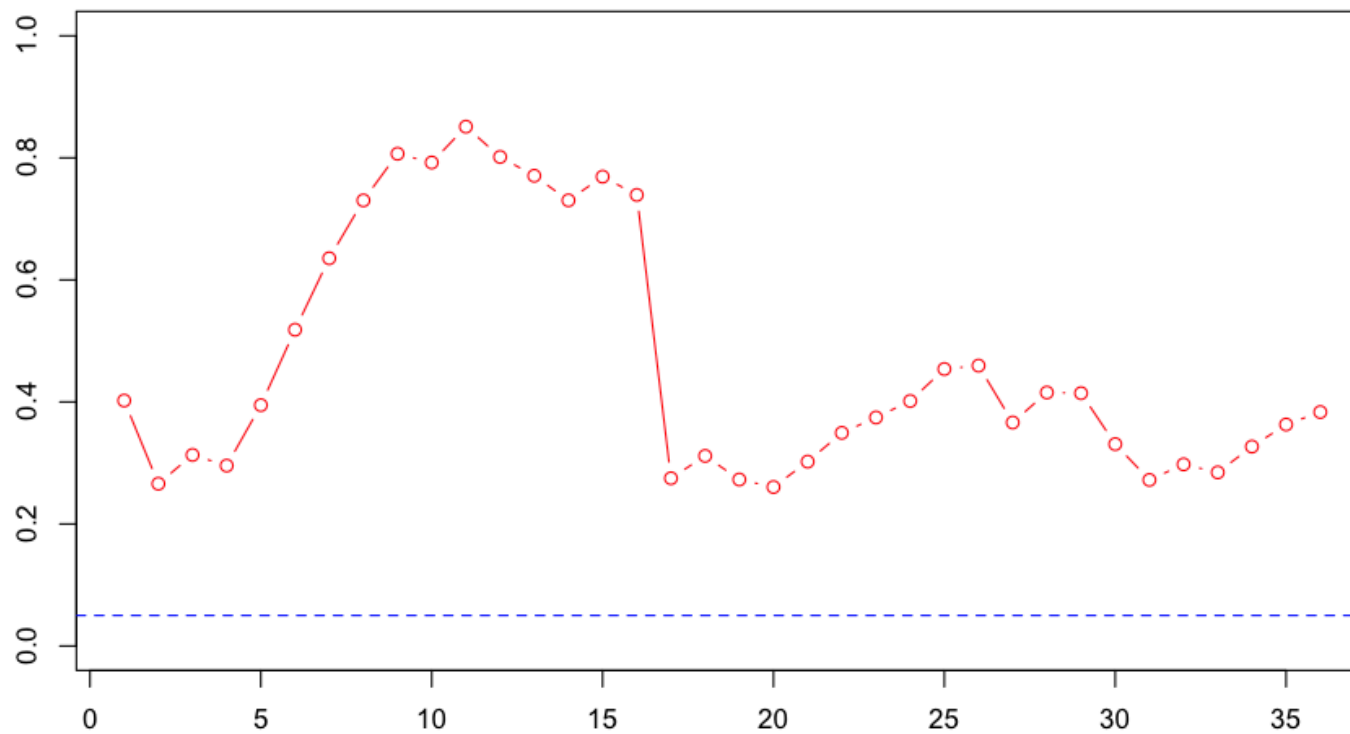
Estimation d'un modèle *ARIMA*

```
> acf(residuals(fit.ar12))  
> pacf(residuals(fit.ar12))
```



Estimation d'un modèle *ARIMA*

```
> BP=function(h) Box.test(residuals(fit.ar12),lag=h, type='Box-Pierce')$p.value  
> plot(1:36,Vectorize(BP)(1:36),type='b',col="red")  
> abline(h=.05,lty=2,col="blue")
```



Estimation d'un modèle *ARIMA*

En manque d'inspiration ?

```
> library(caschrono)
> armaselect(Y,nbmod=5)
      p q      sbc
[1,] 14 1 1635.214
[2,] 12 1 1635.645
[3,] 15 1 1638.178
[4,] 12 3 1638.297
[5,] 12 4 1639.232
```

Estimation d'un modèle *ARIMA*

```
> fit.arma12.1=arima(Y,order=c(12,0,1),include.mean=FALSE)
```

```
> fit.arma12.1
```

```
Series: Y
```

```
ARIMA(12,0,1) with non-zero mean
```

```
Coefficients:
```

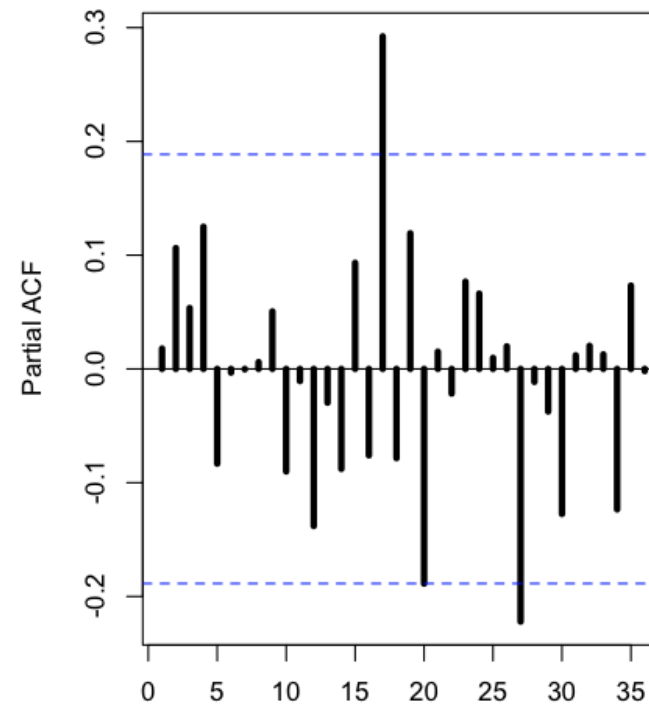
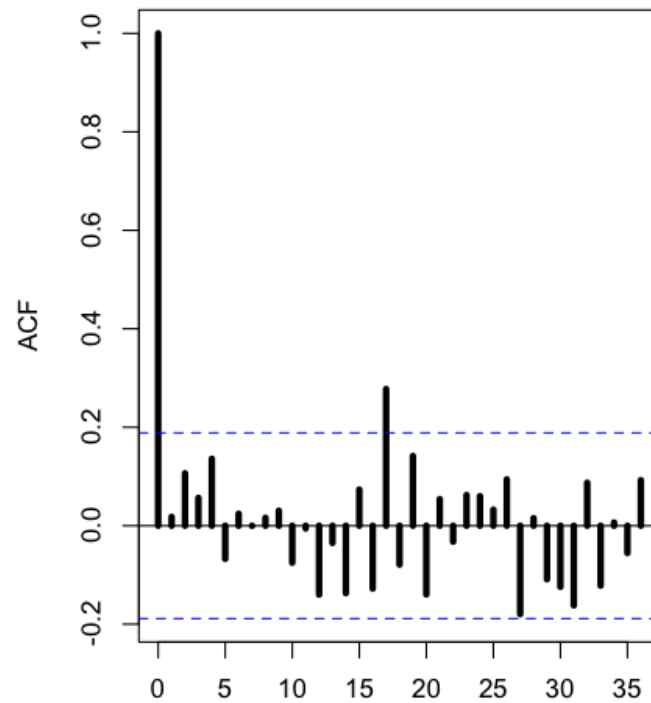
	ar1	ar2	ar3	ar4	ar5	ar6	ar7
	0.0301	0.1558	-0.0941	-0.1461	0.1063	-0.0688	-0.002
s.e.	0.1235	0.0854	0.0757	0.0784	0.0807	0.0774	0.080
	ar9	ar10	ar11	ar12	ma1	intercept	
	-0.0646	-0.0798	0.2538	0.5786	0.2231	-131.3495	
s.e.	0.0802	0.0766	0.0751	0.0861	0.1393	368.8156	

```
sigma^2 estimated as 2127759: log likelihood=-945.65
```

```
AIC=1921.31 AICc=1926.52 BIC=1961.54
```

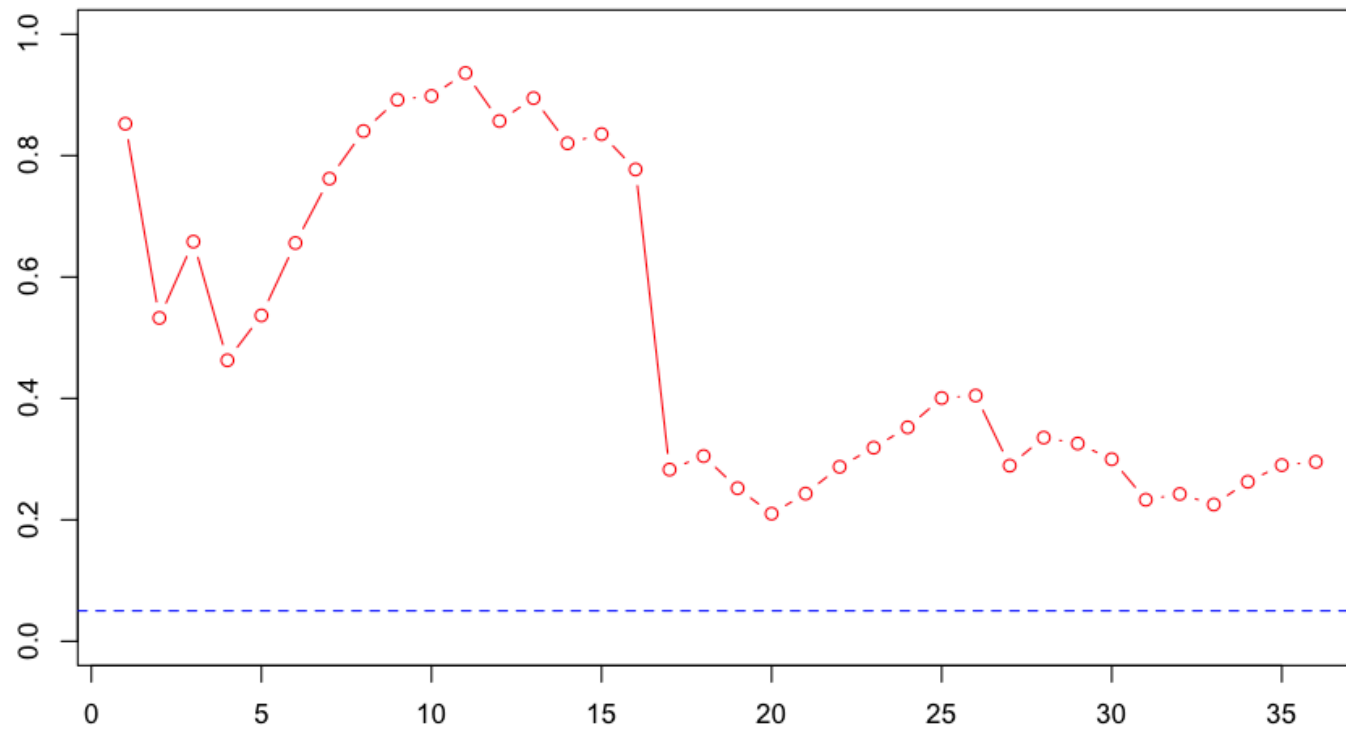
Estimation d'un modèle *ARIMA*

```
> acf(residuals(fit.arma12.1))  
> pacf(residuals(fit.arma12.1))
```



