

Modèles de prévision

Partie 2 - séries temporelles

Arthur Charpentier

charpentier.arthur@uqam.ca

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



AUTOMNE 2012

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*, ${}_T\hat{X}_{T+h}$

Méthode de Buys Balot - la tendance

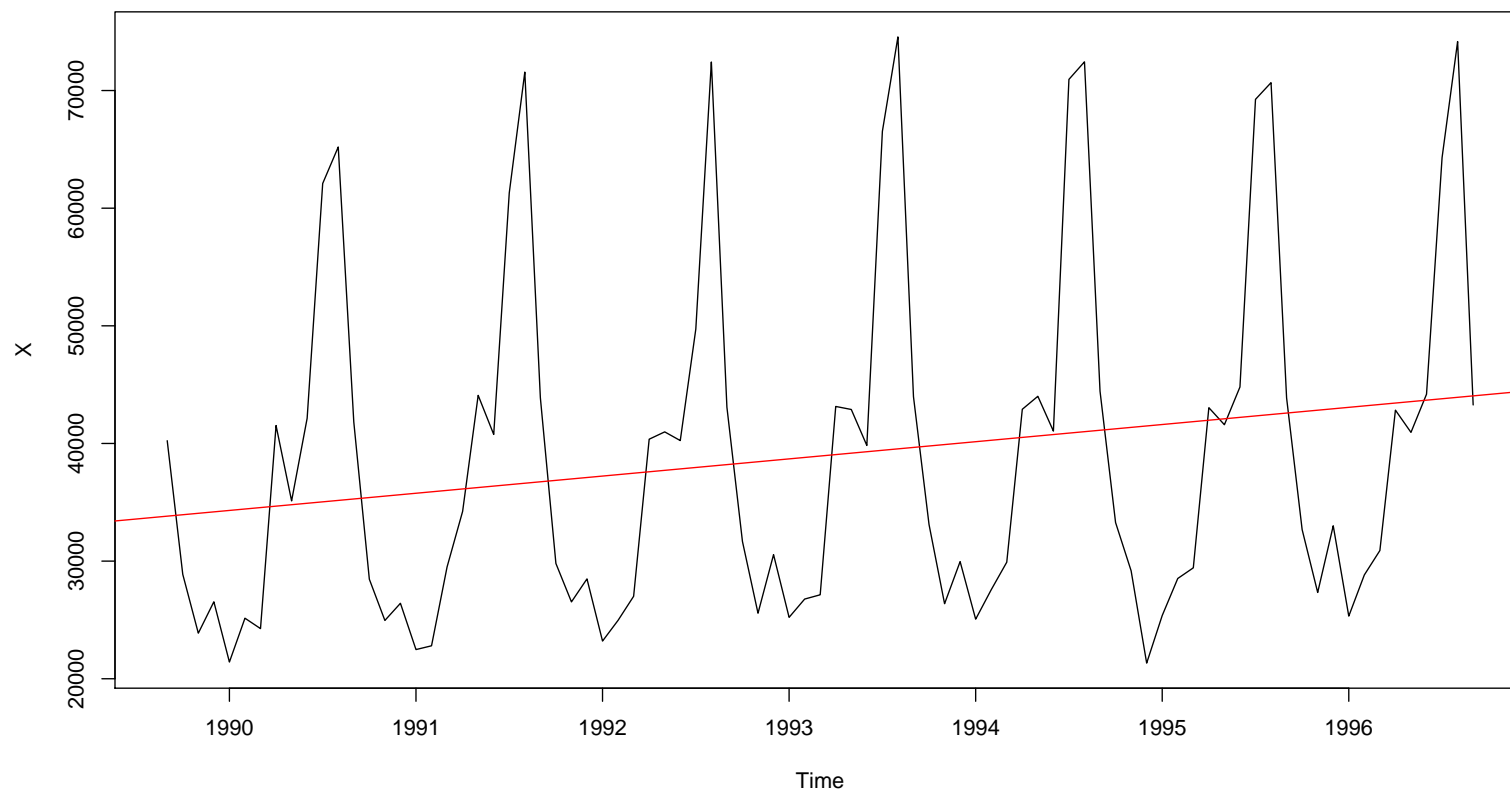
Notons (X_t) une série temporelle, observée jusqu'à la date T .

Si on a une tendance linéaire, on cherche à résoudre

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1) \in \operatorname{argmin} \left\{ \sum_{t=1}^T (X_t - (\beta_0 + \beta_1 \cdot t))^2 \right\}$$

```
> autoroute=read.table(  
+ "http://freakonometrics.blog.free.fr/public/data/autoroute.csv",  
+ header=TRUE,sep=";")  
> a7=autoroute$a007  
> X=ts(a7,start = c(1989, 9), frequency = 12)  
> T=time(X)  
> S=cycle(X)  
> B=data.frame(x=as.vector(X),T=as.vector(T),S=as.vector(S))  
> regT=lm(x~T,data=B)  
> plot(X)  
> abline(regT,col="red")
```

Méthode de Buys Balot - la tendance

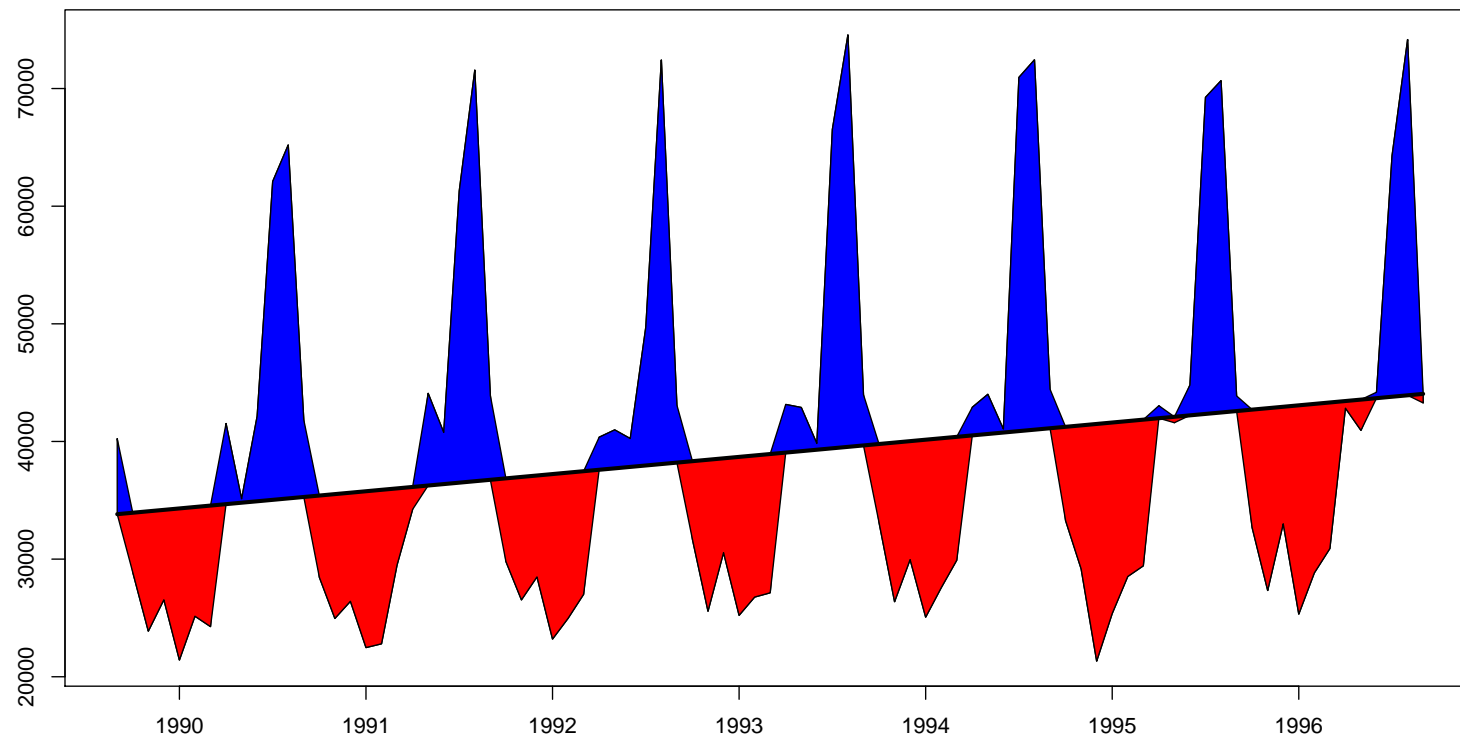


Méthode de Buys Balot - la tendance

Posons $\hat{Y}_t = X_t - (\hat{\beta}_0 + \hat{\beta}_1 t)$

```
> X1=predict(regT)
> B$X1=X1
> Y=B$X1
> plot(X,xlab="",ylab="")
> YU=apply(cbind(X,Y),1,max)
> YL=apply(cbind(X,Y),1,min)
> i=which(is.na(Y)==FALSE)
> polygon(c(T[i],rev(T[i])),c(Y[i],rev(YU[i])),col="blue")
> polygon(c(T[i],rev(T[i])),c(Y[i],rev(YL[i])),col="red")
> lines(B$T,Y,lwd=3)
```

Méthode de Buys Balot - la tendance



Méthode de Buys Balot - le cycle

On cherche un cycle **mensuel** (de période $s = 12$). Supposons

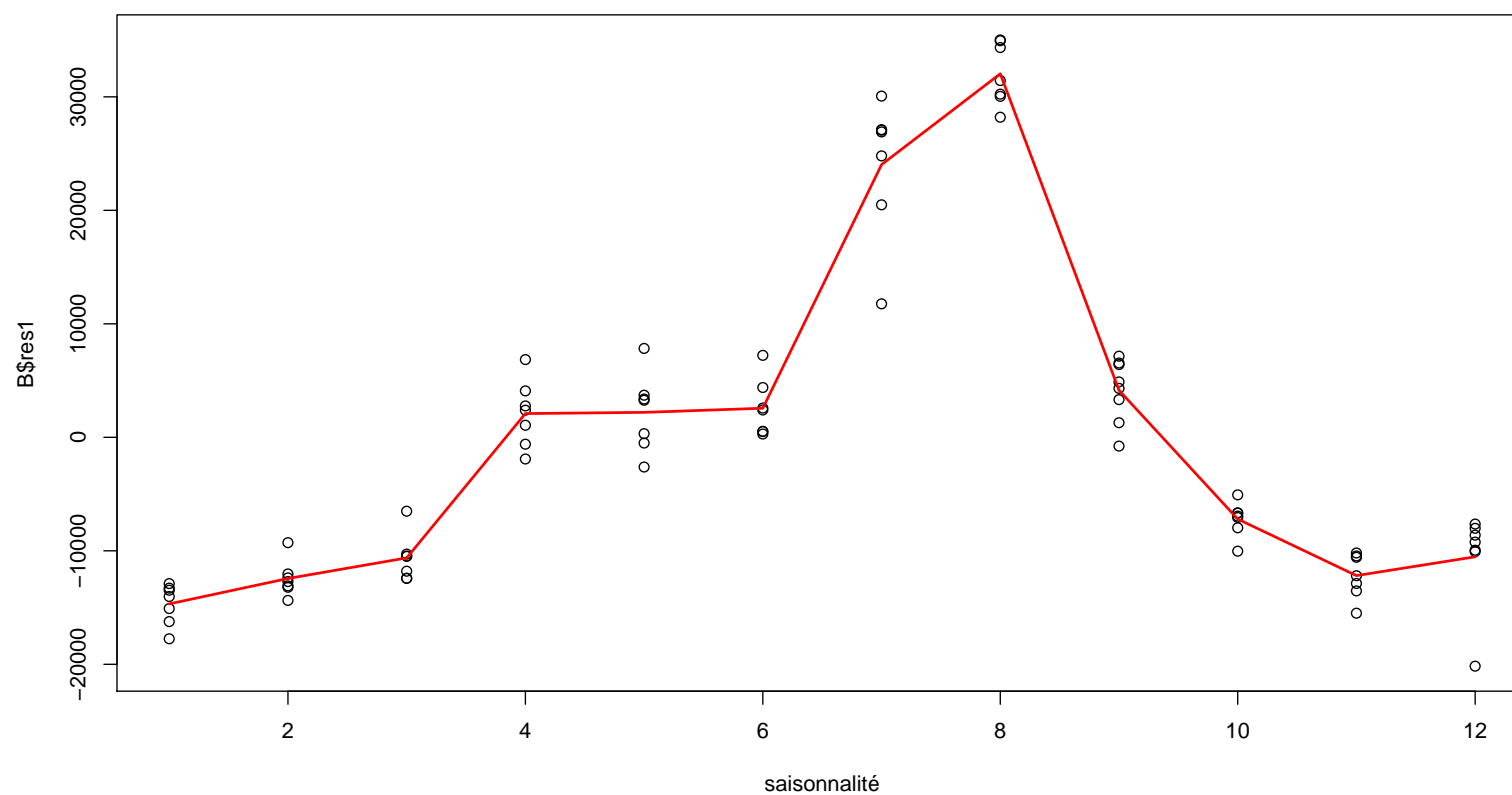
$$\hat{Y}_t = \sum_{k=0}^{s-1} \gamma_k \mathbf{1}(t = k \bmod. s) + Z_t$$

```
> B$res1=X-X1  
> regS=lm(res1~0+as.factor(S),data=B)  
> B$X2=predict(regS)  
> plot(B$S,B$res1,xlab="saisonnalit\'e")  
> lines(B$S[1:4],B$X2[1:4],col="red",lwd=2)  
> lines(B$S[5:13],B$X2[5:13],col="red",lwd=2)
```

Remarque mod. est l'opérateur de calcul du reste de la division euclidienne, ex :

$$27 \bmod. 12 = 27 - (12 \times 2) = 3$$

Méthode de Buys Balot - le cycle



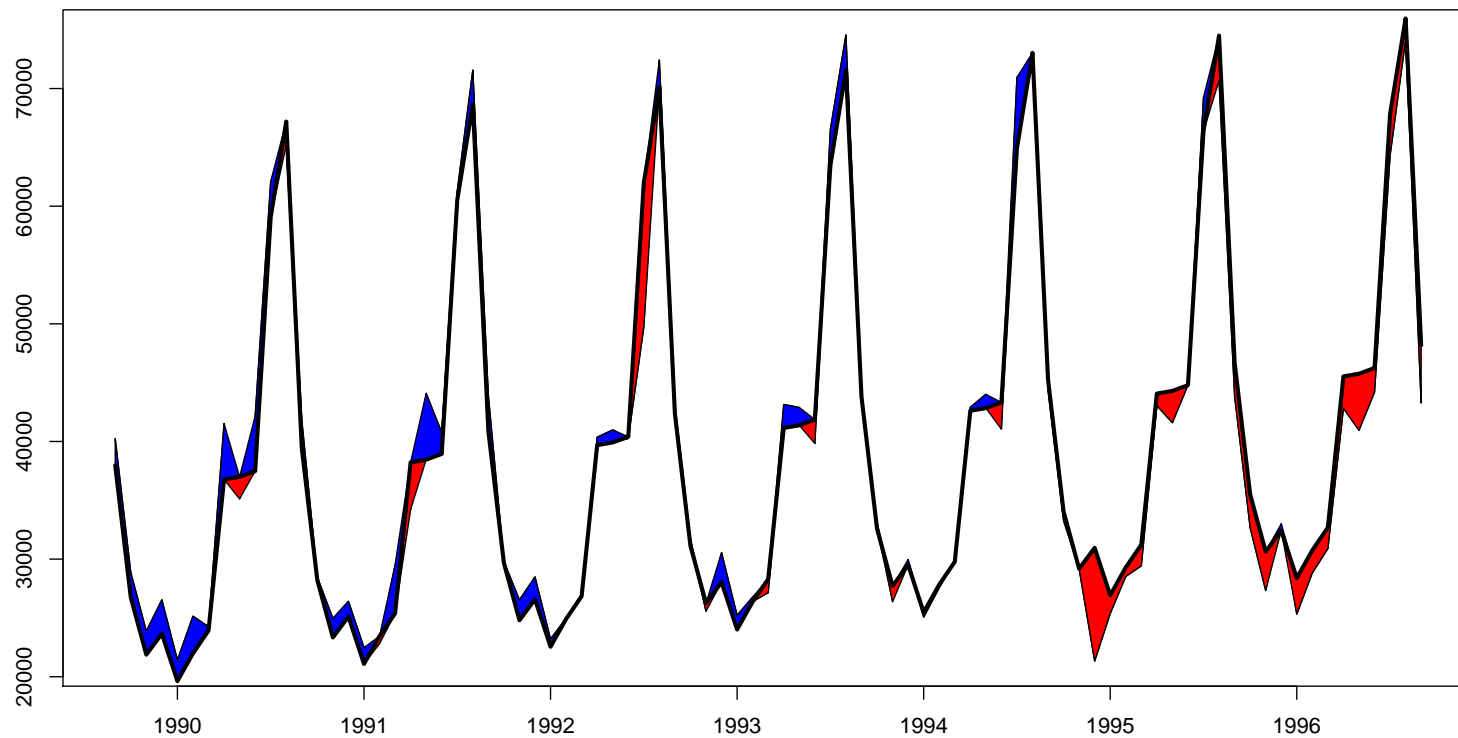
Méthode de Buys Balot - le cycle

Alors

$$X_t = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 t}_{\text{tendance linéaire}} + \underbrace{\sum_{k=0}^{s-1} \hat{\gamma}_k \mathbf{1}(t = k \bmod s)}_{\text{cycle}} + Z_t$$

```
> plot(X,xlab="",ylab="")
> YU=apply(cbind(X,Y),1,max)
> YL=apply(cbind(X,Y),1,min)
> i=which(is.na(Y)==FALSE)
> polygon(c(T[i],rev(T[i])),c(Y[i],rev(YU[i])),col="blue")
> polygon(c(T[i],rev(T[i])),c(Y[i],rev(YL[i])),col="red")
> lines(B$T,Y,lwd=3)
```