Estimation d'un modèle *ARIMA*

```
> BP=function(h) Box.test(residuals(fit.arma12.1),lag=h, type='Box-Pierce')$p.value  
> plot(1:36,Vectorize(BP)(1:36),type='b',col="red")  
> abline(h=.05,lty=2,col="blue")
```



Estimation d'un modèle *ARIMA*

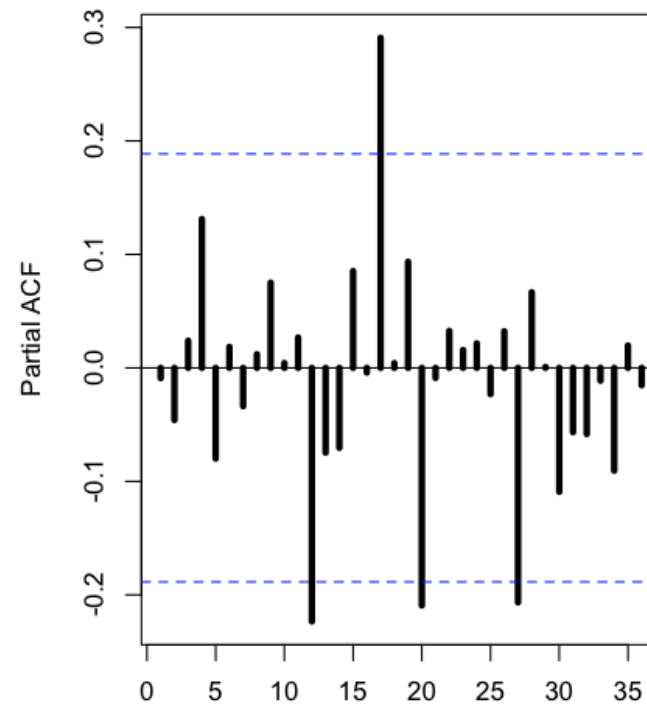
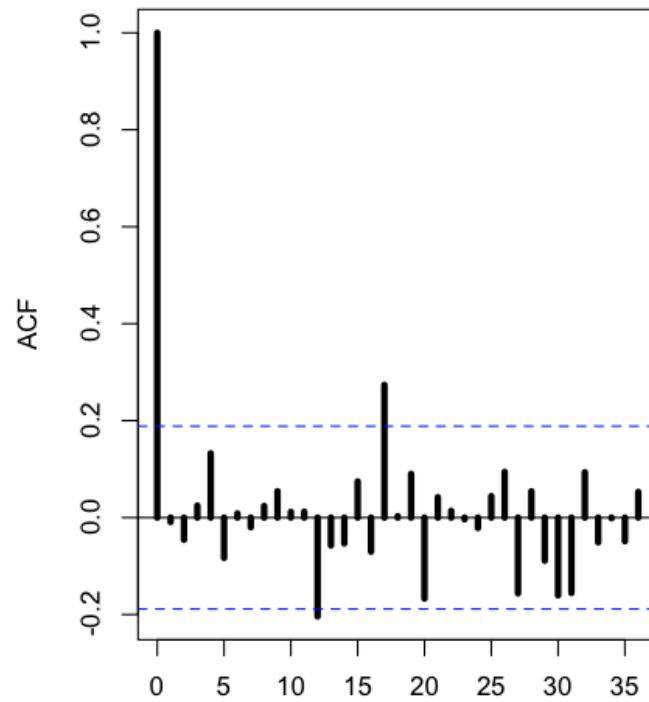
```
> fit.ar14=arima(Y,order=c(14,0,0),method="CSS",include.mean=FALSE)
> fit.ar14
Series: Y
ARIMA(14,0,0) with non-zero mean

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ar6      ar7
 0.2495  0.2105 -0.0584 -0.1569  0.1282 -0.1152  0.0268
s.e.  0.0956  0.0972  0.0854  0.0830  0.0838  0.0840  0.0847
      ar9      ar10      ar11      ar12      ar13      ar14  intercept
-0.0327 -0.1116  0.2649  0.5887 -0.1575 -0.1572      80.5
s.e.  0.0855  0.0851  0.0853  0.0886  0.1031  0.0999  338.9

sigma^2 estimated as 2218612:  part log likelihood=-942.31
```

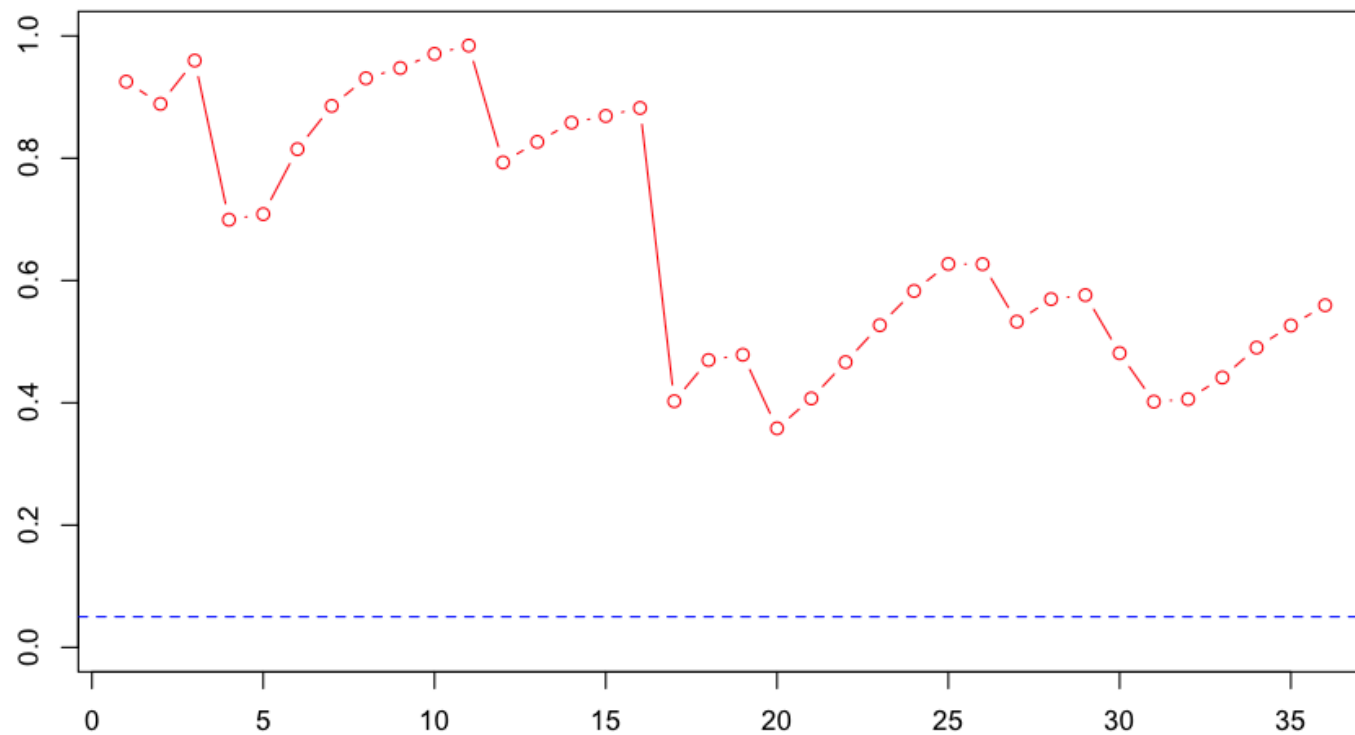
Estimation d'un modèle *ARIMA*

```
> acf(residuals(fit.ar14))  
> pacf(residuals(fit.ar14))
```



Estimation d'un modèle *ARIMA*

```
> BP=function(h) Box.test(residuals(fit.ar14),lag=h, type='Box-Pierce')$p.value  
> plot(1:36,Vectorize(BP)(1:36),type='b',col="red")  
> abline(h=.05,lty=2,col="blue")
```

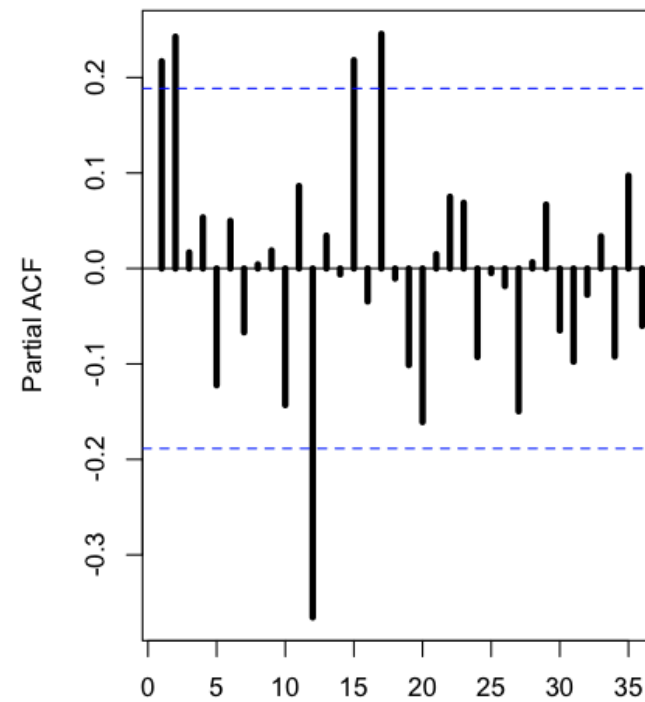
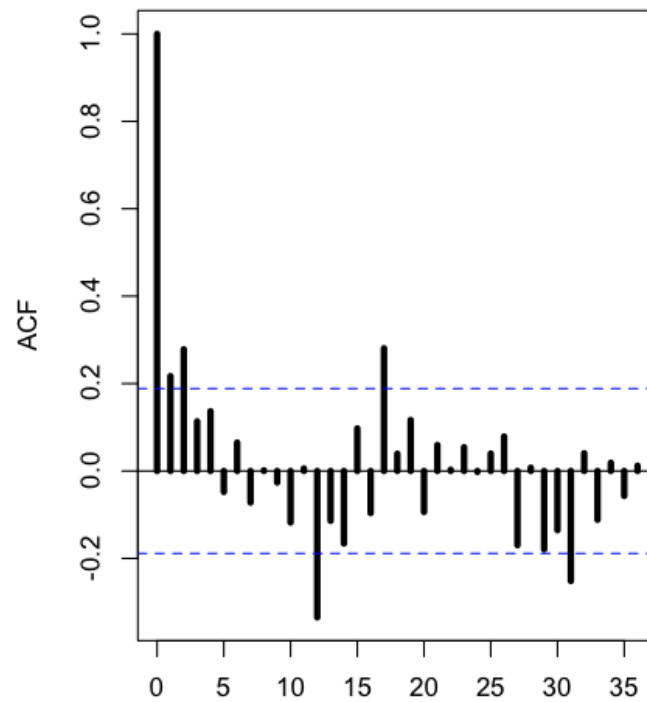


Estimation d'un modèle *SARIMA*

```
> fit.s=arima(Y, order = c(p=0, d=0, q=0),  
+ seasonal = list(order = c(0, 1, 0), period = 12))  
> summary(fit.s)  
Series: Y  
ARIMA(0,0,0)(0,1,0)[12]  
  
sigma^2 estimated as 3089371:  log likelihood=-853.51  
AIC=1709.01   AICc=1709.05   BIC=1711.57  
  
In-sample error measures:  
      ME      RMSE      MAE      MPE      MAPE  
27.28323 1657.13868 1268.25155 65.46664 121.74515
```

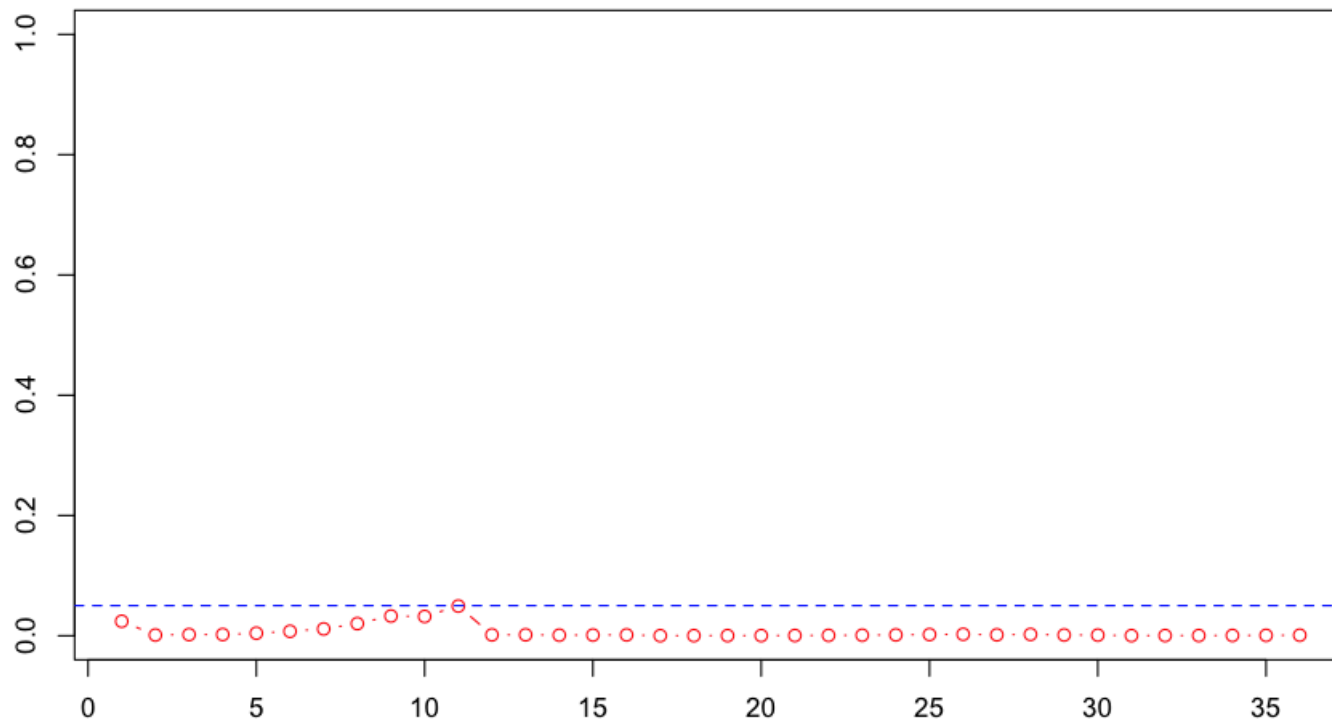
Estimation d'un modèle *SARIMA*

```
> acf(residuals(fit.s))  
> pacf(residuals(fit.s))
```



Estimation d'un modèle *SARIMA*

```
> BP=function(h) Box.test(residuals(fit.s),lag=h, type='Box-Pierce')$p.value  
> plot(1:36,Vectorize(BP)(1:36),type='b',col="red")  
> abline(h=.05,lty=2,col="blue")
```



Estimation d'un modèle *ARIMA*

Envie de davantage d'inspiration ?

```
> auto.arima(Y,max.p=15, max.q=15)
```

```
Series: Y
```

```
ARIMA(2,0,3) with zero mean
```

```
Coefficients:
```

	ar1	ar2	ma1	ma2	ma3
	0.9498	-0.9059	-0.2812	0.7111	0.3391
s.e.	0.0452	0.0444	0.0922	0.0630	0.0924

```
sigma^2 estimated as 5330739: log likelihood=-991.73
```

```
AIC=1995.47 AICc=1996.3 BIC=2011.56
```

Estimation d'un modèle *ARIMA*

```
> fit.arma2.3=arima(Y, order = c(p=2, d=0, q=3),include.mean=FALSE)
```

```
>
```

```
> summary(fit.arma2.3)
```

```
Series: Y
```

```
ARIMA(2,0,3) with zero mean
```

```
Coefficients:
```

	ar1	ar2	ma1	ma2	ma3
	0.9498	-0.9059	-0.2812	0.7111	0.3391
s.e.	0.0452	0.0444	0.0922	0.0630	0.0924

```
sigma^2 estimated as 5330739: log likelihood=-991.73
```

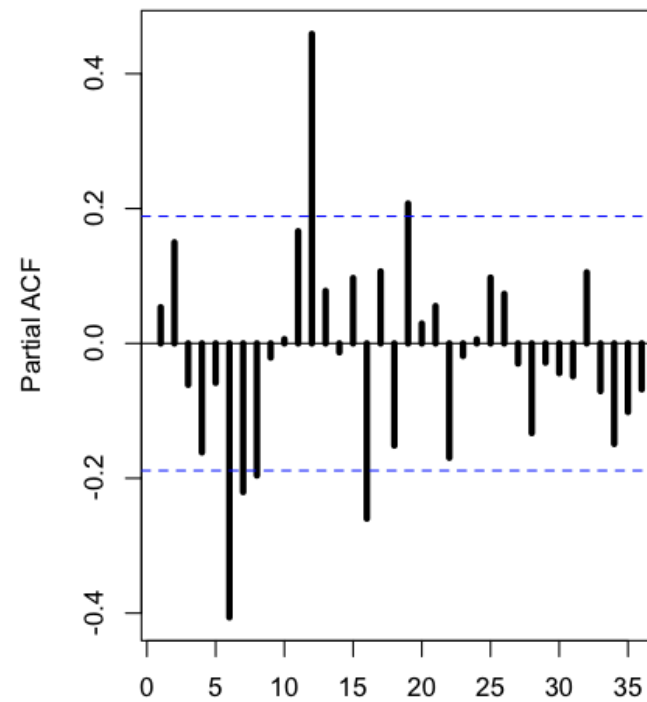
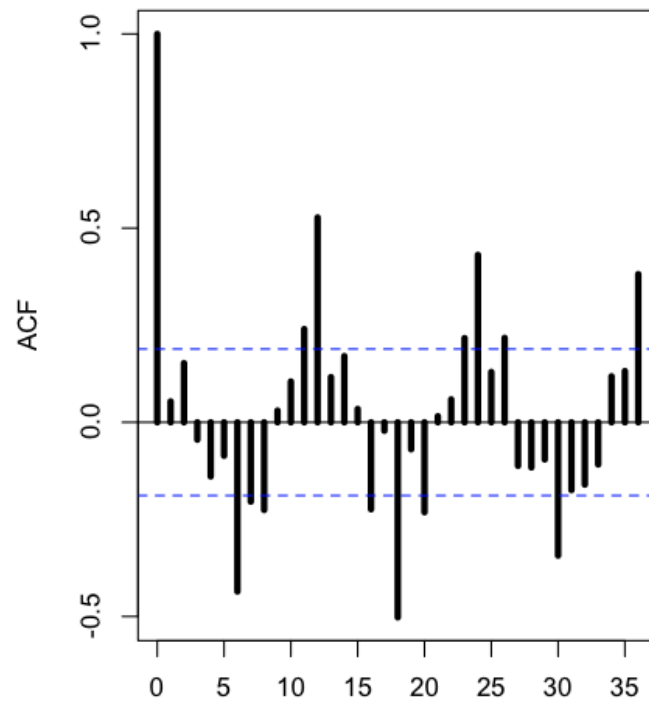
```
AIC=1995.47 AICc=1996.3 BIC=2011.56
```

```
In-sample error measures:
```

	ME	RMSE	MAE	MPE	MAPE
	-25.41080	2308.83936	1811.98016	84.45004	117.63386

Estimation d'un modèle *ARIMA*

```
> acf(residuals(fit.arma2.3))  
> pacf(residuals(fit.arma2.3))
```



Estimation d'un modèle $SARIMA(s, p, d, q)$

Nous avons vu que sous R, un modèle `arima(p,d,q)(ps,ds,qs)[s]` (sans constante) s'écrit

$$(1 - L)^d (1 - L^s)^{ds} \Phi(L) \Phi_s(L^s) X_t = \Theta(L) \Theta_s(L^s)$$

avec $\deg(\Phi) = p$, $\deg(\Theta) = q$, $\deg(\Phi_s) = ps$, et $\deg(\Theta_s) = qs$.

Par exemple `arima(1,0,2)(1,0,0)[4]` est un modèle

$$(1 - \phi L)(1 - \phi_4 L^4) X_t = (1 + \theta_1 L + \theta_2 L^2) \varepsilon_t$$

L'estimation se fait avec

```
> arima(X,order=c(1,0,2),seasonal=list(order=c(1,0,0),period=4))
```

Estimation d'un modèle $SARIMA(s, p, d, q)$

```
> arima(X,order=c(1,0,2),seasonal=list(order=c(1,0,0),period=4))
```

```
Series: X
```

```
ARIMA(1,0,2)(1,0,0)[4] with non-zero mean
```

```
Coefficients:
```

	ar1	ma1	ma2	sar1	intercept
	-0.4485	0.0147	0.5428	0.6901	0.2328
s.e.	0.0894	0.0851	0.0647	0.0552	0.2186

```
sigma^2 estimated as 1.026: log likelihood=-345.03
```

```
AIC=702.06 AICc=702.43 BIC=722.95
```

i.e.