Méthode de Buys Balot - tendance et cycle



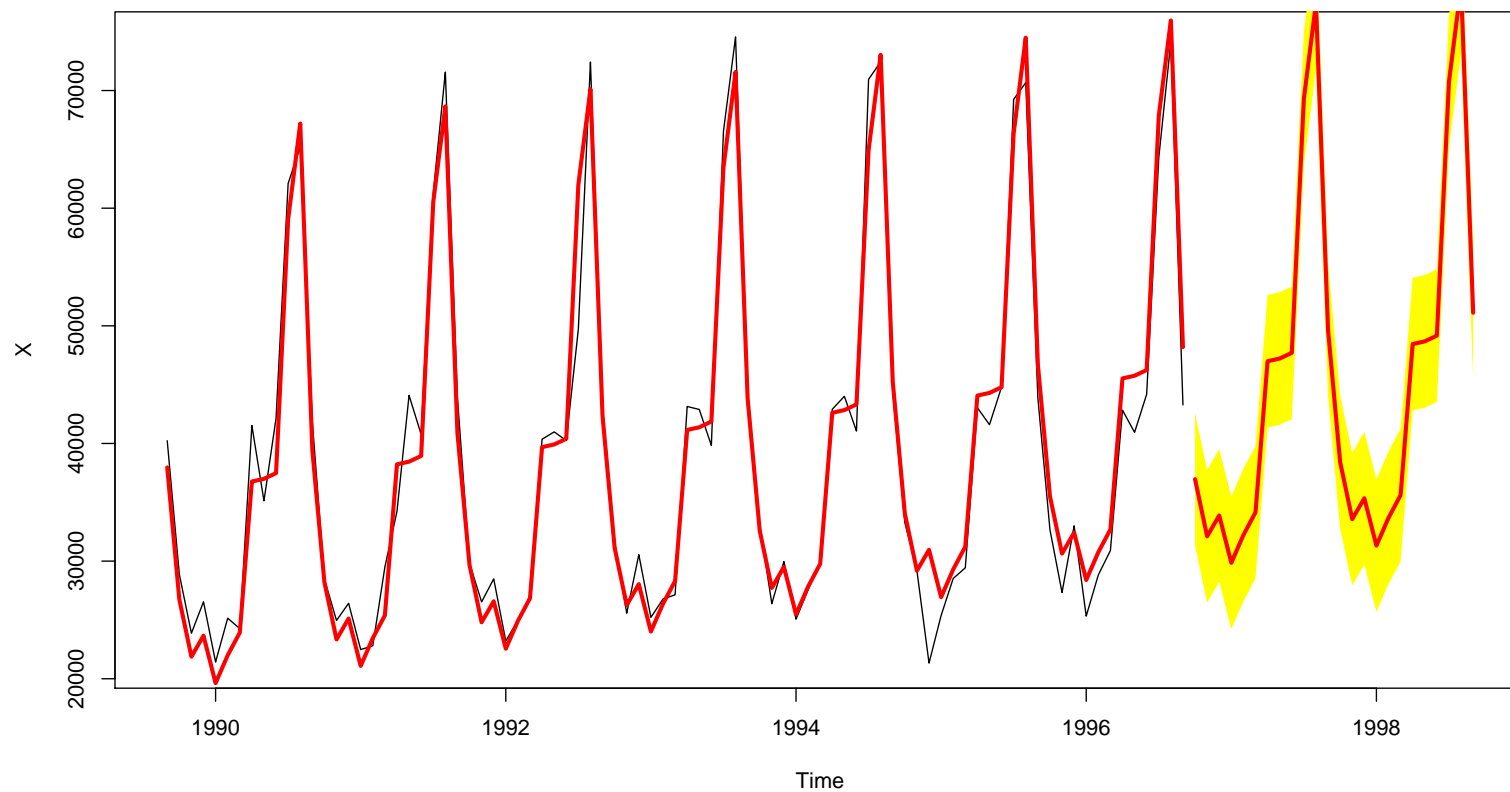
Méthode de Buys Balot - prévision

On peut alors faire de la prévision, en supposant le bruit Z_t i.i.d., i.e.

$${}_T\hat{X}_{T+h} = \mathbb{E}(X_{T+h}|X_1 \cdots, X_T) = \hat{\beta}_0 + \hat{\beta}_1[T+h] + \sum_{k=0}^{s-1} \hat{\gamma}_k \mathbf{1}(T+h = k \bmod s)$$

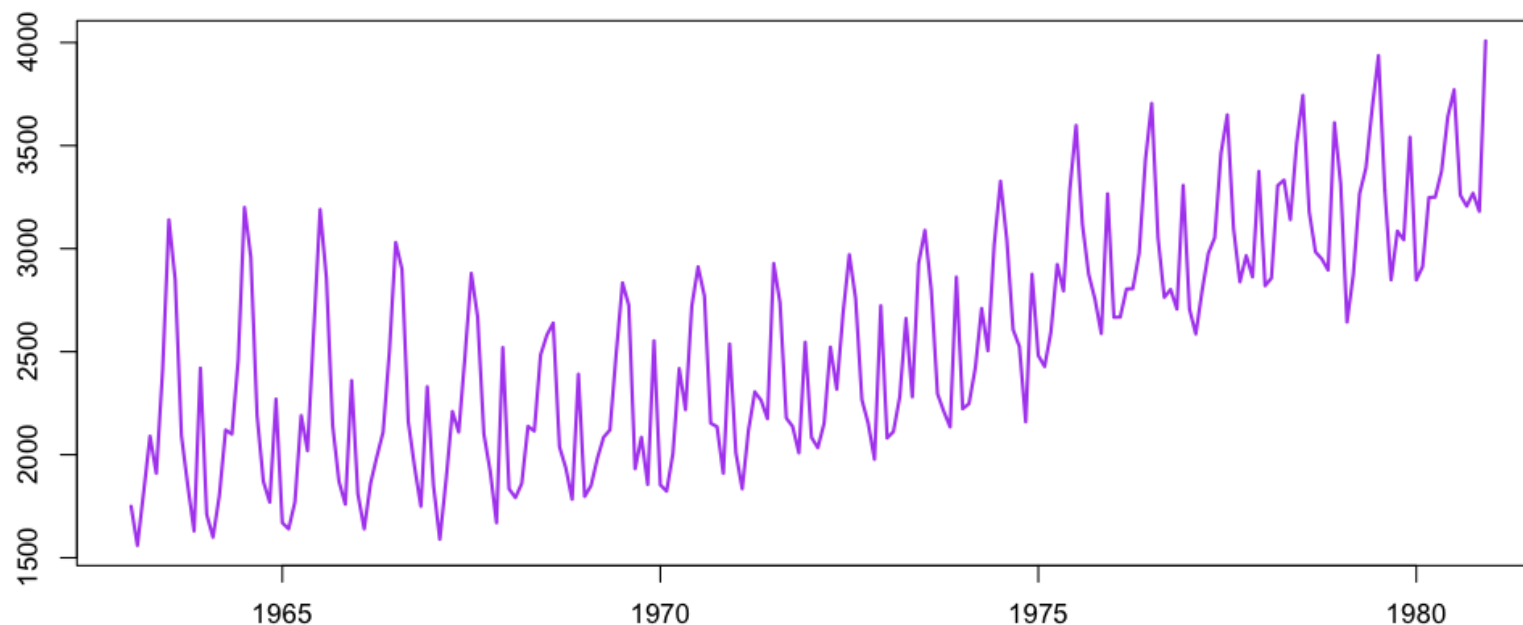
```
> n=length(T)
> T0=T[(n-23):n]+2
> plot(X,xlim=range(c(T,T0)))
> X1p=predict(regT,newdata=data.frame(T=T0))
> Yp=X1p+B$X2[(n-23):n]
> se=sd(X-Y)
> YpU=Yp+1.96*se
> YpL=Yp-1.96*se
> polygon(c(T0,rev(T0)),c(YpU,rev(YpL)),col="yellow",border=NA)
> lines(T0,Yp,col="red",lwd=3)
> lines(B$T,Y,lwd=3,col="red")
```

Méthode de Buys Balot - prévision



Formalisation de Buys-Ballot (1847)

```
> sncf=read.table(  
+ "http://freakonometrics.blog.free.fr/public/data/sncf.csv",  
+ header=TRUE,sep=";")  
> SNCF=ts(as.vector(t(as.matrix(sncf[,2:13]))),  
+ ,start = c(1963, 1), frequency = 12)  
> plot(SNCF,lwd=2,col="purple")
```



Formalisation de Buys-Ballot (1847)

La série X_t est la somme de 2 composantes **déterministes** : une **tendance** Z_t , d'une **saisonnalité** S_t et d'une **composante aléatoire** ε_t

$$X_t = Z_t + S_t + \varepsilon_t.$$

On suppose que Z_t et S_t sont des combinaisons linéaires de fonctions connues dans le temps, Z_t^i et S_t^j , i.e.

$$\begin{cases} Z_t = \beta_0 + Z_t^1 \beta_1 + Z_t^2 \beta_2 + \dots + Z_t^m \beta_m & (\text{ex : } \beta_0 + \beta_1 \cdot t) \\ S_t = S_t^1 \gamma_1 + S_t^2 \gamma_2 + \dots + S_t^s \gamma_s. \end{cases}$$

Le but est d'estimer les β_1, \dots, β_m et $\gamma_1, \dots, \gamma_s$ à partir des T observations.

$$X_t = \sum_{i=1}^m Z_t^i \beta_i + \sum_{j=1}^s S_t^j \gamma_j + \varepsilon_t \text{ pour } t = 1, \dots, T.$$

Formalisation de Buys-Ballot (1847)

La forme de S_t dépend du type de données, et de la forme de la saisonnalité. On considèrera ici des fonctions S_t^i indicatrices,

$$S_t^i = \begin{cases} 0 & \text{si } t = \text{mois } i \\ 1 & \text{si } t \neq \text{mois } i \end{cases} \quad \text{ou} \quad S_t^i = \begin{cases} 0 & \text{si } t = 0 \text{ [modulo } i] \\ 1 & \text{si } t \neq 0 \text{ [modulo } i] . \end{cases}$$

Exemple : Considérons des données [trimestrielles](#),

$$X_t = \underbrace{\alpha + \beta \cdot t}_{Z_t} + \underbrace{\gamma_1 S_t^1 + \gamma_2 S_t^2 + \gamma_3 S_t^3 + \gamma_4 S_t^4}_{S_t} + \varepsilon_t,$$

Que l'on peut écrire de façon matricielle,

$$\begin{pmatrix} 5130 \\ 6410 \\ 8080 \\ 5900 \\ 5110 \\ 6680 \\ 8350 \\ 5910 \\ 5080 \\ \vdots \\ X_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 4 & 0 & 0 & 0 & 1 \\ 1 & 5 & 1 & 0 & 0 & 0 \\ 1 & 6 & 0 & 1 & 0 & 0 \\ 1 & 7 & 0 & 0 & 1 & 0 \\ 1 & 8 & 0 & 0 & 0 & 1 \\ 1 & 9 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t & S_t^1 & S_t^2 & S_t^3 & S_t^4 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \vdots \\ \varepsilon_t \end{pmatrix}$$

Formalisation de Buys-Ballot (1847)

Remark : pour transformer les données en données trimestrielles (et non plus mensuelles)

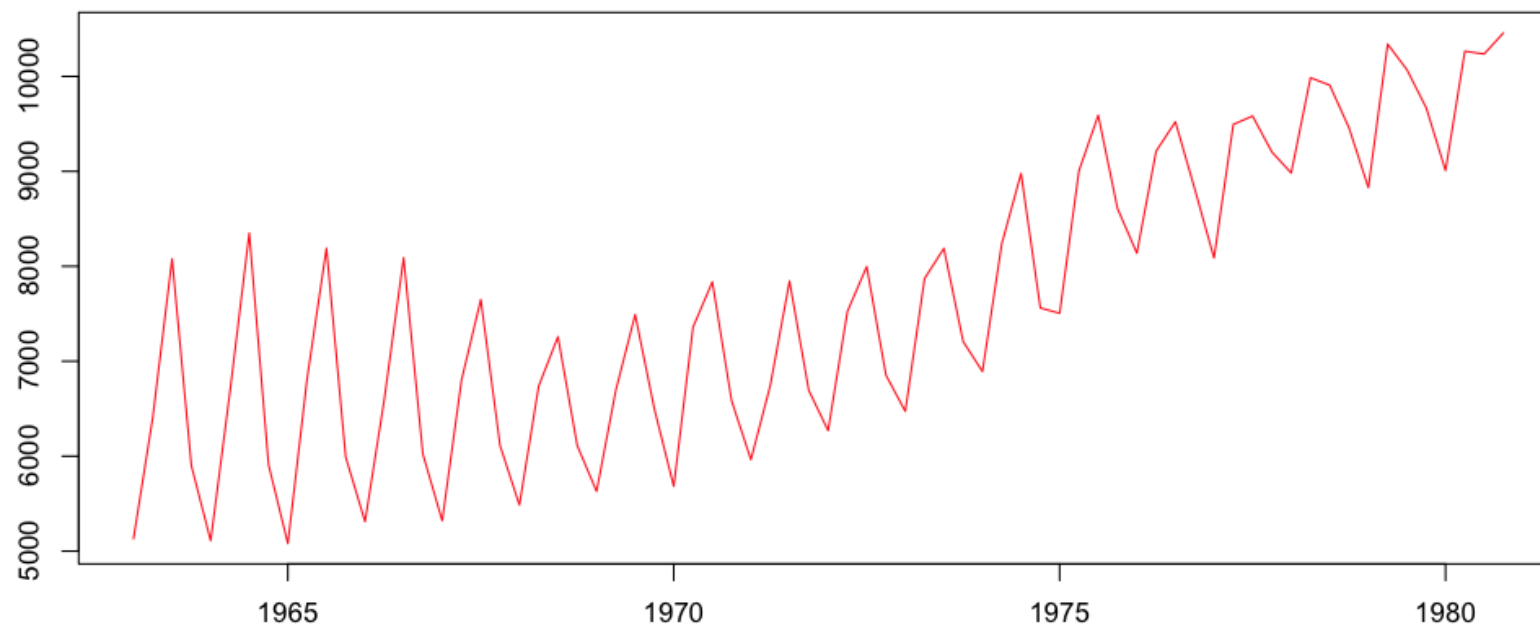
```
> SNCFQ= ts(apply(matrix(as.numeric(SNCF),3,length(SNCF)/3),2,sum),  
+ start = c(1963, 1), frequency = 4)  
> plot(SNCFQ,col="red")  
> SNCFQ
```

	Qtr1	Qtr2	Qtr3	Qtr4
1963	5130	6410	8080	5900
1964	5110	6680	8350	5910
1965	5080	6820	8190	5990
1966	5310	6600	8090	6020

... etc.

Formalisation de Buys-Ballot (1847)

Le graphique des données trimestrielles est le suivant



Formalisation de Buys-Ballot (1847)

On considère un modèle de la forme

$$X_t = \sum_{i=1}^m Z_t^i \beta_i + \sum_{j=1}^s S_t^j \gamma_j + \varepsilon_t \text{ pour } t = 1, \dots, T.$$

La méthode des moindres carrés ordinaires consiste à choisir les β_i et γ_j de façon à minimiser le carré des erreurs

$$\begin{aligned} (\hat{\beta}_i, \hat{\gamma}_j) &= \arg \min \left\{ \sum_{t=1}^T \varepsilon_t^2 \right\} \\ &= \arg \min \left\{ \sum_{t=1}^T \left[X_t - \sum_{i=1}^m Z_t^i \beta_i + \sum_{j=1}^s S_t^j \gamma_j \right]^2 \right\}. \end{aligned}$$

Formalisation de Buys-Ballot (1847)

```
> T = seq(from=1963,to=1980.75,by=.25)
> Q = rep(1:4,18)
> reg=lm(SNCFQ~0+T+as.factor(Q))
> summary(reg)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
T	231.87	12.55	18.47	<2e-16 ***
as.factor(Q)1	-450526.26	24752.39	-18.20	<2e-16 ***
as.factor(Q)2	-449257.44	24755.53	-18.15	<2e-16 ***
as.factor(Q)3	-448644.19	24758.67	-18.12	<2e-16 ***
as.factor(Q)4	-449880.94	24761.81	-18.17	<2e-16 ***

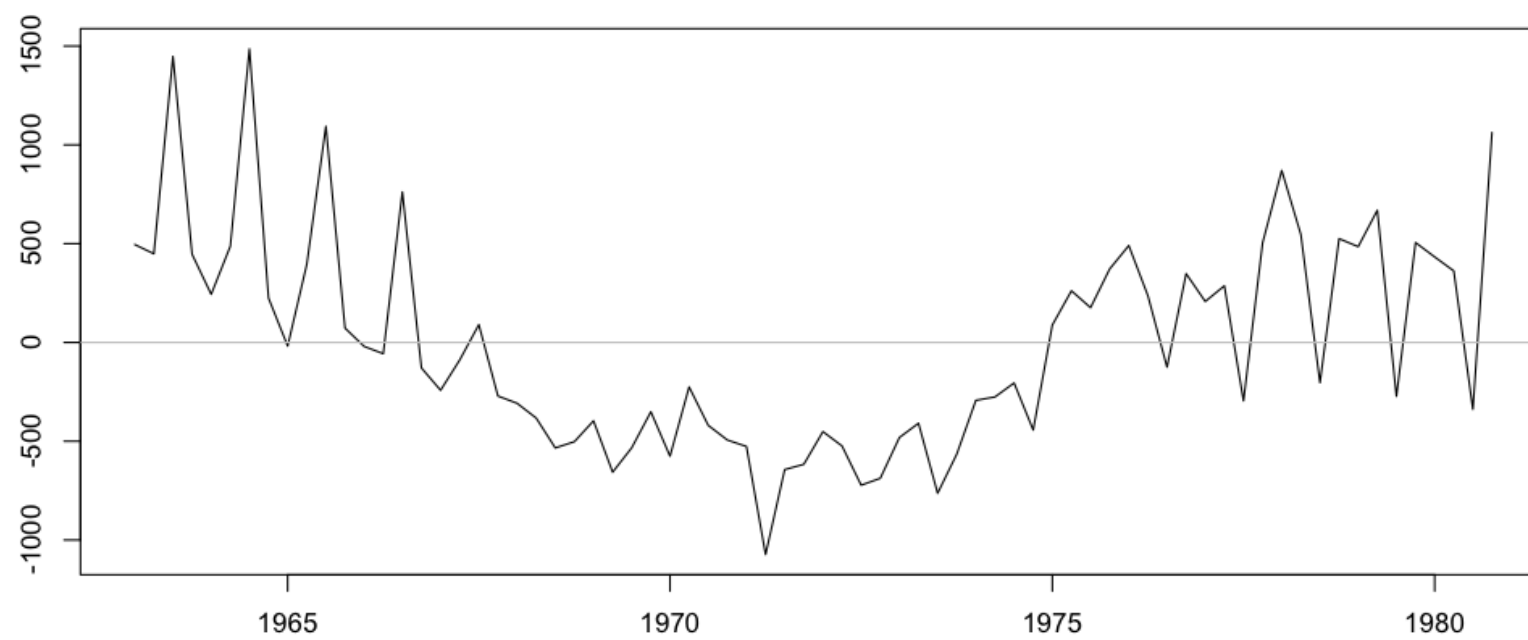
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 552.7 on 67 degrees of freedom

Multiple R-squared: 0.9953, Adjusted R-squared: 0.995

F-statistic: 2846 on 5 and 67 DF, p-value: < 2.2e-16

```
> plot(T,residuals(reg),type="l")
```



Formalisation de Buys-Ballot (1847) : prévision

Soit $h \geq 1$. On suppose que le modèle reste valide en $T + h$ c'est à dire que

$$X_{T+h} = \sum_{i=1}^m Z_{T+h}^i \beta_i + \sum_{j=1}^s S_{T+h}^j \gamma_j + \varepsilon_{T+h},$$

avec $\mathbb{E}(\varepsilon_{T+h}) = 0$, $V(\varepsilon_{T+h}) = \sigma^2$ et $\text{cov}(\varepsilon_t, \varepsilon_{T+h}) = 0$ pour $t = 1, \dots, T$. La variable X_{T+h} peut être approchée par

$${}_T\hat{X}_{T+h} = \sum_{i=1}^m Z_{T+h}^i \hat{\beta}_i + \sum_{j=1}^n S_{T+h}^j \hat{\gamma}_j.$$

Cette prévision est la meilleur (*au sens de l'erreur quadratique moyenne*) prévision, linéaire en X_1, \dots, X_T et sans biais.

Formalisation de Buys-Ballot (1847) : prévision

Un intervalle de confiance de cette prévision est de la forme

$$\left[\hat{X}_T(h) - \phi_{1-\alpha/2} \sqrt{\hat{e}_h}; \hat{X}_T(h) + \phi_{1-\alpha/2} \sqrt{\hat{e}_h} \right],$$

où $\phi_{1-\alpha/2}$ est le quantile d'ordre α de la loi de Student à $T - m - n$ degrés de liberté, et où

$$\begin{aligned} \hat{e}_h &= \hat{\mathbb{E}} \left(\left[\hat{X}_T(h) - X_{T+h} \right]^2 \right) = \hat{V} \left(\sum_{i=1}^m Z_{T+h}^i \hat{\beta}_i + \sum_{j=1}^n S_{T+h}^j \hat{\gamma}_j - \varepsilon_{T+h} \right) \\ &= \left[\hat{\beta}' | \hat{\gamma}' \right] \left[\hat{V} \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \right] \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} + s^2. \end{aligned}$$

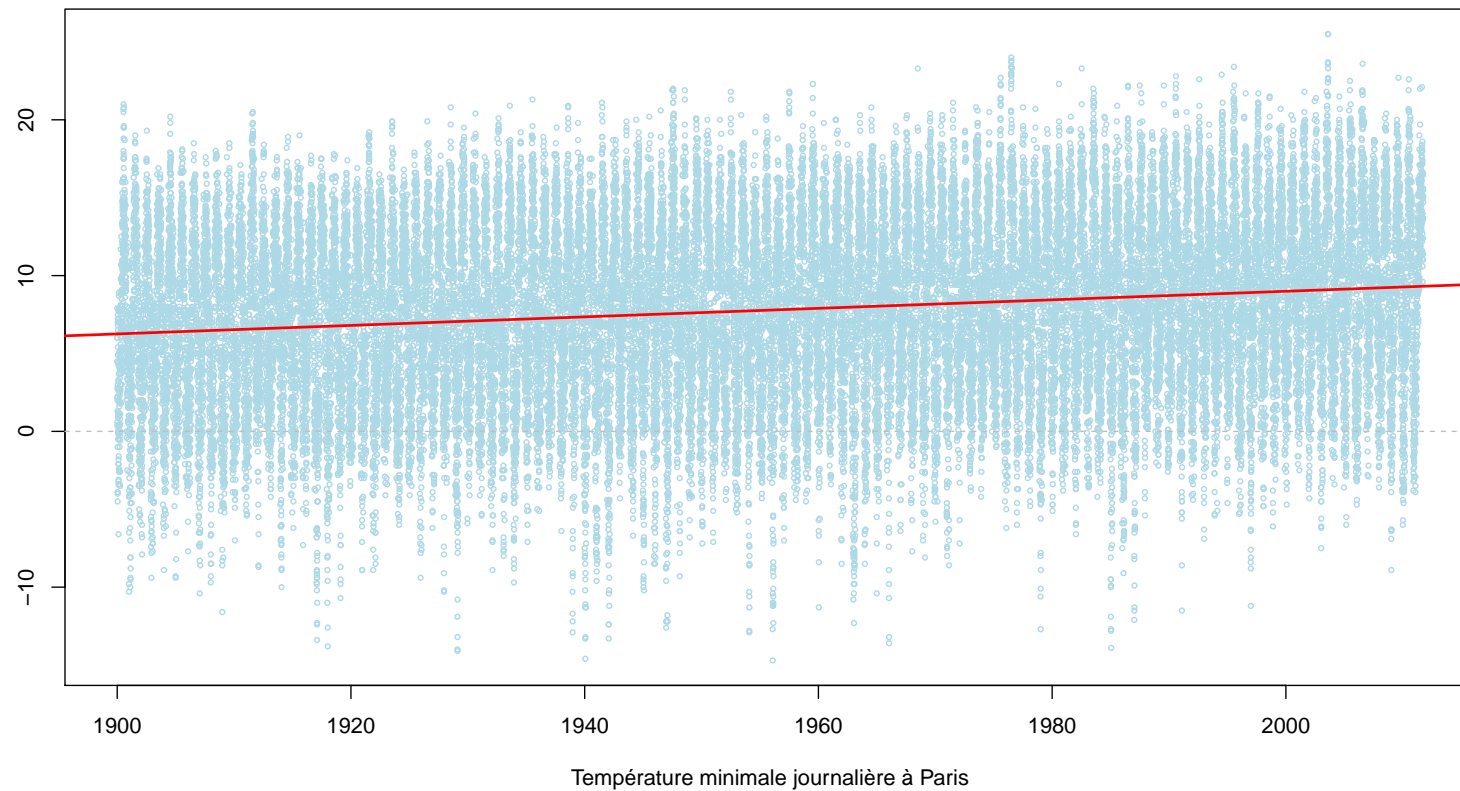
Méthode de Buys Balot et température

On peut le faire sur la modélisation de la température

```
> TEMP=read.table("http://freakonometrics.blog.free.fr/public  
/data/TN_STAID000038.txt",header=TRUE,sep=",")  
> D=as.Date(as.character(TEMP$DATE),"%Y%m%d")  
> T=TEMP$TN/10  
> plot(D,T,col="light blue",xlab="Temp\'erature minimale  
journali\'ere Paris",ylab="",cex=.5)
```

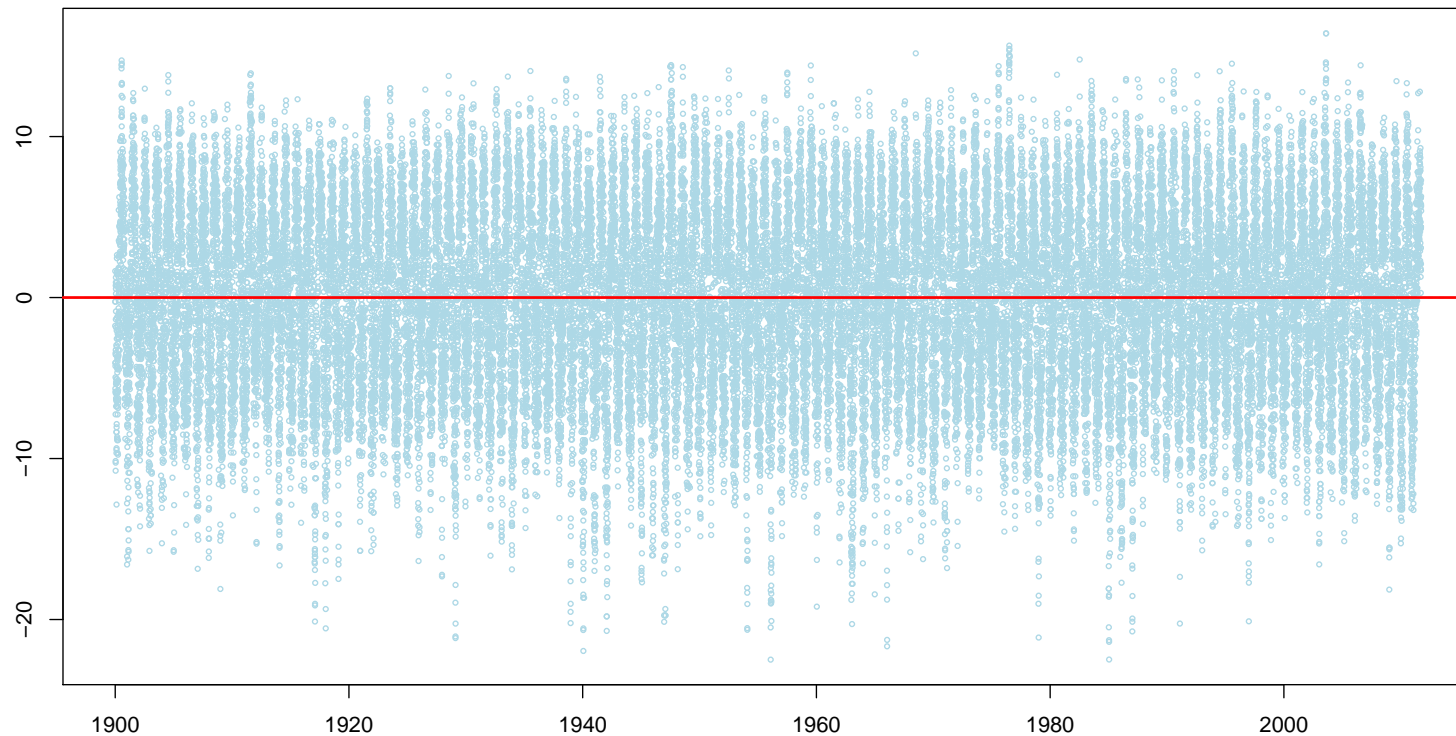
1. Estimer la tendance linéaire $X_t = [\beta_0 + \beta_1 \cdot t] + Y_t$