$$(1 + \underset{(0.0894)}{0.4485L})(1 - \underset{(0.0552)}{0.6901L^4})(X_t - \underset{(0.2186)}{0.2328}) = (1 + \underset{(0.0851)}{0.0147L} + \underset{(0.0647)}{0.5428L^2})\varepsilon_t$$

avec $\hat{\sigma}^2 = 1.026$, ou encore

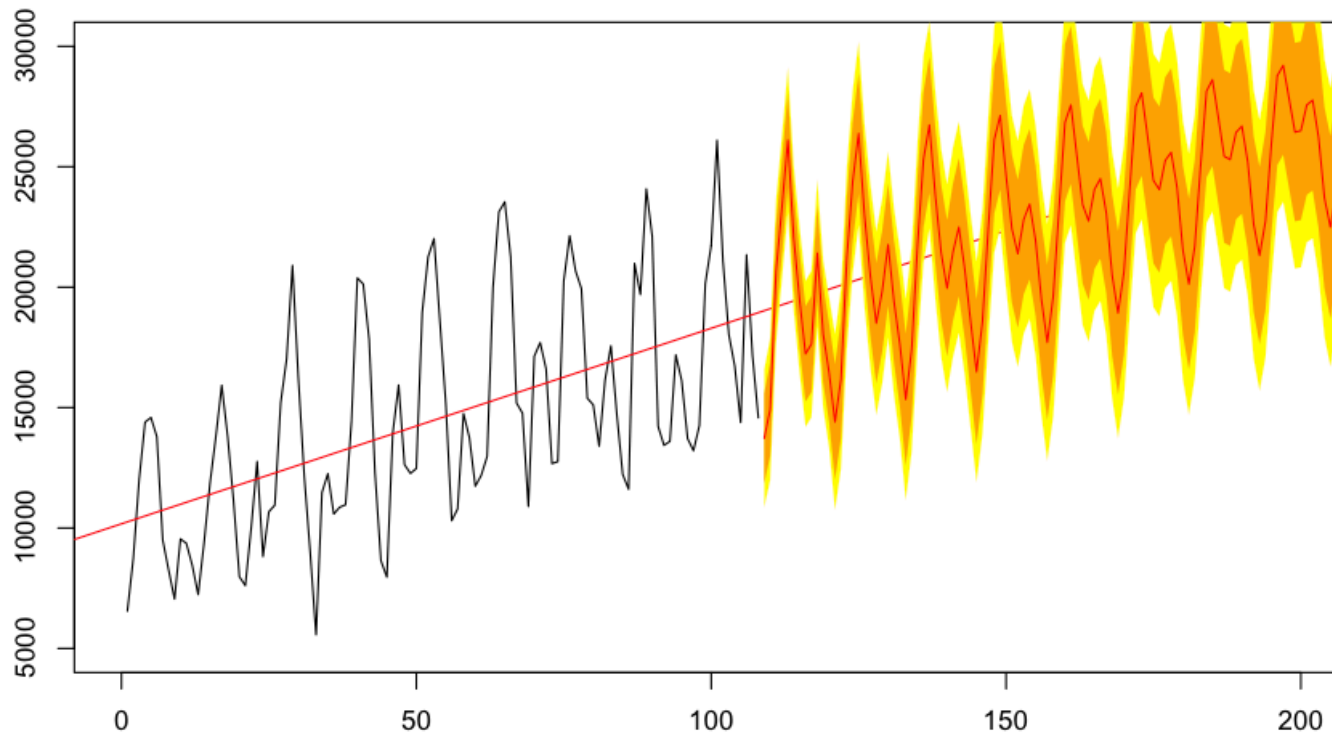
$$(1 + 0.4L)(1 - 0.7L^4)X_t = (1 + 0.5L^2)\varepsilon_t \text{ avec } \text{Var}(\varepsilon_t) = 1.$$

Prévision à l'aide d'un modèle *ARIMA*

Pour finir, on veut faire de la prévision, i.e. pour $h \geq 1$, construire ${}_T\hat{X}_{T+h}$, prévision de X_{t+h} , i.e.

$${}_T\hat{X}_{T+h} = \mathbb{E}(X_{t+h} | X_T, X_{T-1}, \dots, X_1)$$

et calculer $\text{Var}({}_T\hat{X}_{T+h} | X_T, X_{T-1}, \dots, X_1)$.



Prévision ${}_T\hat{X}_{T+h}$ de X_{t+h}

On a observé $\{X_1, \dots, X_T\}$ modélisé par un processus $ARIMA(p, d, q)$

- si (X_t) est un processus autorégressif $AR(p)$.

On suppose que (X_t) vérifie

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t \text{ ou } \Phi(L) X_t = \varepsilon_t$$

La prévision optimale pour la date $T + 1$, faite à la date T est

${}_T\hat{X}_{T+1} = EL(X_{T+1} | X_T, X_{T-1}, \dots) = EL(X_{T+1} | \varepsilon_T, \varepsilon_{T-1}, \dots)$ car (ε_t) est le processus d'innovation. Aussi,

$${}_T\hat{X}_{T+1} = \phi_1 X_T + \dots + \phi_p X_{T-p} + 0$$

Prévision ${}_T\hat{X}_{T+h}$ de X_{t+h}

Pour un horizon $h \geq 1$, ${}_T\hat{X}_{T+h} = EL(X_{T+h} | X_T, X_{T-1}, \dots)$ et donc, de façon récursive par

$${}_T\hat{X}_{T+h} = \begin{cases} \phi_{1T}\hat{X}_{T+h-1} + \dots + \phi_{h-1T}\hat{X}_{T+1} + \phi_h X_T + \dots + \phi_p X_{T+h-p} & \text{pour } h \leq p \\ \phi_{1T}\hat{X}_{T+h-1} + \dots + \phi_{pT}\hat{X}_{T+h-p} & \text{pour } h > p \end{cases}$$

Par exemple pour un $AR(1)$ (avec constante)

- ${}_T\hat{X}_{T+1} = \mu + \phi X_T$,
- ${}_T\hat{X}_{T+2} = \mu + \phi {}_T\hat{X}_{T+1} = \mu + \phi [\mu + \phi X_T] = \mu [1 + \phi] + \phi^2 X_T$,
- ${}_T\hat{X}_{T+3} = \mu + \phi {}_T\hat{X}_{T+2} = \mu + \phi [\mu + \phi [\mu + \phi X_T]] = \mu [1 + \phi + \phi^2] + \phi^3 X_T$,
- et récursivement, on peut obtenir ${}_T\hat{X}_{T+h}$ de la forme

$${}_T\hat{X}_{T+h} = \mu + \phi {}_T\hat{X}_{T+h-1} = \mu [1 + \phi + \phi^2 + \dots + \phi^{h-1}] + \phi^h X_T.$$

Prévision ${}_T\hat{X}_{T+h}$ de X_{t+h}

- si (X_t) est un processus autorégressif $MA(q)$.

On suppose que (X_t) vérifie

$$X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q} = \Theta(L)\varepsilon_t.$$

La prévision optimale pour la date $T+1$, faite à la date T est

${}_T\hat{X}_{T+1} = EL(X_{T+1}|X_T, X_{T-1}, \dots) = EL(X_{T+1}|\varepsilon_T, \varepsilon_{T-1}, \dots)$ car (ε_t) est le processus d'innovation. Aussi,

$${}_T\hat{X}_{T+1} = 0 + \theta_1\varepsilon_T + \dots + \theta_q\varepsilon_{T+1-q}$$

Plus généralement,

${}_T\hat{X}_{T+h} = EL(X_{T+h}|X_T, X_{T-1}, \dots) = EL(X_{T+h}|\varepsilon_T, \varepsilon_{T-1}, \dots)$, et donc

$${}_T\hat{X}_{T+h} = \begin{cases} \theta_h\varepsilon_T + \dots + \theta_q\varepsilon_{T+h-q} & \text{pour } h \leq q \\ 0 & \text{pour } h > q. \end{cases} \quad (9)$$

Le soucis est qu'il faut connaître (ε_t) (qui non observé). La stratégie est d'inverser Θ (si c'est possible), i.e.

$$X_t = \Theta(L)\varepsilon_t \quad \rightarrow \quad \Theta(L)^{-1}X_t = \varepsilon_t$$

Si

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \varepsilon_t \text{ et donc } X_{t+h} = \sum_{k=1}^{\infty} a_k X_{t+h-k} + \varepsilon_{t+h} \text{ pour tout } h \geq 0$$

et donc, ${}_T\hat{X}_{T+h}$ peut être écrit de façon itérative

$${}_T\hat{X}_{T+h} = \sum_{k=1}^{h-1} a_k {}_T\hat{X}_{T+h-k} + \sum_{k=h}^{\infty} a_k X_{t+h-k}$$

mais qui fait qui fait intervenir des valeurs non-observées (pour $t \leq 0$). On

suppose alors que l'on peut poser

$${}_T\hat{X}_{T+h} = \sum_{k=1}^{h-1} a_{kT} \hat{X}_{T+h-k} + \sum_{k=h}^{T+h} a_k X_{T+h-k} + \underbrace{\sum_{k=T+h+1}^{\infty} a_k X_{T+h-k}}_{\text{Négligeable (hyp.)}},$$

Erreur de prévision $X_{t+h} - {}_T\hat{X}_{T+h}$

Le plus simple est d'utiliser une écriture $MA(q)$, voire $MA(\infty)$ (si c'est possible)

$$X_{T+h} = \sum_{i=0}^{\infty} \theta_i \varepsilon_{T+h-i} = \sum_{i=0}^{T+h} \theta_i \varepsilon_{T+h-i} + \sum_{i=T+h+1}^{\infty} \theta_i \varepsilon_{T+h-i},$$

et donc

$${}_T\Delta_h = X_{t+h} - {}_T\hat{X}_{T+h} \approx \sum_{i=0}^h \theta_i \varepsilon_{T+h-i}.$$

Sous l'hypothèse de normalité des résidus (ε_t), i.e. ε_t i.i.d., $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, alors

$${}_T\Delta_h = X_{t+h} - {}_T\hat{X}_{T+h} \sim \mathcal{N}\left(0, \sigma^2 \sum_{i=0}^h b_i^2\right),$$

Erreur de prévision $X_{t+h} - {}_T\hat{X}_{T+h}$

... d'où l'intervalle de confiance pour X_{T+h} au niveau $1 - \alpha$

$$\left[{}_T\hat{X}_{T+h} \pm u_{1-\alpha/2} \times \hat{\sigma} \times \sqrt{\sum_{i=0}^h \hat{\theta}_i^2} \right],$$

où les $\hat{\theta}_i$ sont des estimateurs des coefficients de la forme moyenne mobile, et $\hat{\sigma}$ est un estimateur de la variance du résidu.

Prévision ${}_T\hat{X}_{T+h}$ pour un $AR(1)$

Considérons le cas d'un processus $AR(1)$ (avec constante), $X_t = \phi_1 X_{t-1} + \mu + \varepsilon_t$.

La prévision à horizon 1, faite à la date T , s'écrit

$${}_T\hat{X}_{T+1} = \mathbb{E}(X_{T+1} | X_T, X_{T-1}, \dots, X_1) = \phi_1 X_T + \mu,$$

et de façon similaire, à horizon 2,

$${}_T\hat{X}_{T+2} = \phi_1 {}_T\hat{X}_{T+1} + \mu = \phi_1^2 X_T + [\phi_1 + 1] \cdot \mu.$$

De façon plus générale, à horizon $h \geq 1$, par récurrence

$${}_T\hat{X}_{T+h} = \phi_1^h X_T + [\phi_1^{h-1} + \dots + \phi_1 + 1] \cdot \mu. \quad (10)$$

Remarque A long terme ($h \rightarrow \infty$) ${}_T\hat{X}_{T+h}$ converge vers $\delta (1 - \phi_1)^{-1}$ (i.e. $\mathbb{E}(X_t)$).