```
> abline(h=0,lty=2,col="grey")  
> abline(lm(T~D),lwd=2,col="red")
```

```
> X=T-predict(lm(T~D))
> plot(D,X,col="light blue",xlab="",ylab="",cex=.5)
> abline(h=0,lty=2,col="grey")
> abline(lm(X~D),lwd=2,col="red")
```

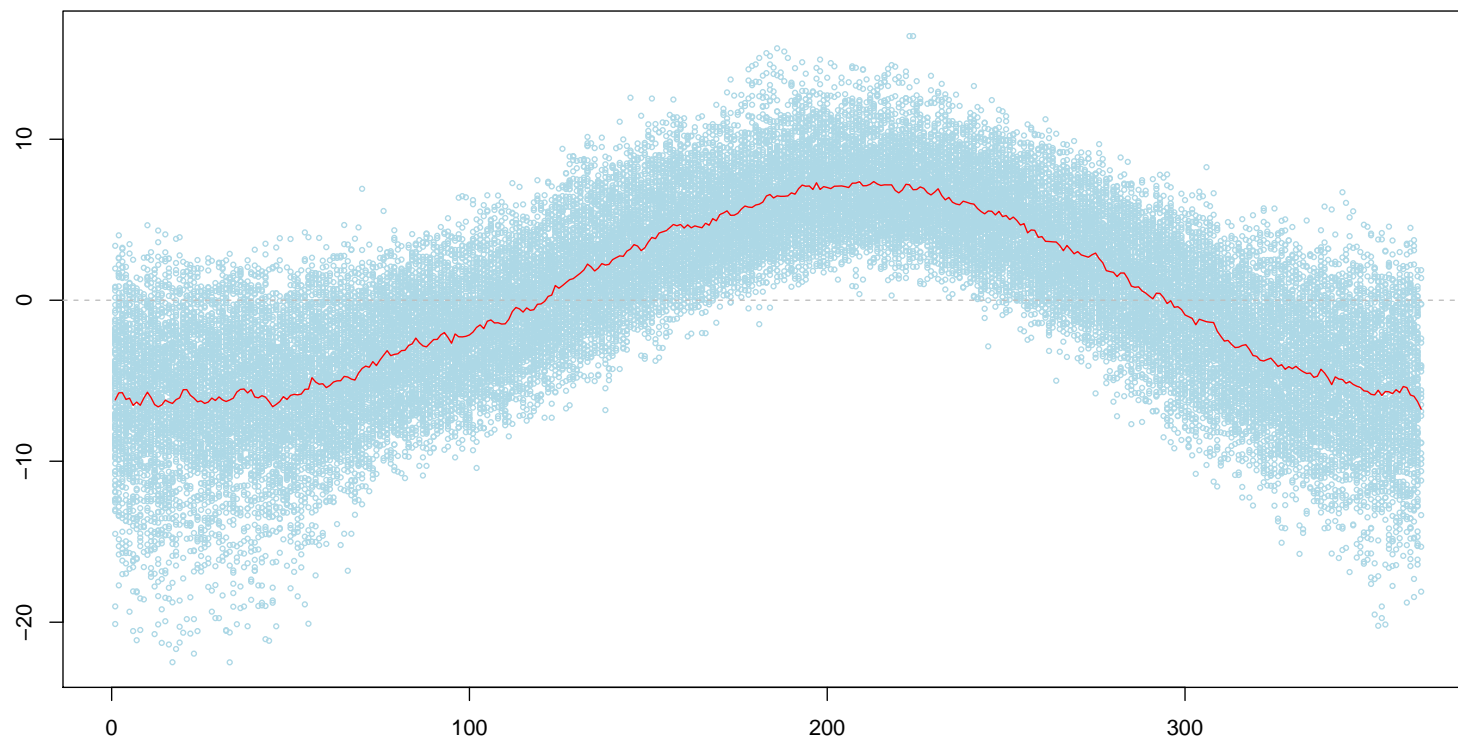
2. Représenter la série résiduelle $Y_t = X_t - [\beta_0 + \beta_1 \cdot t]$



```
> day=as.POSIXlt(D)$yday+1
> plot(day,X,col="light blue",xlab="",ylab="",cex=.5)
> abline(h=0,lty=2,col="grey")
```

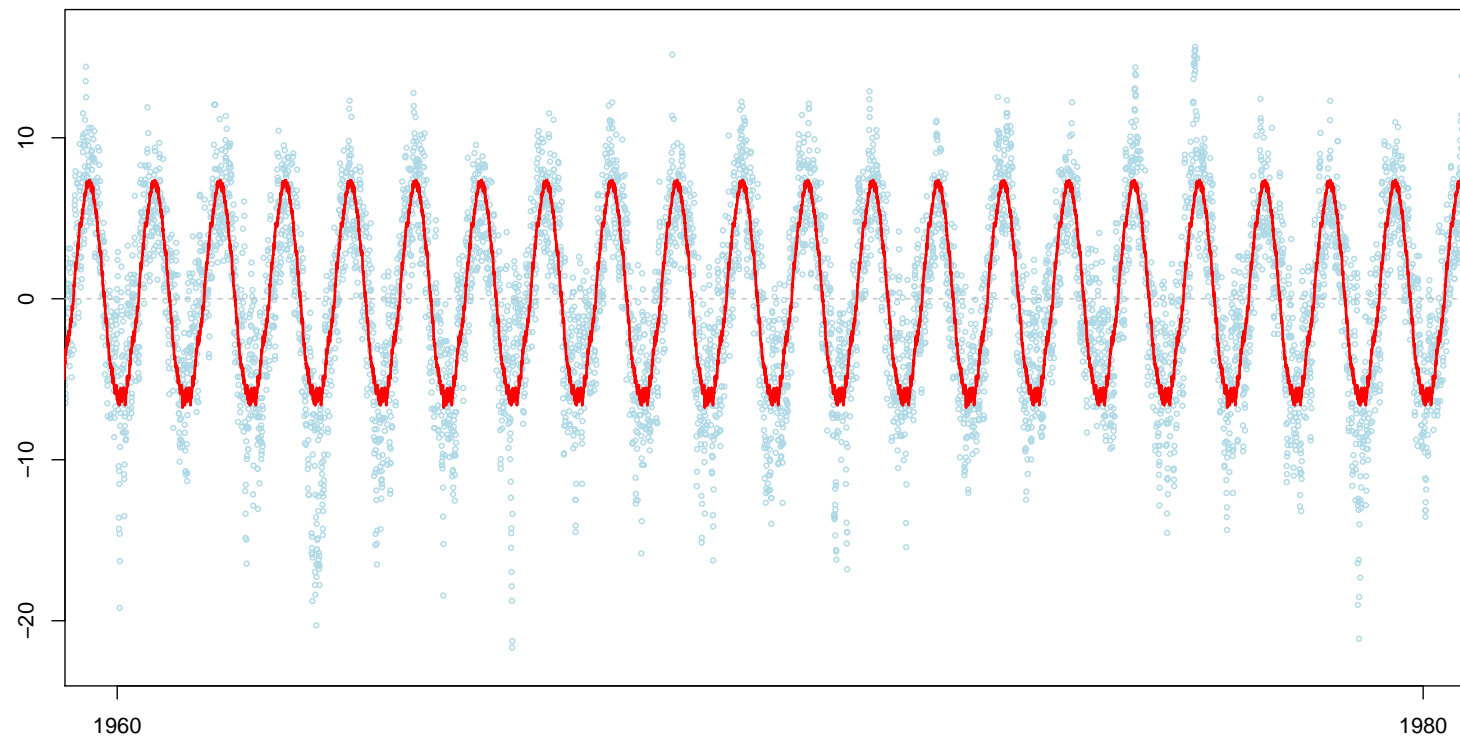
```
> s=tapply(X,as.factor(day),mean)
> lines(1:366,s,col="red")
> cycle=as.vector(s[day])
```

3. Représenter la série résiduelle en fonction de la période supposée Y_t , $t \bmod. s$,



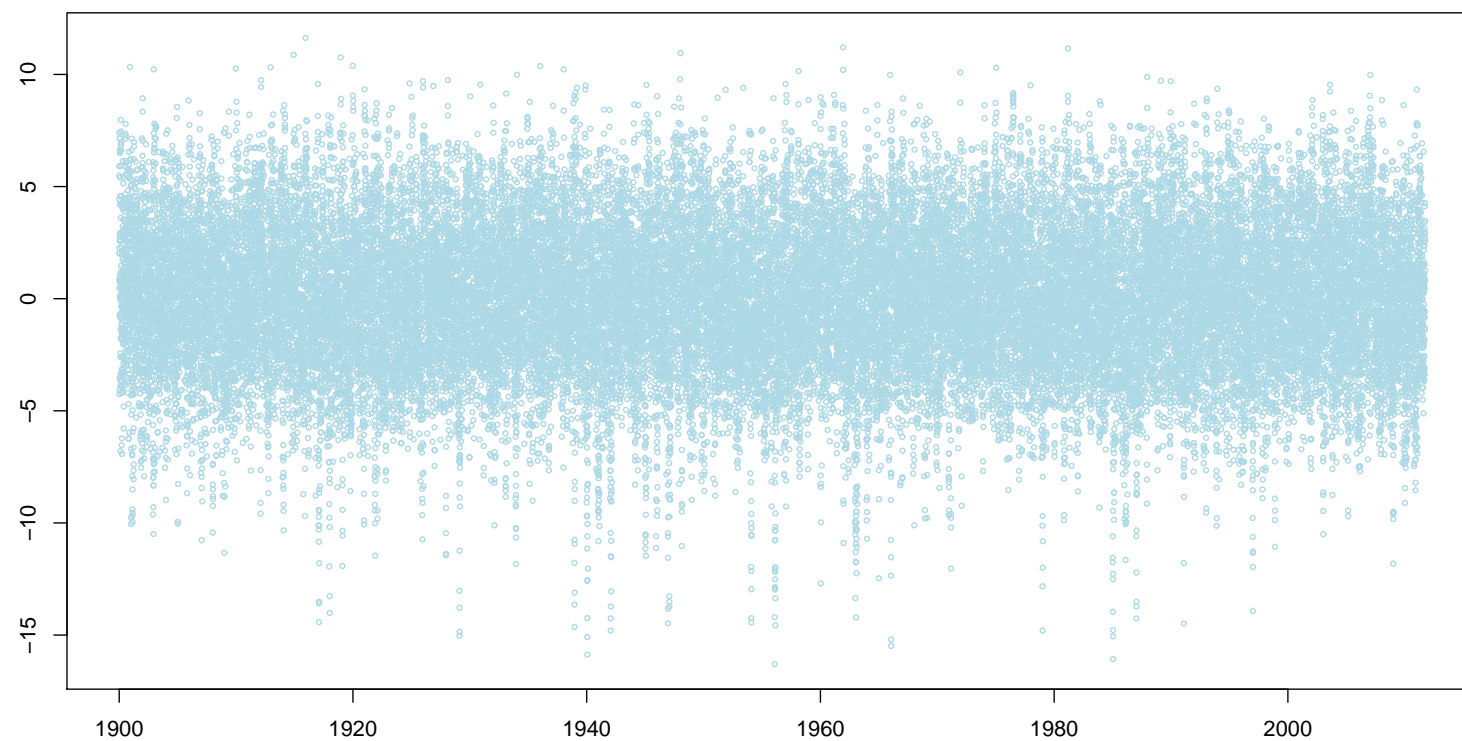
```
> plot(D,X,col="light blue",xlab="",ylab="",cex=.5)
> abline(h=0,lty=2,col="grey")
> lines(D,cycle,lwd=2,col="red")
```

4. Construire la série résiduelle $Y_t = S(t) + Z_t$,



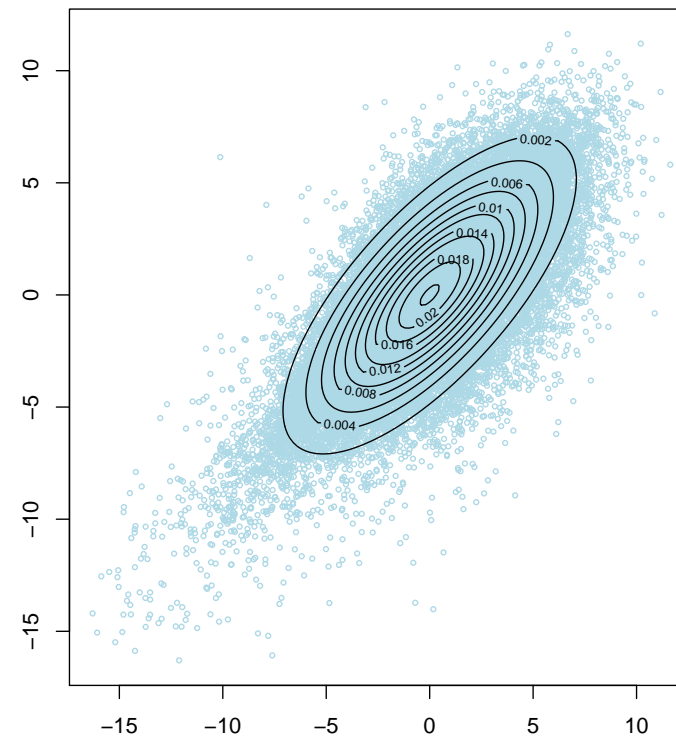
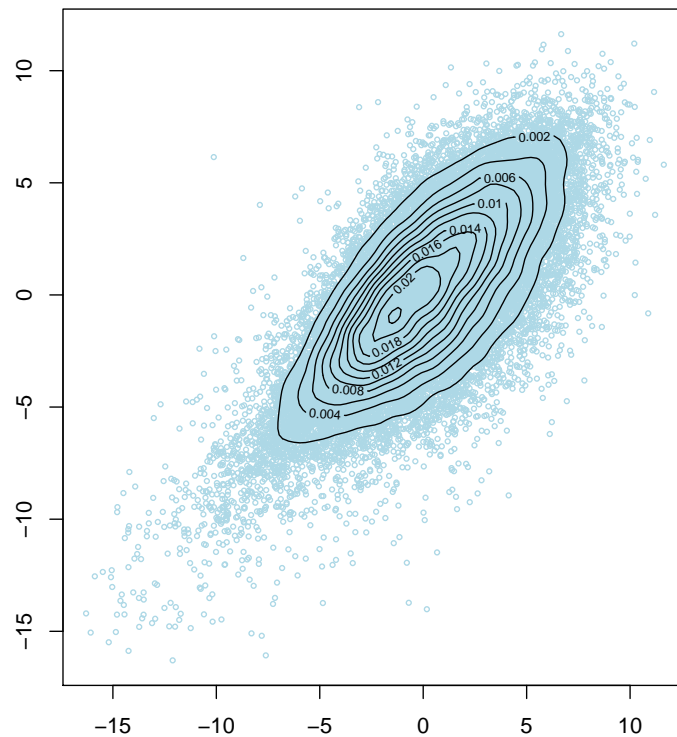
```
> plot(D,X,col="light blue",xlab="",ylab="",cex=.5,  
xlim=as.Date(as.character(c("1990/01/01","2000/01/01"),  
"%Y/%m/%d"))))  
> abline(h=0,lty=2,col="grey")  
> lines(D,cycle,lwd=2,col="red")  
> Z=X-cycle  
> plot(D,Z,col="light blue",xlab="",ylab="",cex=.5)
```

5. ... analyser la série résiduelle Z_t ,



Une (courte) introduction aux autocorrélations

La série résiduelle n'est pas un **bruit blanc** : fortes auto-corrélations....



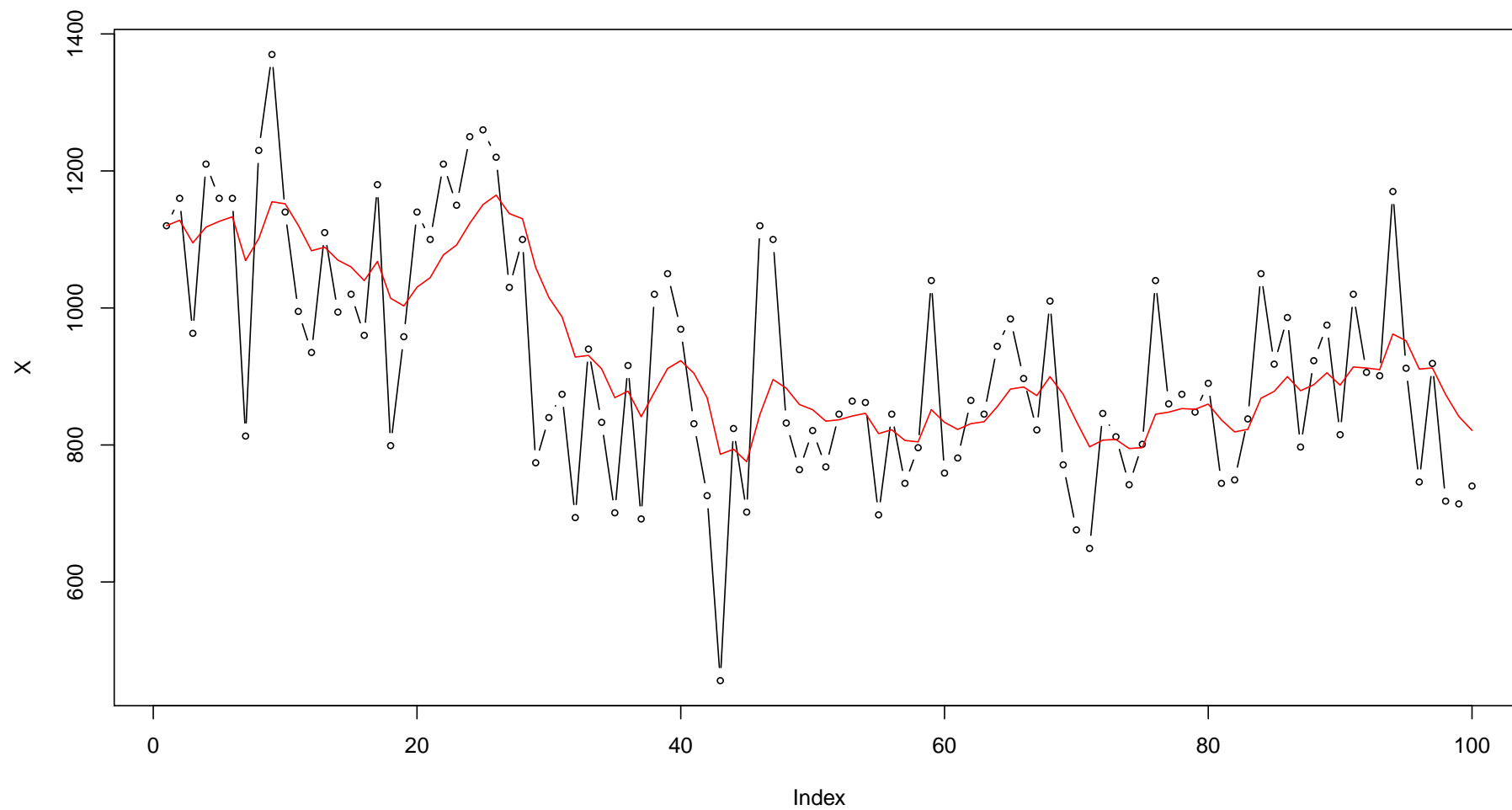
Lissage exponentiel (simple)

À partir d'une série (X_t) on va construire une série (L_t) telle que

$$L_t = \alpha X_t + (1 - \alpha)L_{t-1}, \text{ o ù } \alpha \in (0, 1).$$

avec $L_0 = X_0$.