Erreur de prévision $X_{t+h} - {}_T\hat{X}_{T+h}$ pour un $AR(1)$

$${}_T\Delta_h = {}_T\hat{X}_{T+h} - X_{T+h} = {}_T\hat{X}_{T+h} - [\phi_1 X_{T+h-1} + \mu + \varepsilon_{T+h}]$$

...

$$= {}_T\hat{X}_{T+h} - [\phi_1^h X_T + (\phi_1^{h-1} + \dots + \phi_1 + 1) \mu + \varepsilon_{T+h} + \phi_1 \varepsilon_{T+h-1} + \dots + \phi_1^{h-1} \varepsilon_{T+1}]$$

en substituant (10), on obtient

$${}_T\Delta_h = \varepsilon_{T+h} + \phi_1 \varepsilon_{T+h-1} + \dots + \phi_1^{h-1} \varepsilon_{T+1},$$

qui possède la variance

$$\text{Var}({}_T\Delta_h) = [1 + \phi_1^2 + \phi_1^4 + \dots + \phi_1^{2h-2}] \sigma^2, \text{ où } \text{Var}(\varepsilon_t) = \sigma^2.$$

Remarque $\text{Var}({}_T\Delta_h) \geq \text{Var}({}_T\Delta_{h-1})$

Prévision ${}_T\hat{X}_{T+h}$ pour un $MA(1)$

Considérons le processus stationnaire (X_t) , $MA(1)$ (avec constante)

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

La prévision à **horizon 1**, faite à la date T , s'écrit

$${}_T\hat{X}_{T+1} = \mathbb{E}(X_{T+1} | X_T, X_{T-1}, \dots, X_1) = \mu + \theta_1 \varepsilon_T$$

à **horizon 2**,

$${}_T\hat{X}_{T+2} = \mathbb{E}(X_{T+2} | X_T, X_{T-1}, \dots, X_1) = \mathbb{E}(\mu + \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1}) = \mu$$

et de façon plus générale, à **horizon $h \geq 1$** , par récurrence

$${}_T\hat{X}_{T+h} = \mathbb{E}(X_{T+h} | X_T, X_{T-1}, \dots, X_1) = \mathbb{E}(\mu + \varepsilon_{T+h} + \theta_1 \varepsilon_{T+h-1}) = \mu \quad (11)$$

Erreur de prévision $X_{t+h} - {}_T\hat{X}_{T+h}$ pour un $AR(1)$

L'erreur de prévision à horizon h est

$${}_T\Delta_h = {}_T\hat{X}_{T+h}X_{T+h} = \varepsilon_{T+h} + \theta_1\varepsilon_{T+h-1}$$

dont la variance est

$$\text{Var}({}_T\Delta_h) = \begin{cases} \theta_1^2\sigma^2 & \text{pour } h = 1 \\ (1 + \theta_1^2)\sigma^2 & \text{pour } h \geq 2 \end{cases}$$

Prévision ${}_T\hat{X}_{T+h}$ pour un $ARIMA(1, 1, 0)$

On a ici un modèle $AR(1)$ intégré, i.e. (X_t) est solution de

$$\begin{cases} Y_t = X_t - X_{t-1} \text{ ou } X_t = X_{t-1} + Y_t \\ Y_t = \phi_1 Y_{t-1} + \mu + \varepsilon_t \end{cases}$$

Comme pour tout t , $X_{t+1} = X_t + Y_{t+1}$, en particulier pour $t = T$, et donc

$${}_T\hat{X}_{T+1} = X_T + {}_T\hat{Y}_{T+1}$$

et comme $X_{t+h} = X_t + Y_{t+1} + Y_{t+2} + \dots + Y_{t+h}$, i.e.

$${}_T\hat{X}_{T+h} = X_T + {}_T\hat{Y}_{T+1} + {}_T\hat{Y}_{T+2} + \dots + {}_T\hat{Y}_{T+h}$$

Or (Y_t) suit un processus $AR(1)$, et on sait le prédire. En utilisant les transparents précédents, la prévision à [horizon 1](#), faite à la date T , s'écrit

$${}_T\hat{X}_{T+1} = X_T + {}_T\hat{Y}_{T+1} = X_T + \phi_1 Y_T + \mu = X_T + \phi_1 [X_T - X_{T-1}] + \mu$$

i.e.

$${}_T\widehat{X}_{T+1} = (1 + \phi_1) X_T - \phi_1 X_{T-1} + \mu$$

à horizon 2,

$${}_T\widehat{X}_{T+2} = (1 + \phi_1 + \phi_1^2) X_T - (\phi_1 + \phi_1^2) X_{T-1} + (\phi_1 + 1) \mu$$

et de façon plus générale, à horizon $h \geq 1$, par récurrence

$$\begin{cases} {}_T\widehat{Y}_{T+h} = \phi_1^h Y_T + [\phi_1^{h-1} + \dots + \phi_1 + 1] \mu \\ {}_T\widehat{X}_{T+h} = {}_T\widehat{X}_{T+h-1} + \phi_1 {}_T\widehat{Y}_{T+h-1} + \mu. \end{cases} \quad (12)$$

Erreur de prévision $X_{t+h} - {}_T\hat{X}_{T+h}$ pour un $ARIMA(1, 1, 0)$

L'erreur faite sur la prévision à horizon 1 est

$${}_T\Delta_1 = {}_T\hat{X}_{T+1} - X_{T+1} = {}_T\hat{Y}_{T+1} - Y_{T+1} = \varepsilon_{T+1}$$

i.e. $\text{Var}({}_T\Delta_1) = \sigma^2$.

A horizon 2, l'erreur de prévision est

$${}_T\Delta_2 = {}_T\hat{X}_{T+2} - X_{T+2} = \left({}_T\hat{Y}_{T+1} - Y_{T+1} \right) + \left({}_T\hat{Y}_{T+2} - Y_{T+2} \right)$$

i.e.

$${}_T\Delta_2 = (1 + \theta_1) \varepsilon_{T+1} + \varepsilon_{T+2},$$

et $\text{Var}({}_T\Delta_2) = \left[1 + (1 + \phi_1)^2 \right] \sigma^2$.

De façon plus générale, l'erreur de prévision à horizon h est

$$\begin{aligned}
 {}_T\Delta_h &= \left({}_T\widehat{Y}_{T+1} - Y_{T+1} \right) + \left({}_T\widehat{Y}_{T+2} - Y_{T+2} \right) + \left({}_T\widehat{Y}_{T+3} - Y_{T+3} \right) + \cdots + \left({}_T\widehat{Y}_{T+h} - Y_{T+h} \right) \\
 &= \varepsilon_{T+1} + (\varepsilon_{T+2} + \phi_1\varepsilon_{T+1}) + \dots + (\varepsilon_{T+h} + \phi_1\varepsilon_{T+h-1} + \dots + \phi_1^{h-2}\varepsilon_{T+2} + \phi_1^{h-1}\varepsilon_{T+1}) \\
 &= \varepsilon_{T+h} + (1 + \phi_1)\varepsilon_{T+h-1} + \dots + (1 + \phi_1 + \dots + \phi_1^{h-1})\varepsilon_{T+1},
 \end{aligned}$$

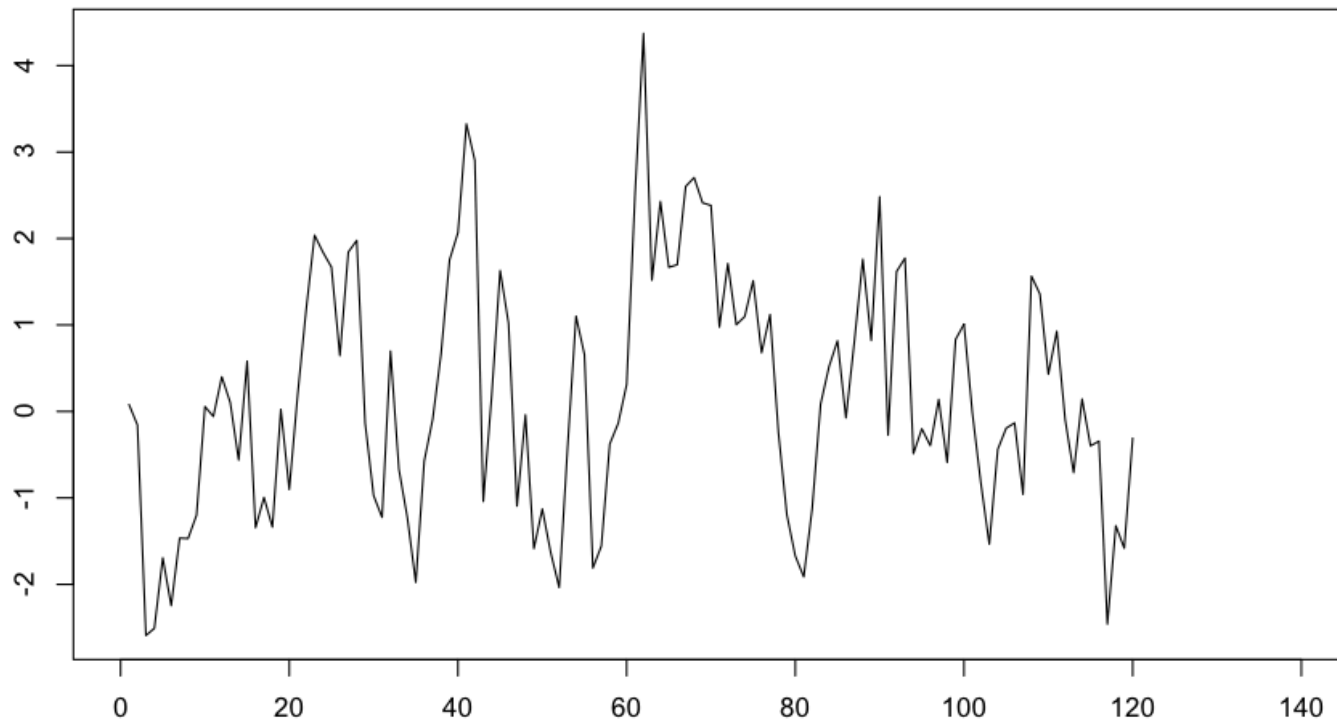
avec

$$\text{Var}({}_T\Delta_h) = \left[\sum_{i=1}^h \left(\sum_{j=0}^{i-1} \phi_1^j \right)^2 \right] \sigma^2.$$

Remarque L'erreur de prévision sur X_{T+h} est alors l'accumulation des erreurs de prévision de Y_{T+1}, \dots, Y_{T+h} .

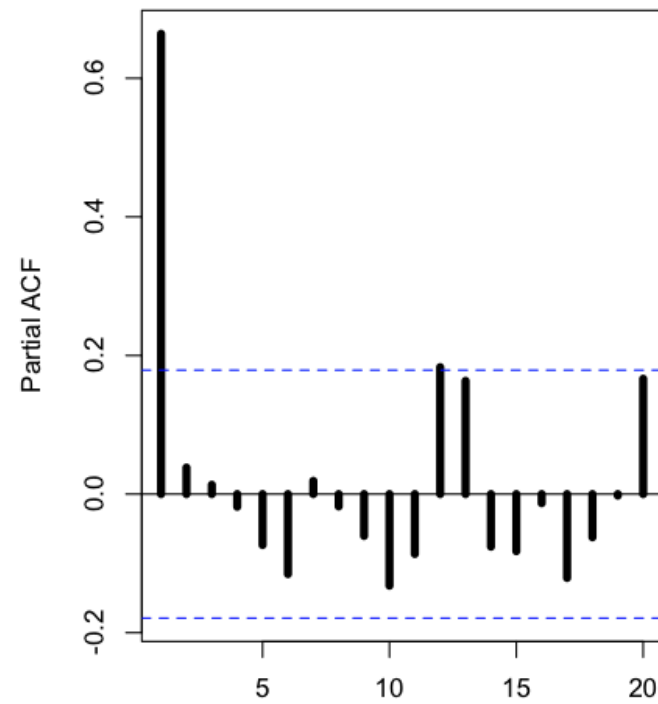
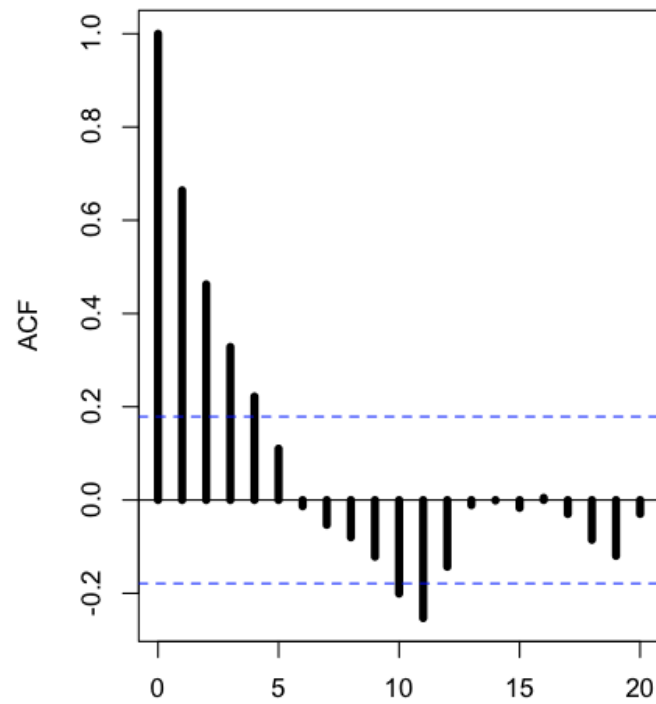
Aspects computationnels, modèle $AR(1)$

```
> X=arima.sim(list(ar=0.8),n=120)  
> plot(X)
```



Aspects computationnels, modèle $AR(1)$

- > `acf(X)`
- > `pacf(X)`



Aspects computationnels, modèle $AR(1)$

```
> model=arima(X,order=c(1,0,0),include.mean=FALSE)
```

```
> summary(model)
```

```
Series: X
```

```
ARIMA(1,0,0) with zero mean
```

```
Coefficients:
```

```
ar1
```

```
0.6644
```

```
s.e. 0.0671
```

```
sigma^2 estimated as 1.094: log likelihood=-175.97
```

```
AIC=355.95 AICc=356.05 BIC=361.52
```

```
In-sample error measures:
```

ME	RMSE	MAE	MPE	MAPE	MASE
0.05603047	1.04612565	0.84647681	143.66955140	207.85668760	0.93652943

Aspects computationnels, modèle $AR(1)$

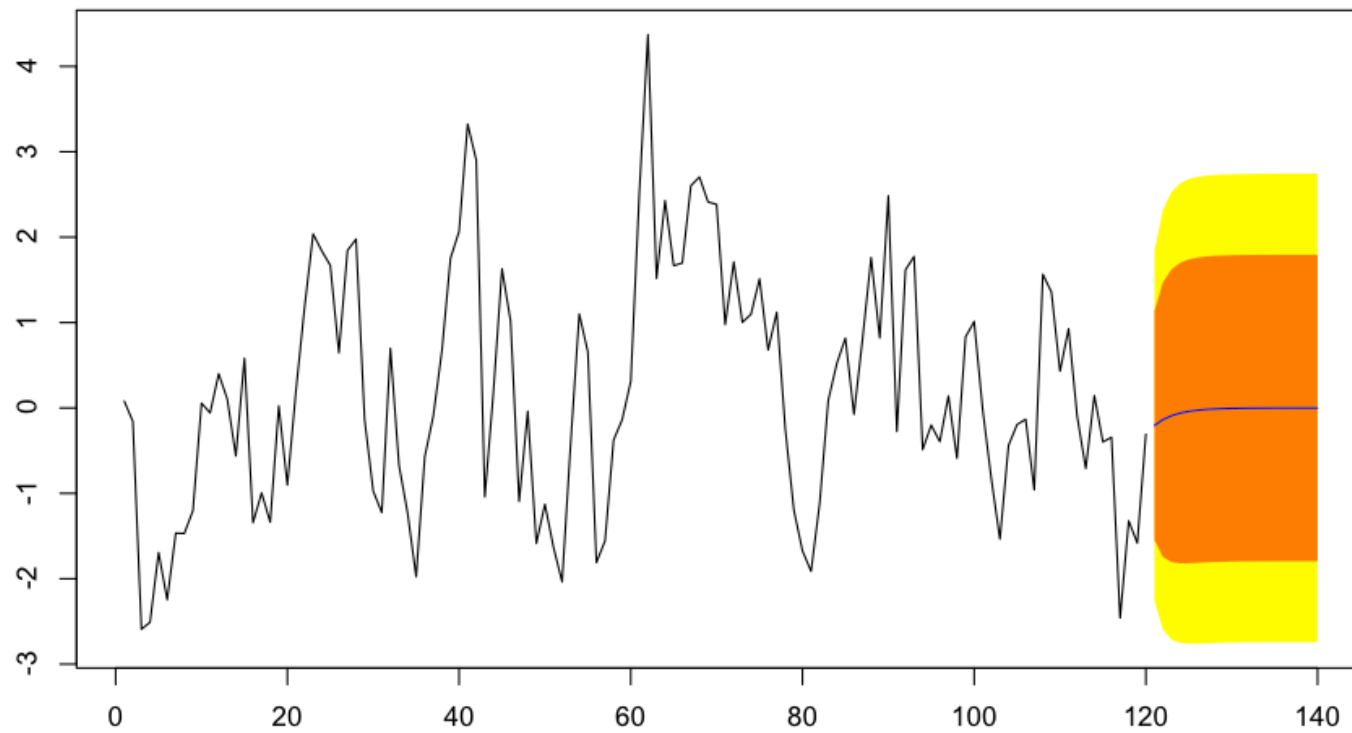
```
> Xp=forecast(model,20)
```

```
> Xp
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
121	-2.049995e-01	-1.545663	1.135665	-2.255368	1.845369
122	-1.361928e-01	-1.745755	1.473369	-2.597805	2.325420
123	-9.048062e-02	-1.805366	1.624405	-2.713172	2.532211
124	-6.011142e-02	-1.819479	1.699256	-2.750832	2.630610
125	-3.993544e-02	-1.818583	1.738712	-2.760142	2.680271
126	-2.653138e-02	-1.813622	1.760559	-2.759650	2.706587
127	-1.762631e-02	-1.808431	1.773178	-2.756425	2.721172
128	-1.171016e-02	-1.804151	1.780731	-2.753012	2.729592
129	-7.779723e-03	-1.800943	1.785383	-2.750186	2.734626
130	-5.168513e-03	-1.798650	1.788313	-2.748062	2.737725
131	-3.433737e-03	-1.797056	1.790188	-2.746542	2.739674
132	-2.281227e-03	-1.795965	1.791403	-2.745484	2.740922
133	-1.515549e-03	-1.795227	1.792196	-2.744761	2.741729
134	-1.006865e-03	-1.794730	1.792717	-2.744270	2.742257
135	-6.689177e-04	-1.794398	1.793060	-2.743941	2.742603
136	-4.444001e-04	-1.794176	1.793287	-2.743720	2.742831

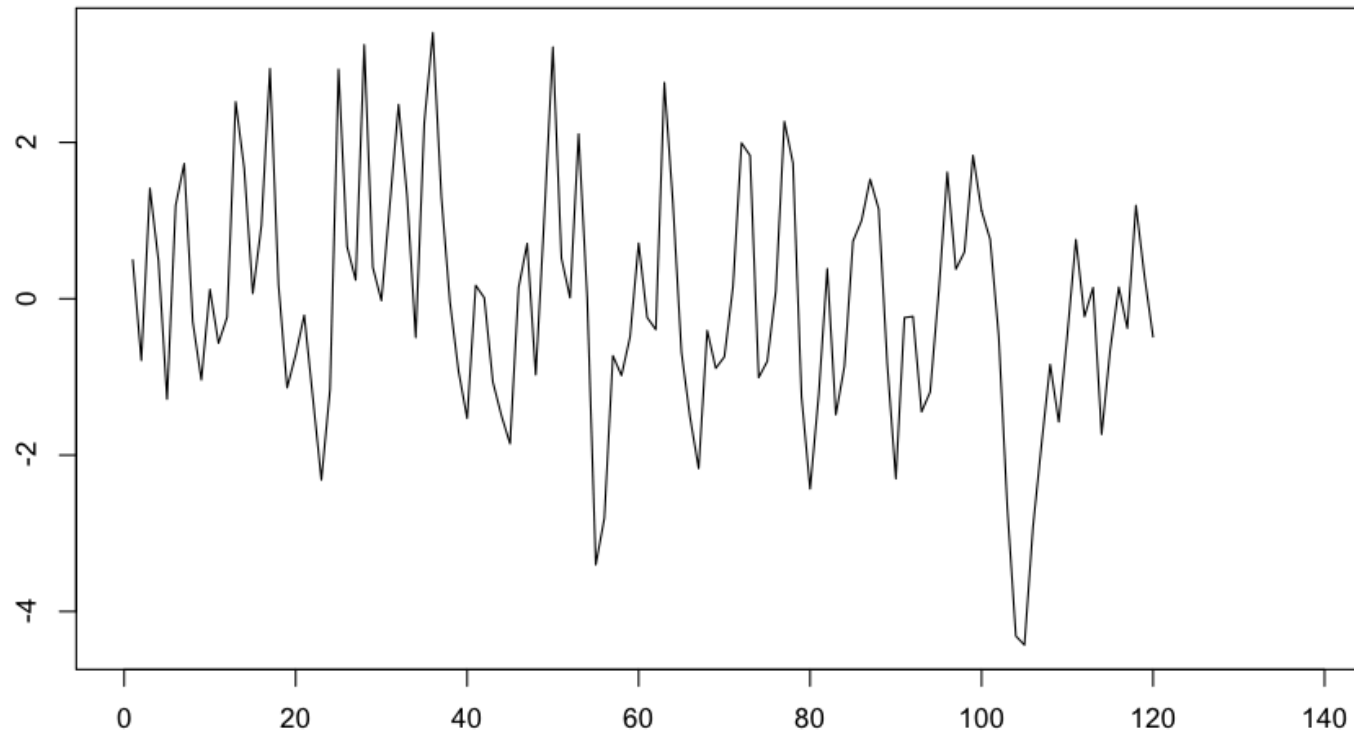
```
137 -2.952403e-04 -1.794028 1.793437 -2.743572 2.742982
138 -1.961449e-04 -1.793929 1.793537 -2.743474 2.743081
139 -1.303102e-04 -1.793863 1.793603 -2.743408 2.743148
140 -8.657248e-05 -1.793820 1.793646 -2.743365 2.743191
> plot(Xp)
```

Aspects computationnels, modèle $AR(1)$



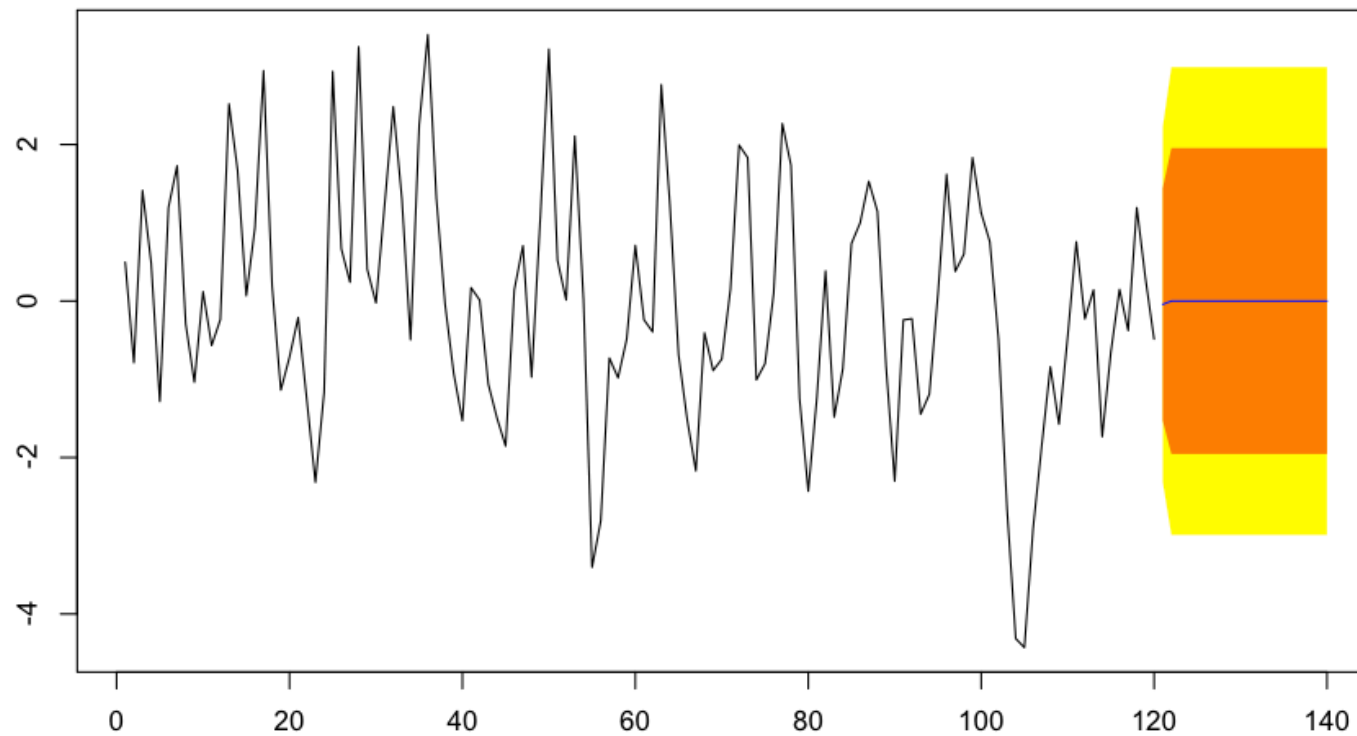
Aspects computationnels, modèle $MA(1)$