```
> library(datasets)
> X=as.numeric(Nile)
> Lissage=function(a){
+   T=length(X)
+   L=rep(NA,T)
+   L[1]=X[1]
+   for(t in 2:T){L[t]=a*X[t]+(1-a)*L[t-1]}
+   return(L)
+ }
> plot(X,type="b",cex=.6)
> lines(Lissage(.2),col="red")
```

Lissage exponentiel (simple)

On parle de **lissage exponentiel** car

$$\begin{aligned} L_t &= \alpha X_t + (1 - \alpha) L_{t-1} \\ &= \alpha (X_t + (1 - \alpha) X_{t-1}) + (1 - \alpha)^2 L_{t-2} \\ &\dots \\ &= \alpha (X_t + (1 - \alpha) X_{t-1} + \dots + (1 - \alpha)^k X_{t-k} + \dots) + (1 - \alpha)^t X_0 \end{aligned}$$

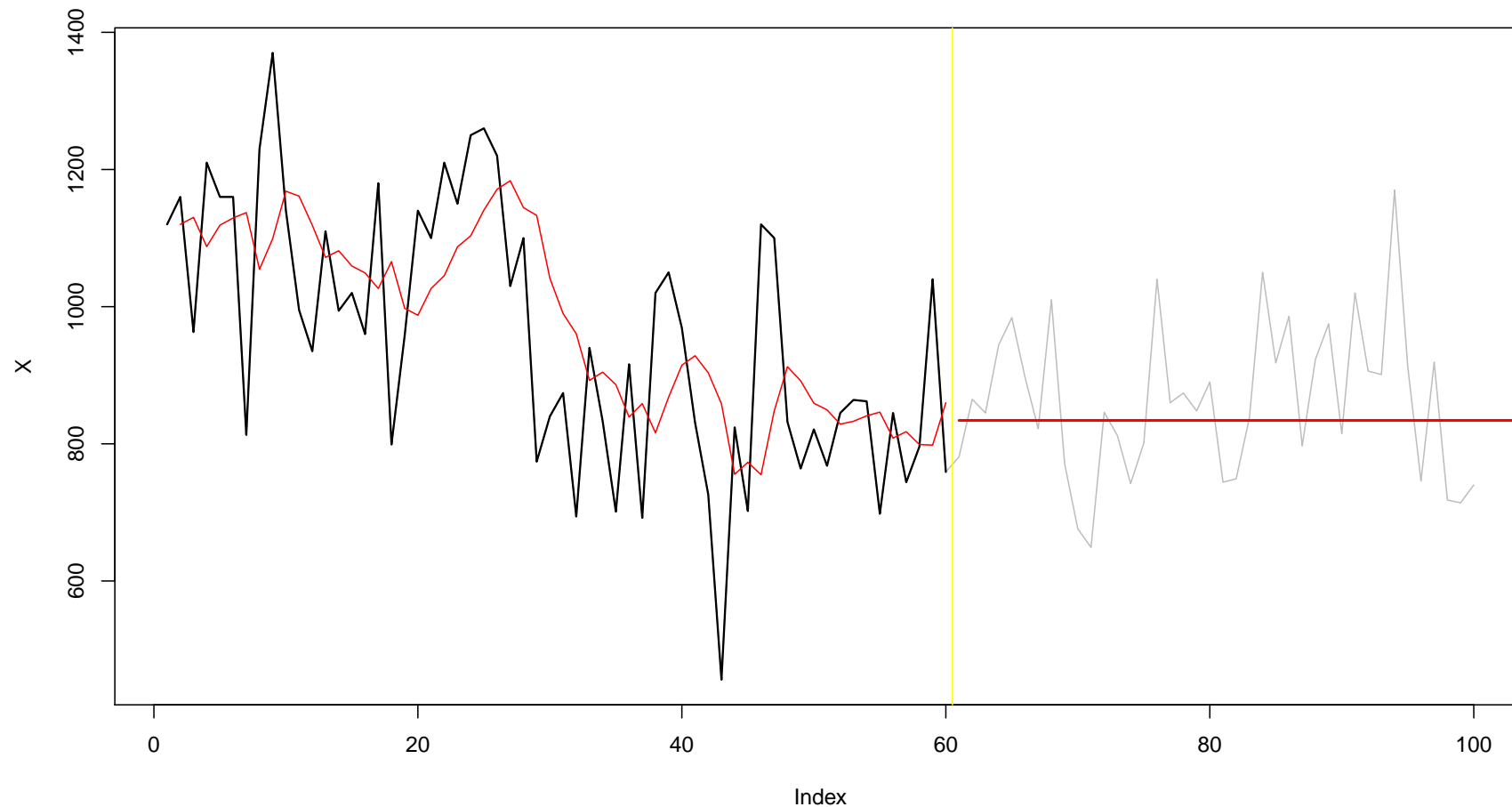
Cette progression géométrique est une discrétisation de la fonction exponentielle...

L'idée est que si l'on cherche à faire une prévision, pour un horizon h , à partir de t , on pose

$${}_t\hat{X}_{t+h} = L_t$$

Lissage exponentiel (simple)

```
> plot(HoltWinters(X,beta=FALSE,gamma=FALSE)$fitted[,1],col="red")  
> lines(predict(hw,n.ahead=50),col="red",lwd=2)
```



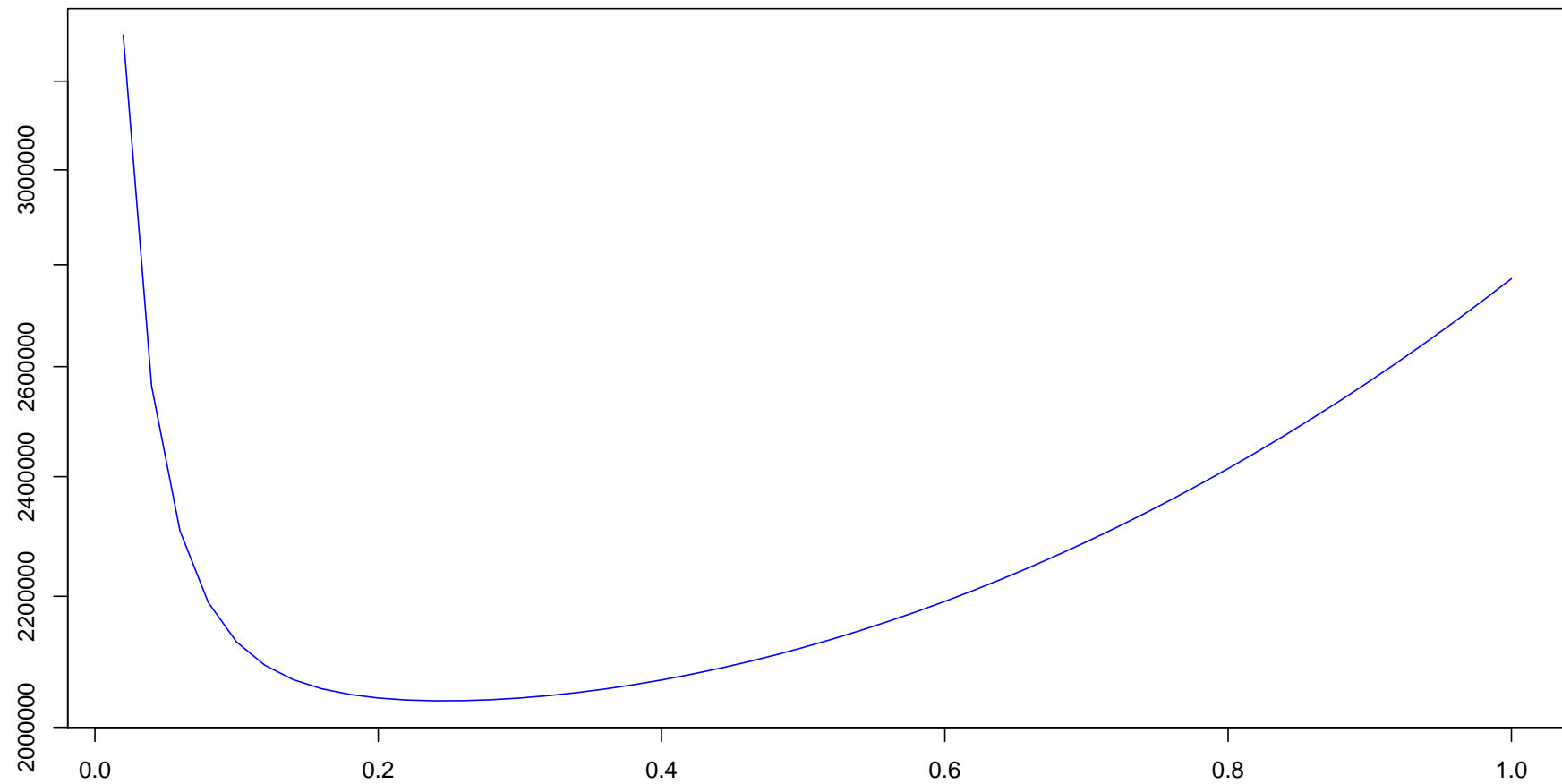
Lissage exponentiel (simple)

Le poids optimal α sera celui qui minimise l'erreur commise à un horizon $h = 1$, i.e.

$$\alpha^* = \operatorname{argmin} \left\{ \sum_{t=0}^{T-1} [L_t - X_{t+1}]^2 \right\}$$

```
> V=function(a){  
+   T=length(X)  
+   L=erreur=rep(NA,T)  
+   erreur[1]=0  
+   L[1]=X[1]  
+   for(t in 2:T){  
+     L[t]=a*X[t]+(1-a)*L[t-1]  
+     erreur[t]=X[t]-L[t-1]  }  
+   return(sum(erreur^2))  
+ }  
> optimize(V,c(0,.5))$minimum  
[1] 0.246581
```

Lissage exponentiel (simple)



Lissage exponentiel (simple)

```
> hw=HoltWinters(X,beta=FALSE,gamma=FALSE)
```

```
> hw
```

Holt-Winters exponential smoothing without trend an seasonal comp.

Call:

```
HoltWinters(x = X, beta = FALSE, gamma = FALSE, l.start = X[1])
```

Smoothing parameters:

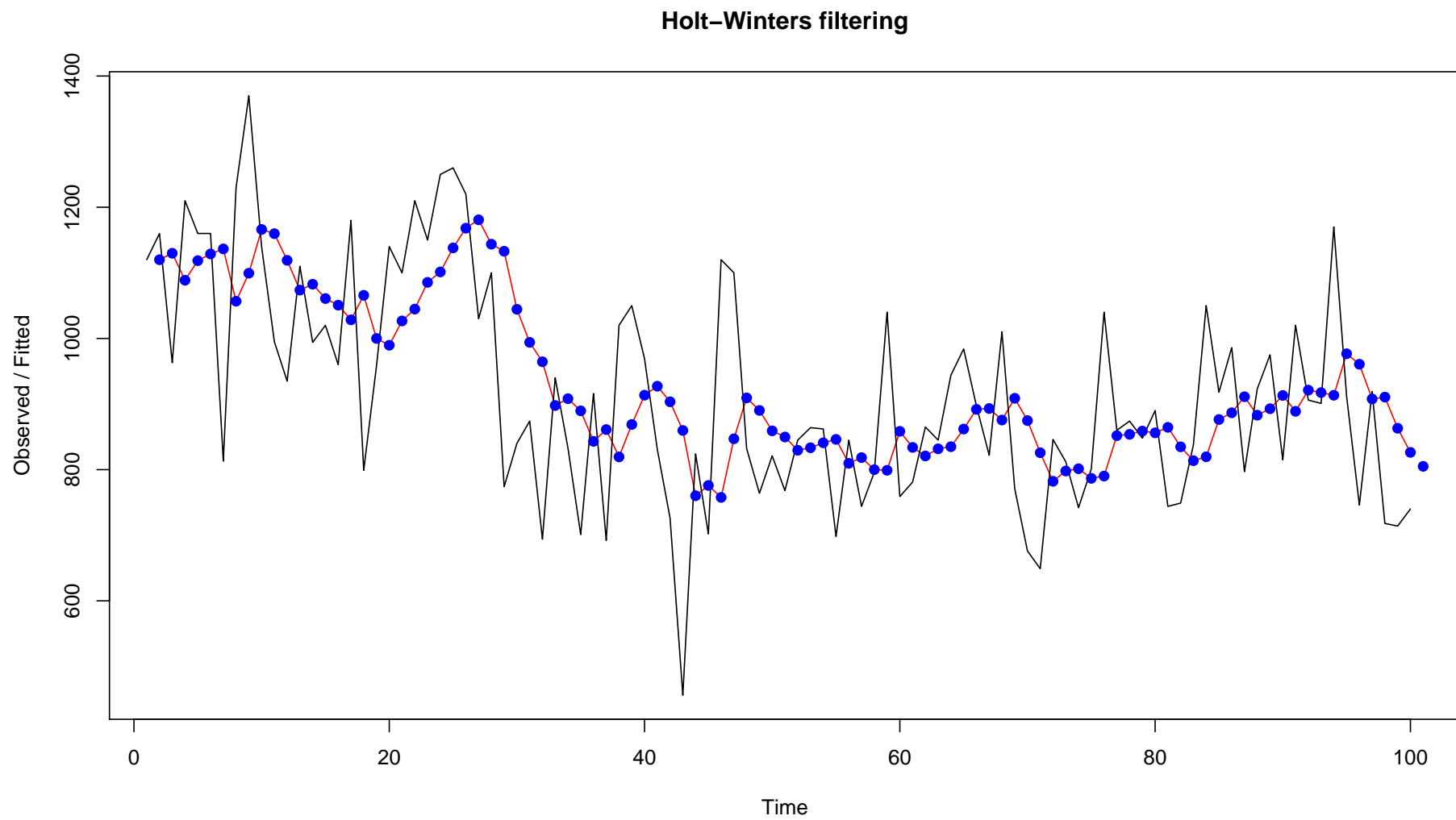
```
alpha: 0.2465579
```