```
> X=arima.sim(list(ma=0.8),n=120)  
> plot(X,xlim=c(0,140),ylab="")
```



Aspects computationnels, modèle $MA(1)$

```
> acf(X)  
> pacf(X)
```



Aspects computationnels, modèle $MA(1)$

```
> model=arima(X,order=c(0,0,1),include.mean=FALSE)
```

```
> summary(model)
```

```
Series: X
```

```
ARIMA(0,0,1) with zero mean
```

```
Coefficients:
```

```
    ma1
```

```
    0.8593
```

```
s.e. 0.0466
```

```
sigma^2 estimated as 1.338:  log likelihood=-188.44
```

```
AIC=380.87  AICc=380.97  BIC=386.45
```

```
In-sample error measures:
```

ME	RMSE	MAE	MPE	MAPE	MASE
-0.02511947	1.15692835	0.92565202	-19.30389457	289.18624726	0.71927487

Aspects computationnels, modèle $MA(1)$

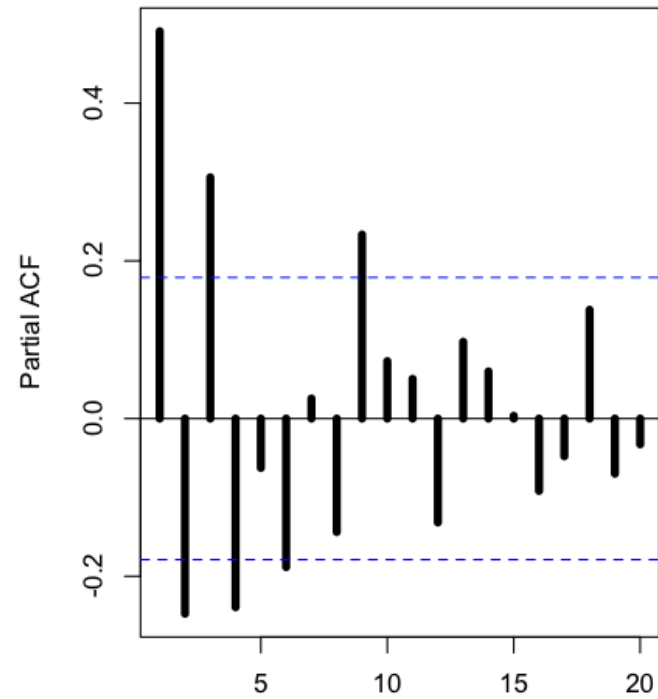
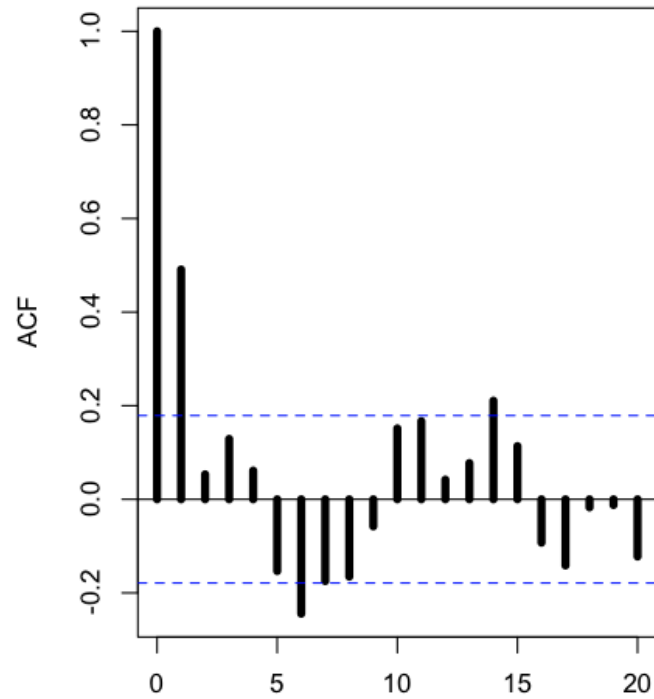
```
> Xp=forecast(model,20)
```

```
> Xp
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
121	-0.04321057	-1.525874	1.439453	-2.310748	2.224327
122	0.00000000	-1.954858	1.954858	-2.989698	2.989698
123	0.00000000	-1.954858	1.954858	-2.989698	2.989698
124	0.00000000	-1.954858	1.954858	-2.989698	2.989698
125	0.00000000	-1.954858	1.954858	-2.989698	2.989698
126	0.00000000	-1.954858	1.954858	-2.989698	2.989698
127	0.00000000	-1.954858	1.954858	-2.989698	2.989698
128	0.00000000	-1.954858	1.954858	-2.989698	2.989698
129	0.00000000	-1.954858	1.954858	-2.989698	2.989698
130	0.00000000	-1.954858	1.954858	-2.989698	2.989698
131	0.00000000	-1.954858	1.954858	-2.989698	2.989698
132	0.00000000	-1.954858	1.954858	-2.989698	2.989698
133	0.00000000	-1.954858	1.954858	-2.989698	2.989698
134	0.00000000	-1.954858	1.954858	-2.989698	2.989698
135	0.00000000	-1.954858	1.954858	-2.989698	2.989698
136	0.00000000	-1.954858	1.954858	-2.989698	2.989698

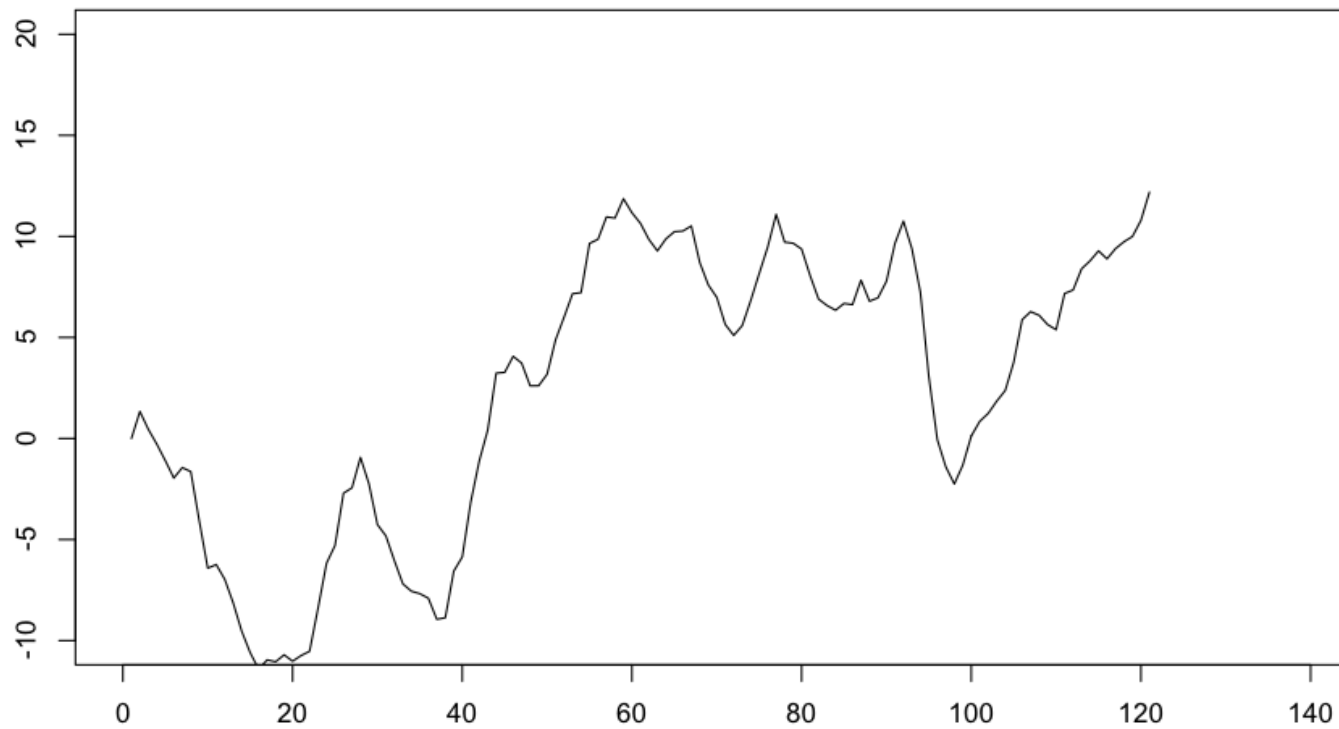
```
137      0.00000000 -1.954858 1.954858 -2.989698 2.989698
138      0.00000000 -1.954858 1.954858 -2.989698 2.989698
139      0.00000000 -1.954858 1.954858 -2.989698 2.989698
140      0.00000000 -1.954858 1.954858 -2.989698 2.989698
> plot(Xp)
```