```
beta : FALSE
```

```
gamma: FALSE
```

```
> plot(hw)
```

```
> points(2:(length(X)+1),Vectorize(Lissage)(.2465),col="blue")
```



Lissage exponentiel (double)

On va maintenant construire deux séries,

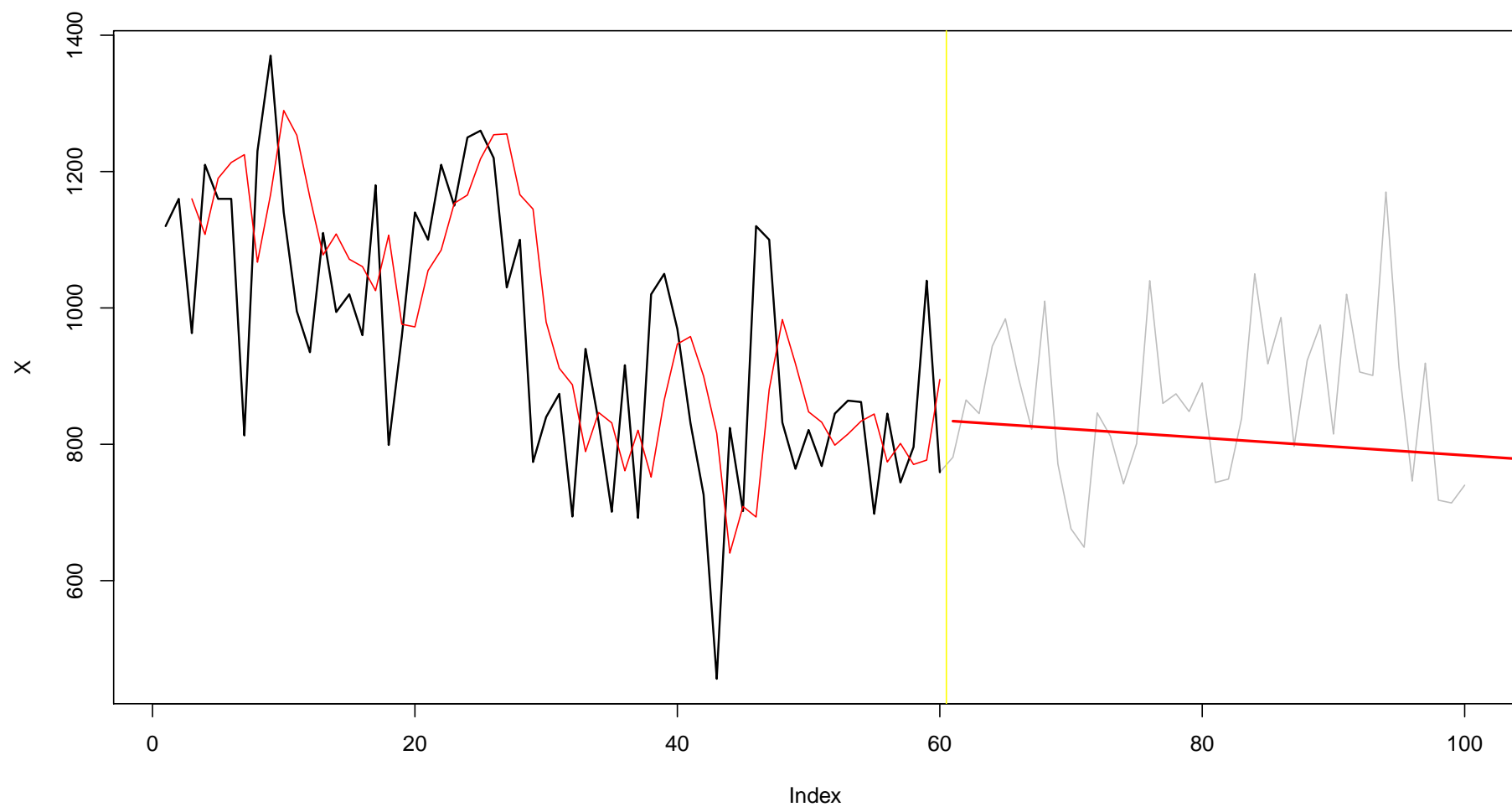
$$\begin{cases} L_t = \alpha X_t + (1 - \alpha)(L_{t-1} + B_{t-1}) \\ B_t = \beta(L_t - L_{t-1}) + (1 - \beta)B_{t-1} \end{cases}$$

où $0 < \alpha, \beta < 1$, avec les conditions initiales $L_0 = X_0$ et $B_0 = X_1 - X_0$. Cette fois, la prédiction s'écrit

$${}_t\hat{X}_{t+h} = L_t + hB_t$$

C'est une généralisation du lissage simple au sens où on autorise la série à avoir une tendance linéaire.

Lissage exponentiel (double)

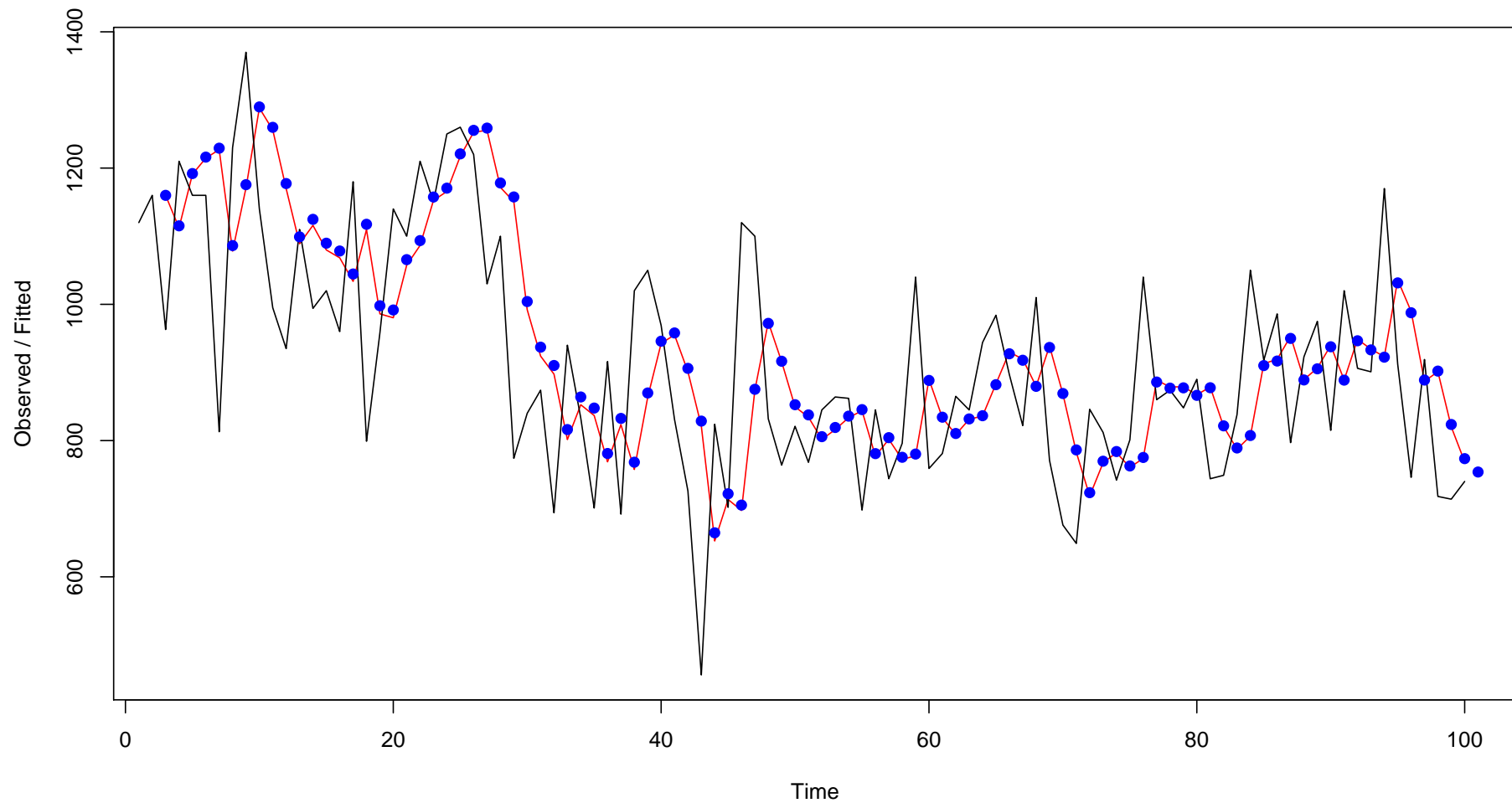


Lissage exponentiel (double)

```
> Lissage=function(a,b){  
+   T=length(X)  
+   L=B=rep(NA,T)  
+   L[1]=X[1]; B[1]=0  
+   for(t in 2:T){  
+     L[t]=a*X[t]+(1-a)*(L[t-1]+B[t-1])  
+     B[t]=b*(L[t]-L[t-1])+(1-b)*B[t-1] }  
+   return(L)  
+ }
```

Lissage exponentiel (double)

Holt-Winters filtering



Lissage exponentiel (double)

Là encore, les poids optimaux (α, β) minimisent l'erreur commise à un horizon $h = 1$, i.e.

$$(\alpha^*, \beta^*) = \operatorname{argmin} \left\{ \sum_{t=0}^{T-1} [\hat{X}_{t+1} - X_{t+1}]^2 \right\}$$

```
> V=function(P){
+   a=P[1];b=P[2]
+   T=length(X)
+   L=B=erreur=rep(NA,T)
+   erreur[1:2]=0
+   L[2]=X[1]; B[2]=X[2]-X[1]
+   for(t in 3:T){
+     L[t]=a*X[t]+(1-a)*(L[t-1]+B[t-1])
+     B[t]=b*(L[t]-L[t-1])+(1-b)*B[t-1]
+     erreur[t]=X[t]-(L[t-1]-B[t-1]) }
+   return(sum(erreur^2))
+ }
> nlm(V,c(.5,.5))$estimate
[1] 0.40972695 0.05007216
```

Lissage exponentiel (double)

```
> hw=HoltWinters(X,gamma=FALSE,l.start=X[1])
```

```
> hw
```

Holt-Winters exponential smoothing with trend and without seasonal component.

Call:

```
HoltWinters(x = X, gamma = FALSE, l.start = X[1])
```

Smoothing parameters:

```
alpha: 0.4200241
```

```
beta : 0.05973389
```

```
gamma: FALSE
```

Lissage exponentiel saisonnier - Holt Winters (1960)

On va maintenant construire trois séries. Le lissage saisonnier peut se faire de manière additive ou multiplicative.

Pour le modèle **saisonnier additif**, de saisonnalité s , pour $t > s$

$$\begin{cases} L_t = \alpha[X_t - S_{t-s}] + (1 - \alpha)(L_{t-1} + B_{t-1}) \\ B_t = \beta(L_t - L_{t-1}) + (1 - \beta)B_{t-1} \\ S_t = \gamma(X_t - L_t) + (1 - \gamma)S_{t-s} \end{cases}$$

où $0 < \alpha, \beta, \gamma < 1$, avec les conditions initiales

$$L_s = \frac{X_1 + \dots + X_s}{s} \text{ et } B_s = \frac{1}{s} \left(\frac{X_{s+1} - X_1}{s} + \dots + \frac{X_{2s} - X_s}{s} \right)$$

avec $S_t = X_t - L_s$ pour $t = 1, \dots, s$. Cette fois, la prédiction s'écrit

$${}_t\hat{X}_{t+h} = L_t + hB_t + S_{t+h-s}$$

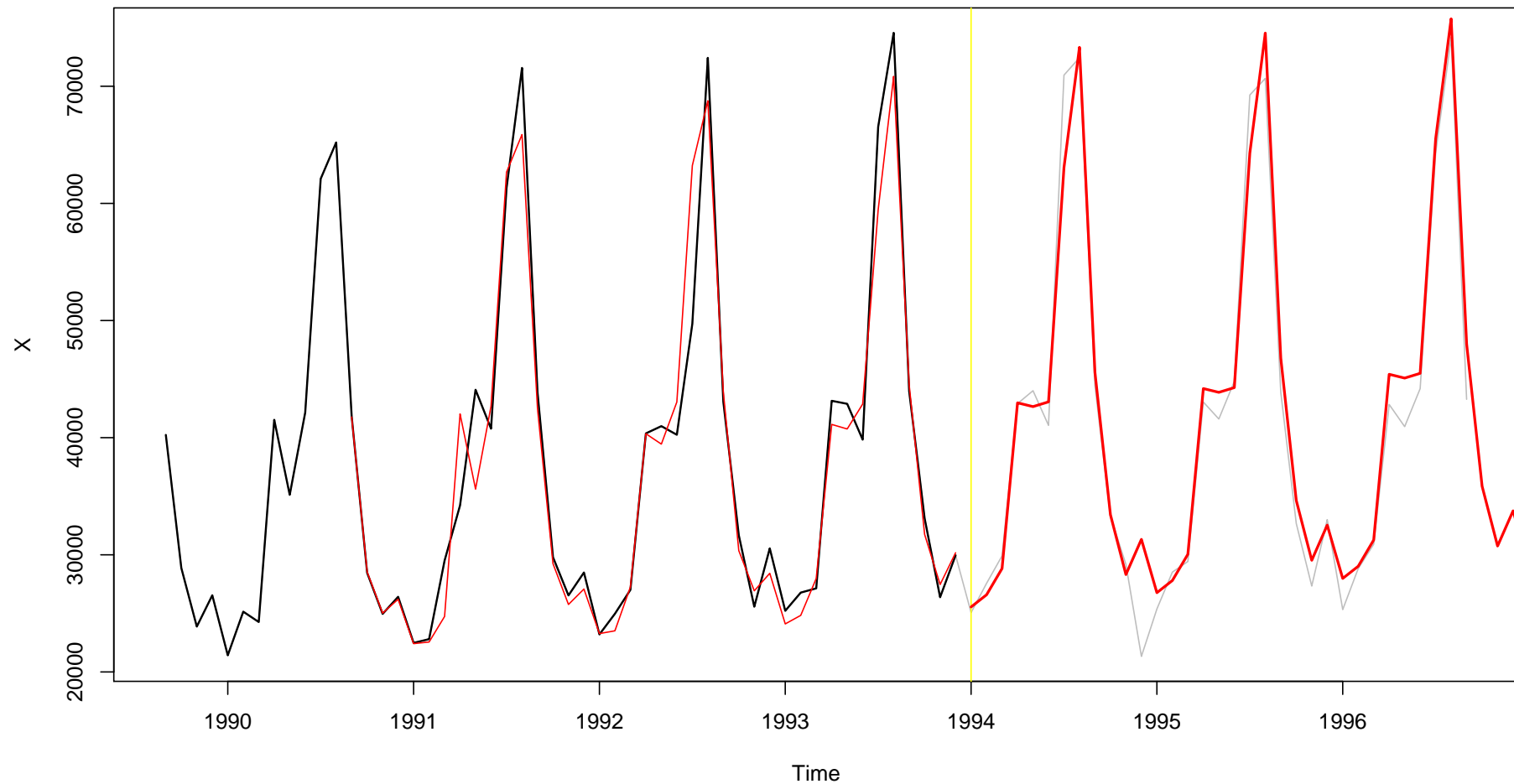
Lissage exponentiel saisonnier - Holt Winters (1960)

Les poids optimaux (α, β, γ) minimisent l'erreur commise à un horizon $h = 1$, en supposant la saisonnalité s donnée

$$(\alpha^*, \beta^*, \gamma^*) = \operatorname{argmin} \left\{ \sum_{t=0}^{T-1} [\hat{X}_{t+1} - X_{t+1}]^2 \right\}$$

```
> Lissage=function(a,b,c,s){
+   T=length(X)
+   L=B=C=pred=rep(NA,T)
+   L[s]=mean(X[1:s]); B[1:s]=mean(X[(s+1):(2*s)]-X[1:s])/s
+   C[1:s]=X[1:s]-mean(X[1:s])
+   for(t in (s+1):(T-1)){
+     L[t]=a*(X[t]-C[t-s])+(1-a)*(L[t-1]+B[t-1]);
+     B[t]=b*(L[t]-L[t-1])+(1-b)*B[t-1];
+     C[t]=c*(X[t]-L[t])+(1-c)*C[t-s];
+     pred[t]=(L[t]+B[t])+C[t-s]};
+   return(pred)}
```

```
> source("http://freakonometrics.blog.free.fr/public/data/sourcets.R")  
> X=autoroute7
```



Lissage exponentiel saisonnier - Holt Winters (1960)