Aspects computationnels, modèle $MA(1)$



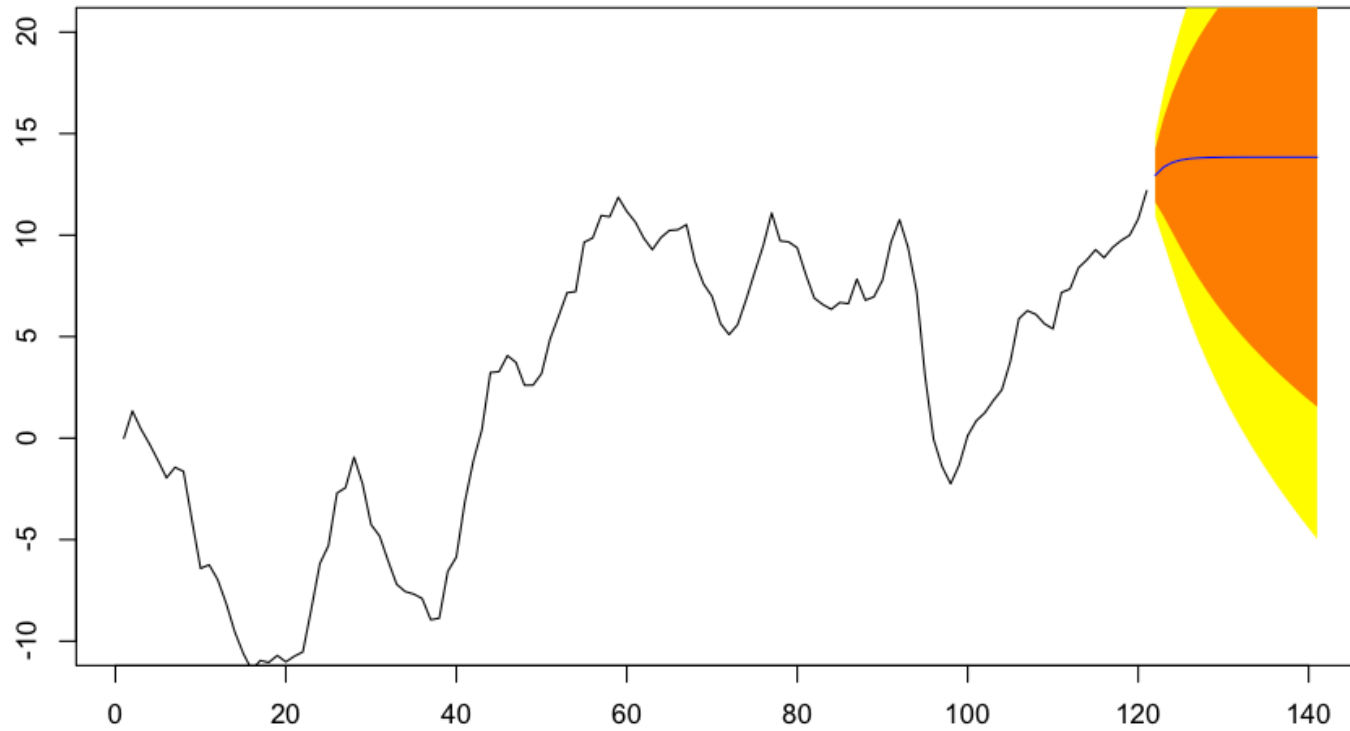
Aspects computationnels, modèle $ARIMA(1, 1, 0)$

```
> X=arima.sim(list(order = c(1,1,0), ar = 0.6),n=120)  
> plot(X)
```



Aspects computationnels, modèle $ARIMA(1, 1, 0)$

- > `acf(X)`
- > `pacf(X)`



Aspects computationnels, modèle *ARIMA*(1, 1, 0)

```
> model=arima(X,order=c(1,1,0),include.mean=FALSE)
```

```
> summary(model)
```

```
Series: X
```

```
ARIMA(1,1,0)
```

```
Coefficients:
```

```
ar1
```

```
0.5429
```

```
s.e. 0.0769
```

```
sigma^2 estimated as 1.067: log likelihood=-174.33
```

```
AIC=352.67 AICc=352.77 BIC=358.24
```

```
In-sample error measures:
```

```
ME
```

```
RMSE
```

```
MAE
```

```
MPE
```

```
MAPE
```

```
MASE
```

```
0.05052021
```

```
1.02864633
```

```
0.80579410
```

```
14.91841852
```

```
39.94896823
```

```
0.83914131
```

Aspects computationnels, modèle $ARIMA(1, 1, 0)$

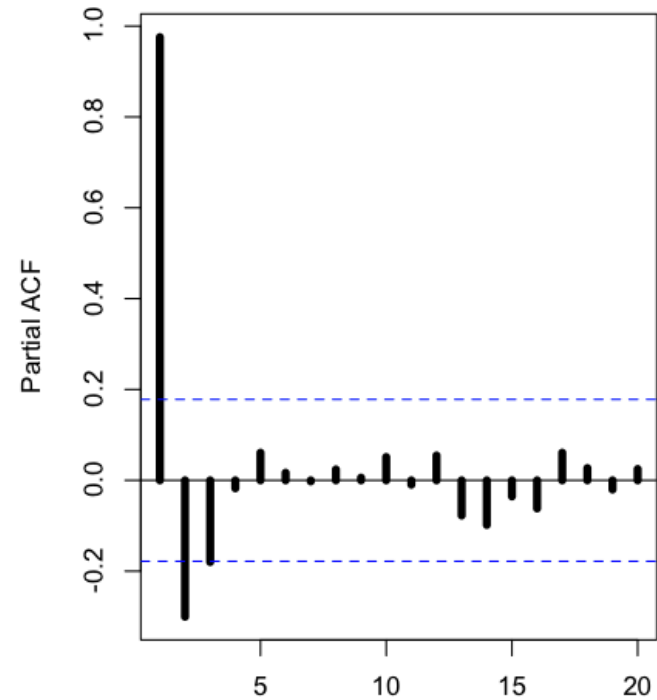
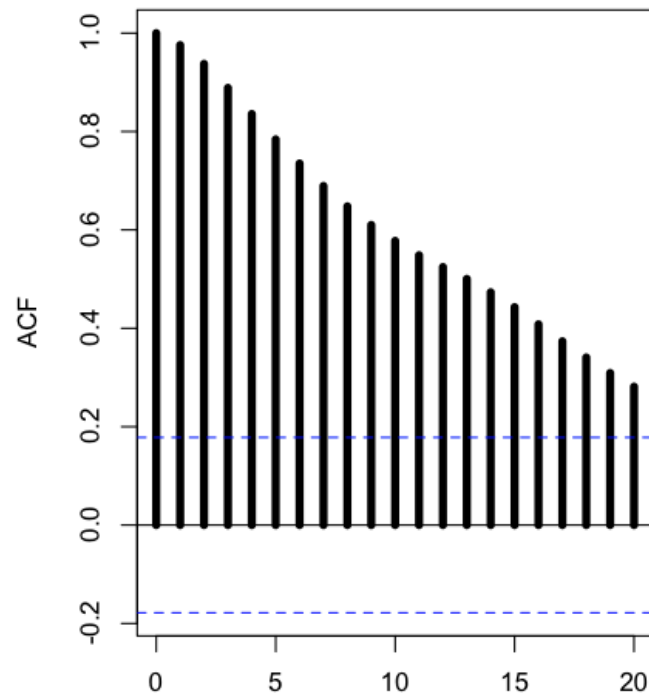
```
> Xp=forecast(model,20)
```

```
> Xp
```

	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
122	12.94419	11.620443	14.26793	10.9196947	14.96868	
123	13.35497	10.921053	15.78888	9.6326174	17.07731	
124	13.57799	10.136796	17.01919	8.3151374	18.84085	
125	13.69908	9.359105	18.03906	7.0616595	20.33651	
126	13.76483	8.621811	18.90784	5.8992633	21.63039	
127	13.80052	7.934143	19.66690	4.8286695	22.77237	
128	13.81990	7.295278	20.34453	3.8413499	23.79845	
129	13.83042	6.700730	20.96012	2.9264973	24.73435	
130	13.83614	6.145108	21.52717	2.0737234	25.59855	
131	13.83924	5.623240	22.05524	1.2739524	26.40453	
132	13.84092	5.130562	22.55129	0.5195745	27.16227	
133	13.84184	4.663199	23.02048	-0.1956796	27.87936	
134	13.84233	4.217914	23.46675	-0.8769461	28.56161	
135	13.84260	3.792020	23.89319	-1.5284378	29.21365	
136	13.84275	3.383281	24.30222	-2.1536274	29.83913	
137	13.84283	2.989832	24.69583	-2.7553978	30.44106	

138	13.84287	2.610106	25.07564	-3.3361619	31.02191
139	13.84290	2.242776	25.44302	-3.8979569	31.58375
140	13.84291	1.886711	25.79911	-4.4425171	32.12833
141	13.84292	1.540941	26.14489	-4.9713316	32.65716

Aspects computationnels, modèle $ARIMA(1, 1, 0)$



Sélection de modèle et backtesting

On dispose d'observations $\{X_1, X_2, \dots, X_T\}$.

- on estime un modèle sur $\{X_1, X_2, \dots, X_{T-k}\}$
- on calcule ${}_{T-k}\hat{X}_{T-k+1}$, ${}_{T-k}\hat{X}_{T-k+2}$, \dots , ${}_{T-k}\hat{X}_{T-k+k}$
- on calcule une erreur de prévision, e.g.

$$\sum_{h=1}^k \left({}_{T-k}\hat{X}_{T-k+h} - X_{T-k+h} \right)^2$$

Sélection de modèle et backtesting

```
> T=length(Y)
> backtest=12
> subY=Y[-((T-backtest+1):T)]
> subtemps=1:(T-backtest)
> fit.ar12.s=arima(subY,order=c(p=12,d=0,q=0),method="CSS")
> fit.arma12.1.s=arima(subY,order=c(p=12,d=0,q=1),method="CSS")
> fit.ar14.s=arima(subY,order=c(p=14,d=0,q=0),method="CSS")
> p.ar12=predict(fit.ar12.s,12)
> pred.ar12=as.numeric(p.ar12$pred)
> p.arma12.1=predict(fit.arma12.1.s,12)
> pred.arma12.1=as.numeric(p.arma12.1$pred)
> p.ar14=predict(fit.ar14.s,12)
> pred.ar14=as.numeric(p.ar14$pred)
```

Sélection de modèle et backtesting

```
> (M=cbind(observ=Y[((T-backtest+1):T]),  
+ modle1=pred.ar12,  
+ modle2=pred.arma12.1,  
+ modle3=pred.ar14))  
      observ      modle1      modle2      modle3  
97  -4836.2174 -5689.3331 -5885.4486 -6364.2471  
98  -3876.4199 -4274.0391 -4287.2193 -4773.8116  
99   1930.3776  1817.8411  2127.9915  2290.1460  
100  3435.1751  4089.3598  3736.1110  4039.4150  
101  7727.9726  6998.9829  7391.6694  7281.4797  
102  2631.7701  3456.8819  3397.5478  4230.5324  
103  -509.4324 -2128.6315 -2268.9672 -2258.7216  
104 -1892.6349 -3877.7609 -3694.9409 -3620.4798  
105 -4310.8374 -3384.0905 -3430.4090 -2881.4942  
106  2564.9600  -504.6883  -242.5018   183.2891  
107 -1678.2425 -1540.9904 -1607.5996  -855.7677  
108 -4362.4450 -3927.4772 -3928.0626 -3718.3922
```

Sélection de modèle et backtesting

```
> sum((M[,1]-M[,2])^2)
[1] 19590931
> sum((M[,1]-M[,3])^2)
[1] 17293716
> sum((M[,1]-M[,4])^2)
[1] 21242230
```