```
> hw=HoltWinters(X,seasonal="additive")
```

```
> hw
```

```
Holt-Winters exponential smoothing with trend and additive seasonal component.
```

```
Call:
```

```
  HoltWinters(x = X, seasonal = "additive")
```

```
Smoothing parameters:
```

```
alpha:  0.01340761
```

```
beta :   0.01853709
```

```
gamma:   0.2473586
```

Lissage exponentiel saisonnier - Holt Winters (1960)

Il existe aussi un modèle **saisonnier multiplicatif**, de saisonnalité s , pour $t > s$

$$\begin{cases} L_t = \alpha X_t \cdot S_{t-s}^{-1} + (1 - \alpha)(L_{t-1} + B_{t-1}) \\ B_t = \beta(L_t - L_{t-1}) + (1 - \beta)B_{t-1} \\ S_t = \gamma(X_t \cdot L_t^{-1}) + (1 - \gamma)S_{t-s} \end{cases}$$

où $0 < \alpha, \beta, \gamma < 1$, avec les conditions initiales

$$L_s = \frac{X_1 + \dots + X_s}{s} \text{ et } B_s = \frac{1}{s} \left(\frac{X_{s+1} - X_1}{s} + \dots + \frac{X_{2s} - X_s}{s} \right)$$

avec $S_t = X_t \cdot L_s^{-1}$ pour $t = 1, \dots, s$. Cette fois, la prédiction s'écrit

$${}_t\hat{X}_{t+h} = (L_t + hB_t) \cdot S_{t+h-s}$$

La stationnarité des séries temporelles

Un processus (Y_t) est **stationnaire au sens fort** si pour tout n, t_1, \dots, t_n et $h \in \mathbb{N}$,

$$\mathcal{L}(Y_{t_1}, \dots, Y_{t_n}) = \mathcal{L}(Y_{t_1+h}, \dots, Y_{t_n+h})$$

Un processus (Y_t) est **stationnaire au sens faible** si

- la moyenne du processus est constante : $\mathbb{E}(Y_t) = m$ pour tout $t \in \mathbb{Z}$,
- les autocovariances ne dépendent que de la différence entre les observations :
 $\text{cov}(X_t, X_s) = \gamma(|t - s|)$

En particulier, la variance de Y_t est constante : $\text{Var}(Y_t) = \sigma^2$.

Exemple : si $(X_t, t \in \mathbb{Z})$ est un processus stationnaire, et si $(a_i, i \in \mathbb{Z})$ est une suite de réels absolument convergente, i.e. $\sum_{i \in \mathbb{Z}} |a_i| < +\infty$, alors, le processus (Y_t) défini par

$$Y_t = \sum_{i \in \mathbb{Z}} a_i X_{t-i}, \text{ pour tout } t \in \mathbb{Z},$$

est un processus stationnaire.

Les non-stationnarités

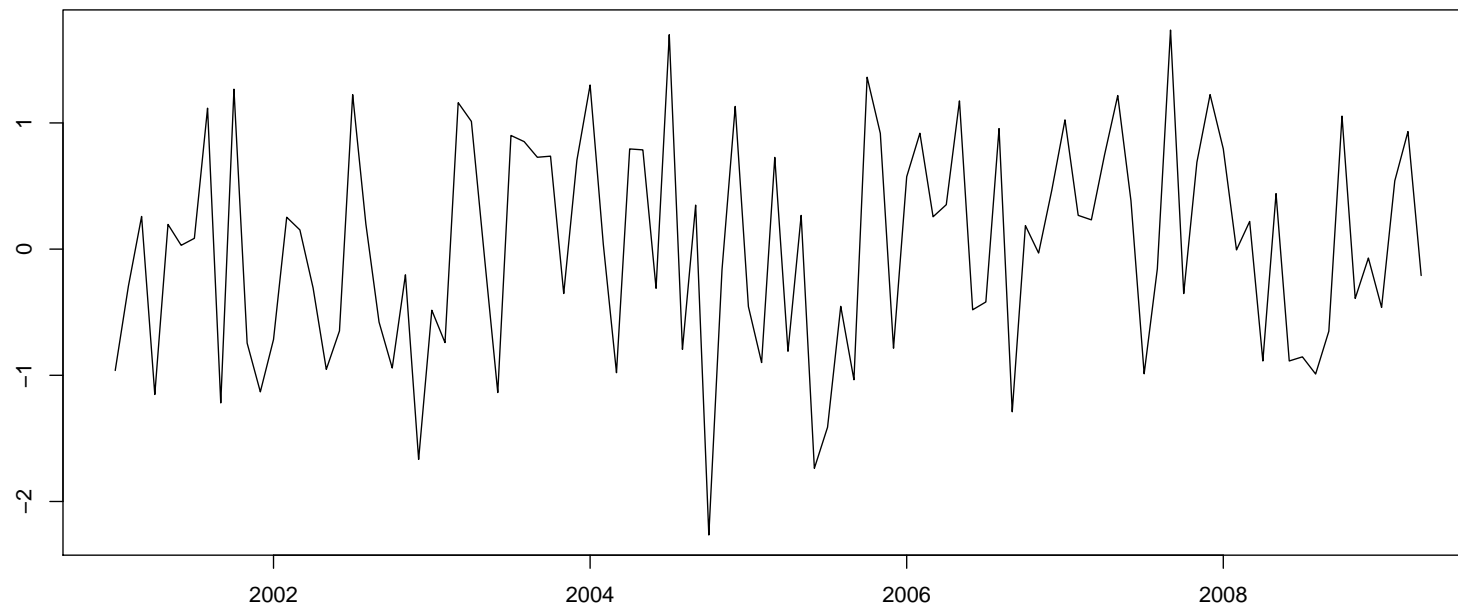
(X_t) est un processus **non-stationnaire TS** s'il peut s'écrire sous la forme $X_t = f(t) + Z_t$ où $f(t)$ est une fonction (déterministe) du temps, et (Z_t) est un processus stationnaire.

Example : $X_t = \alpha + \beta t + \varepsilon_t$

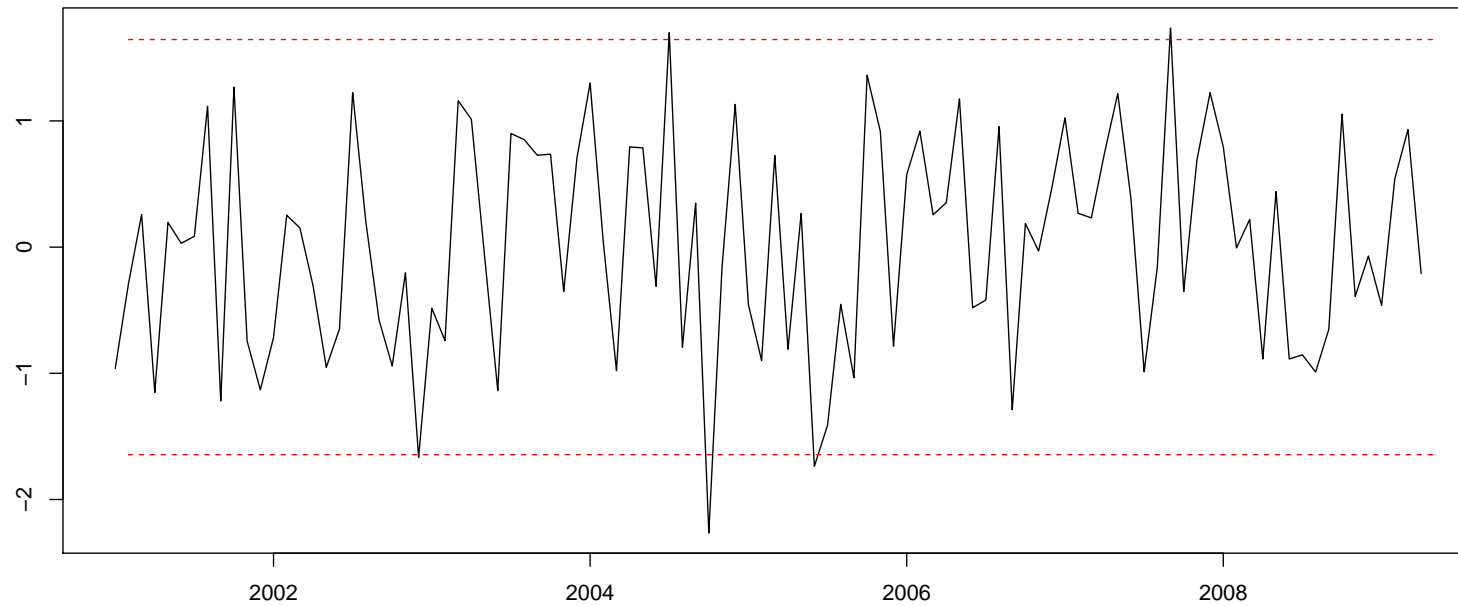
(X_t) est un processus **non-stationnaire DS** - ou intégré d'ordre d , noté $I(d)$ - si le processus obtenu après d différenciation est stationnaire : $Z_t = \Delta^d X_t$ est stationnaire, où $\Delta X_t = X_t - X_{t-1}$.

Example : $X_t = X_{t-1} + \varepsilon_t$.

Stationnarité, série # 1

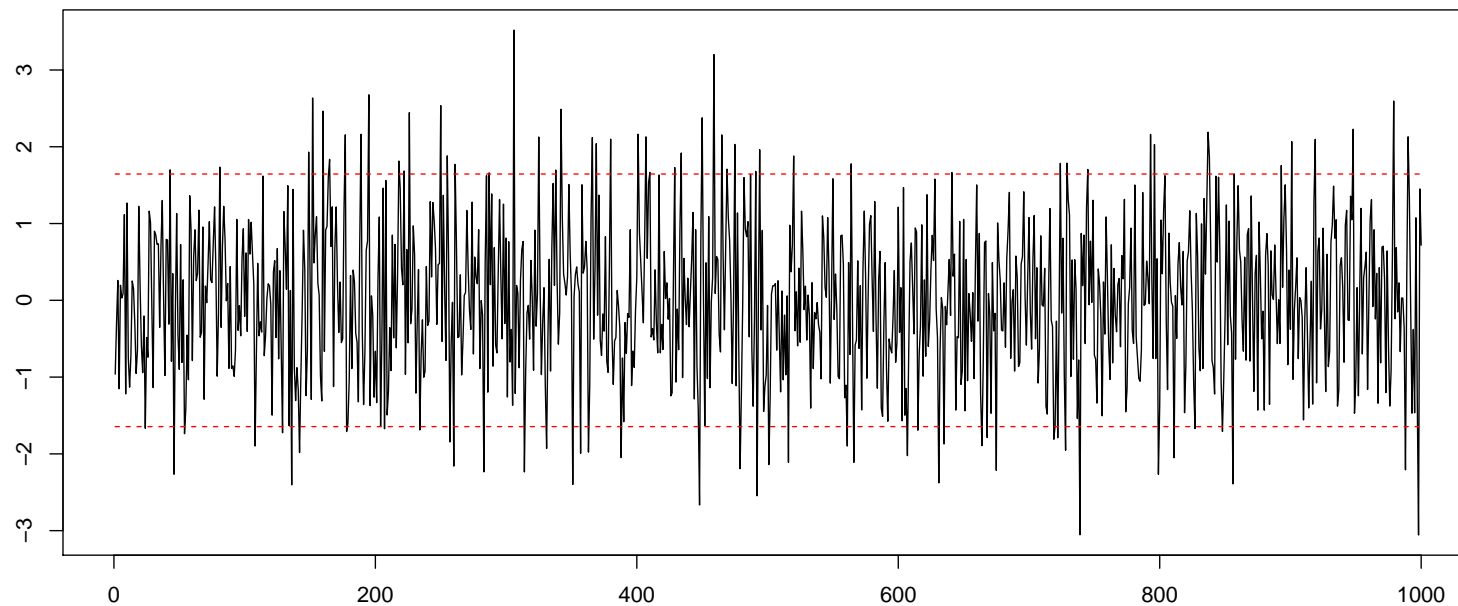


Stationnarité, série # 1



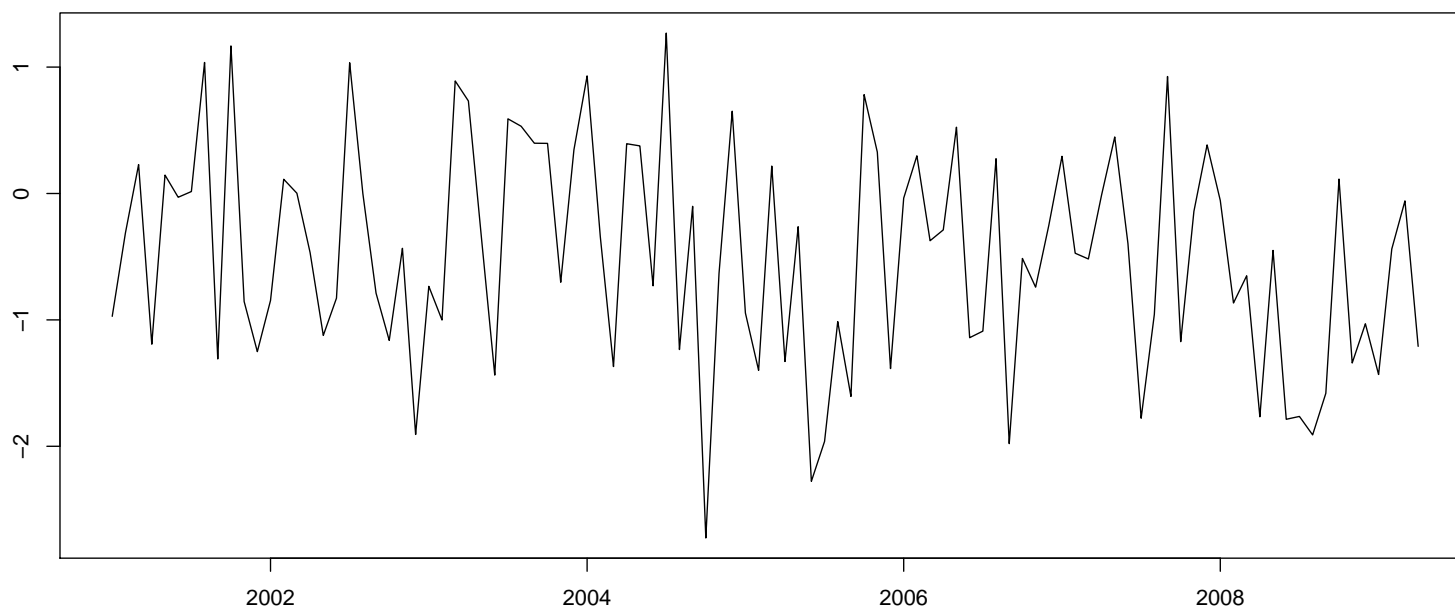
$$X_t = \varepsilon_t \text{ où } \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

Stationnarité, série # 1

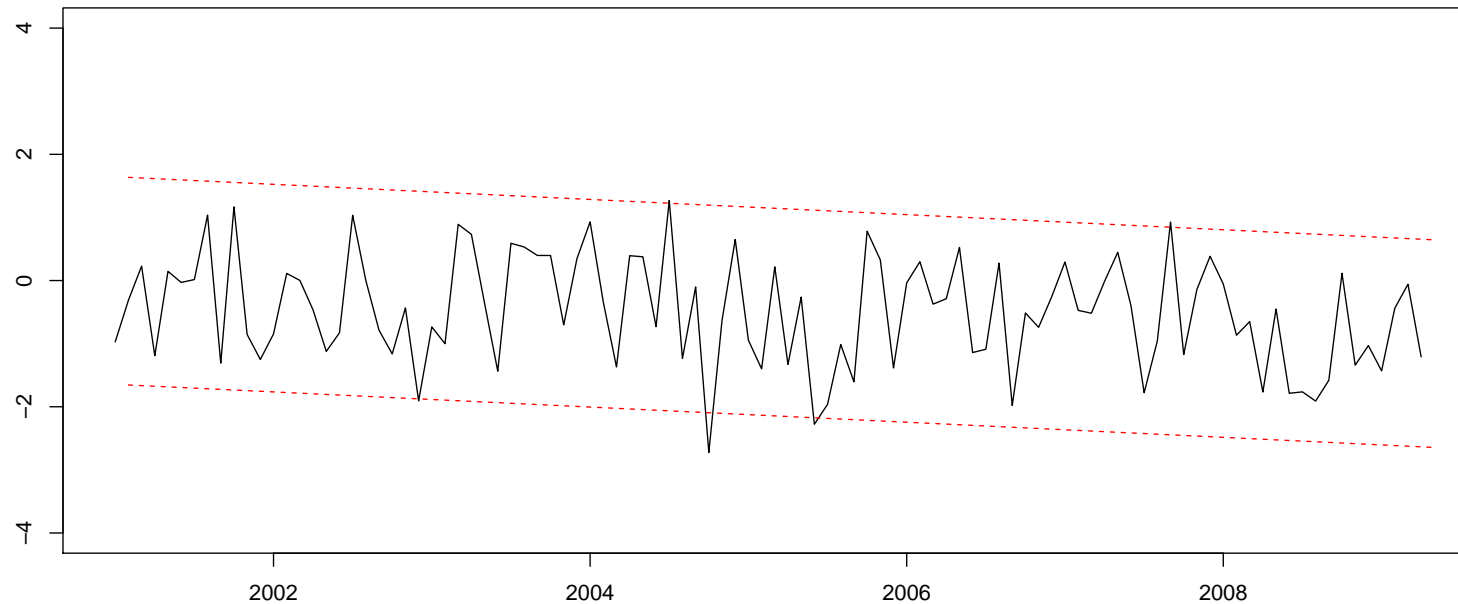


$$X_t = \varepsilon_t \text{ où } \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

Stationnarité, série # 2

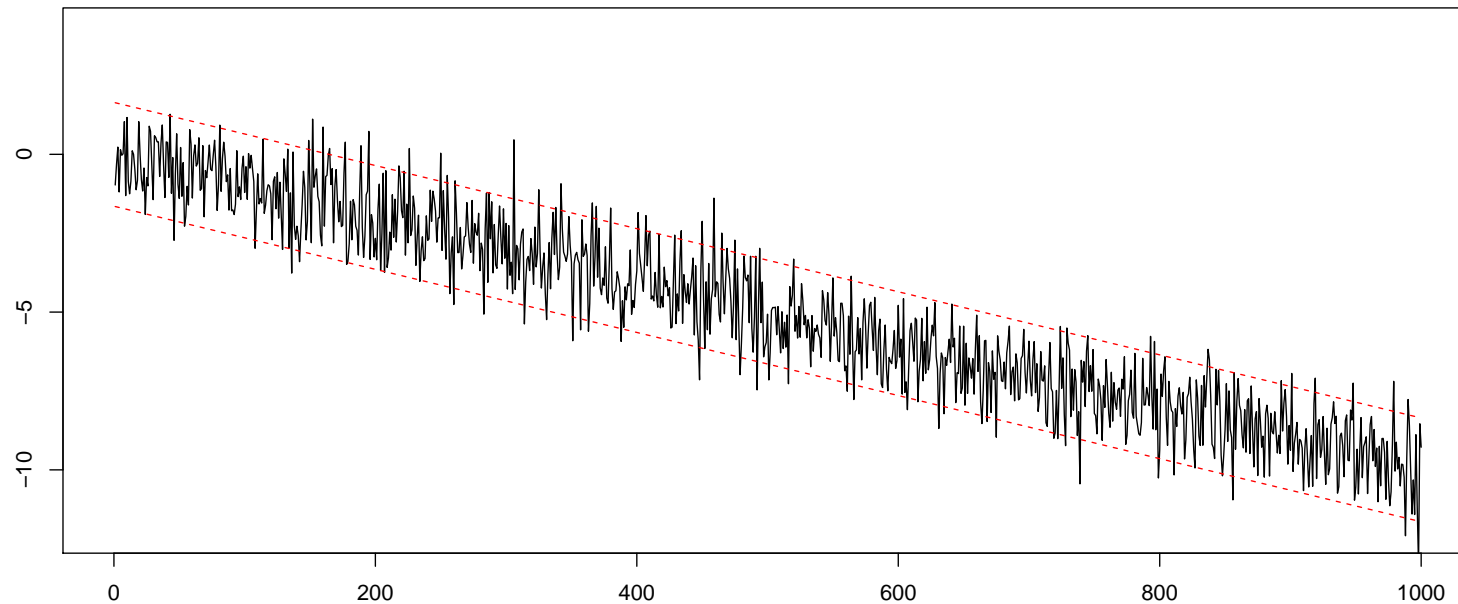


Stationnarité, série # 2



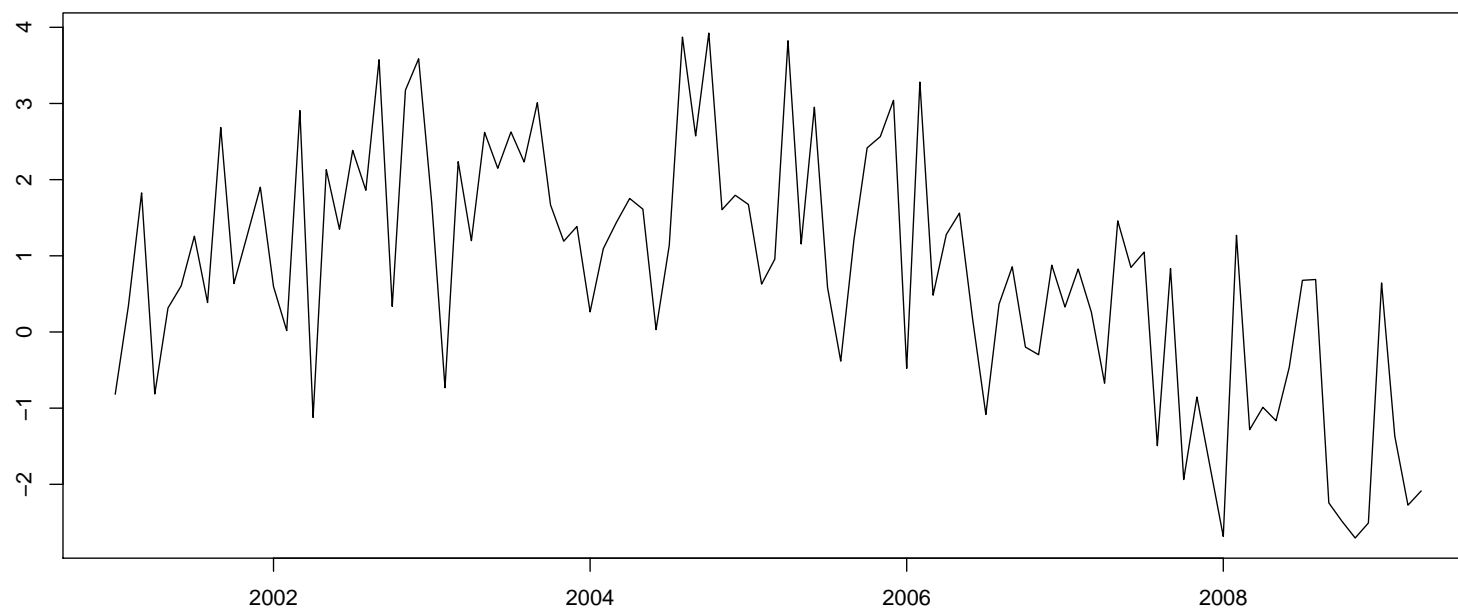
$$X_t = a + b \cdot t + \varepsilon_t \text{ où } \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

Stationnarité, série # 2

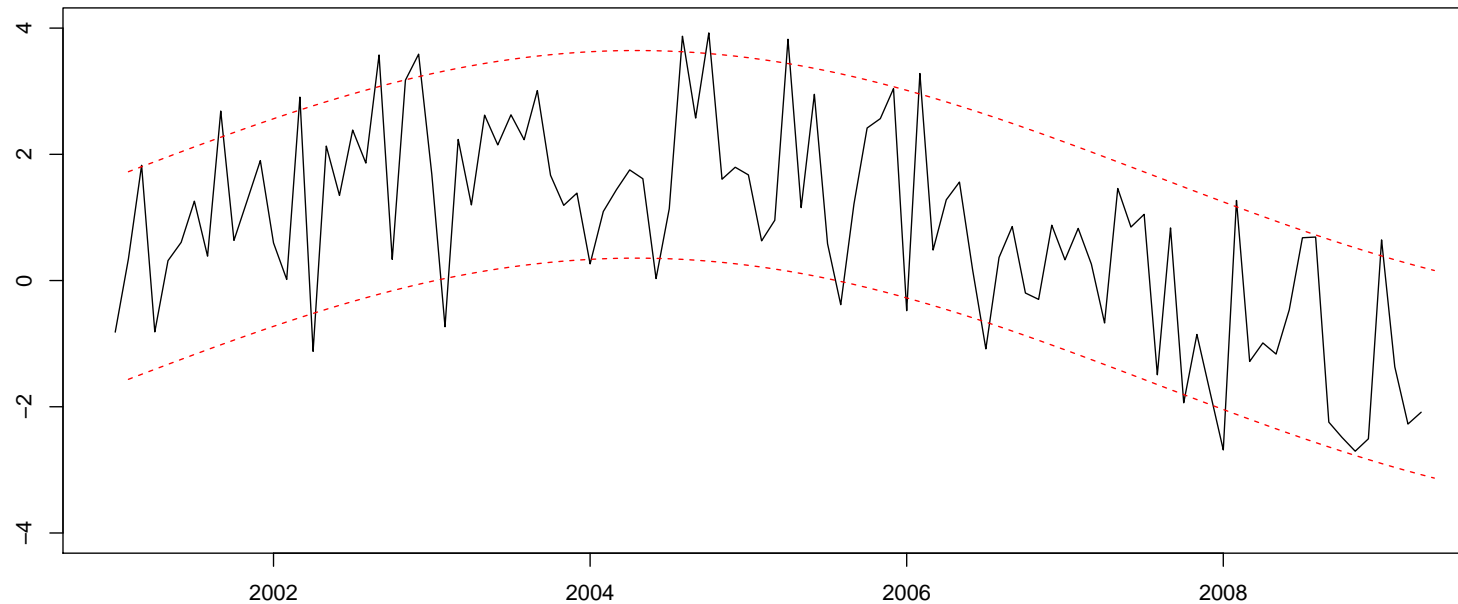


$$X_t = a + b \cdot t + \varepsilon_t \text{ où } \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

Stationnarité, série # 3

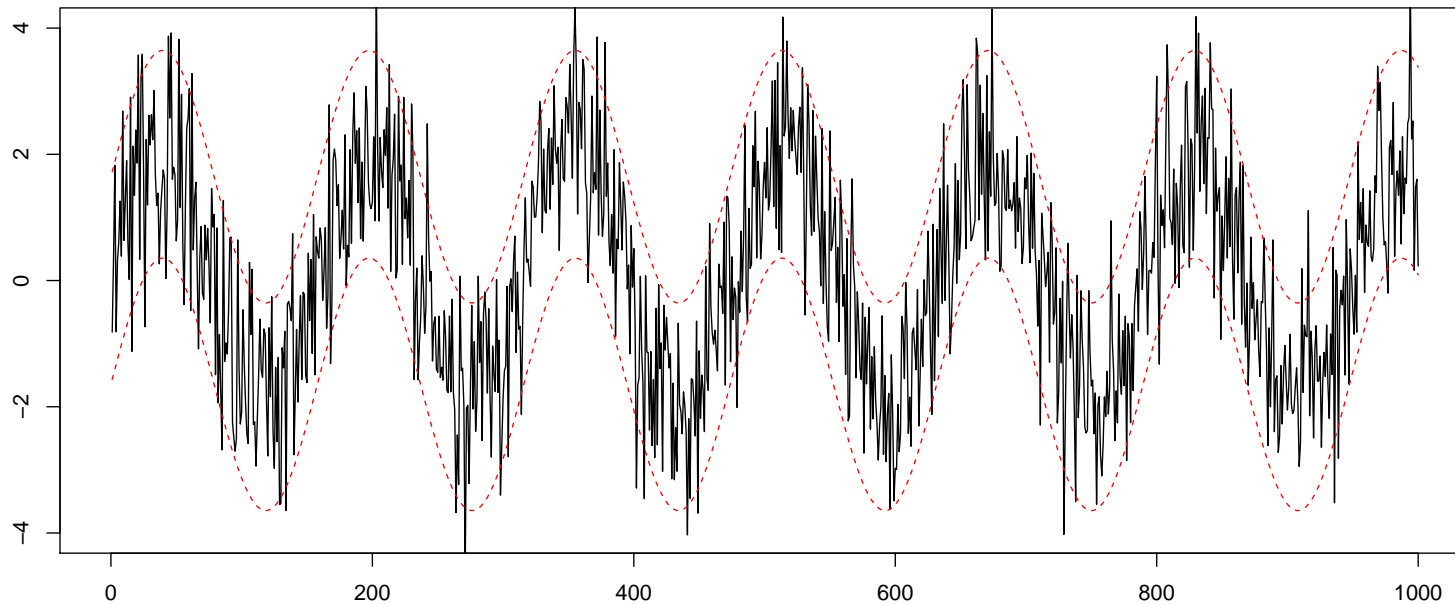


Stationnarité, série # 3



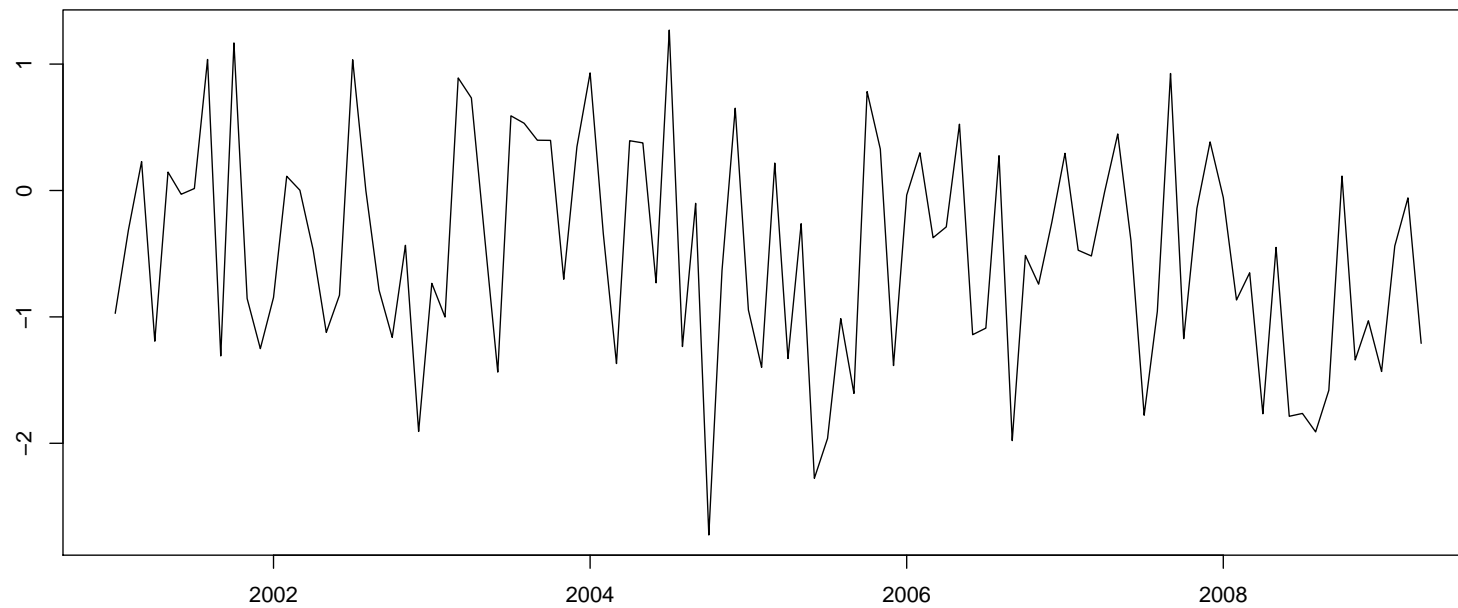
$$X_t = S(t) + \varepsilon_t \text{ où } \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

Stationnarité, série # 3

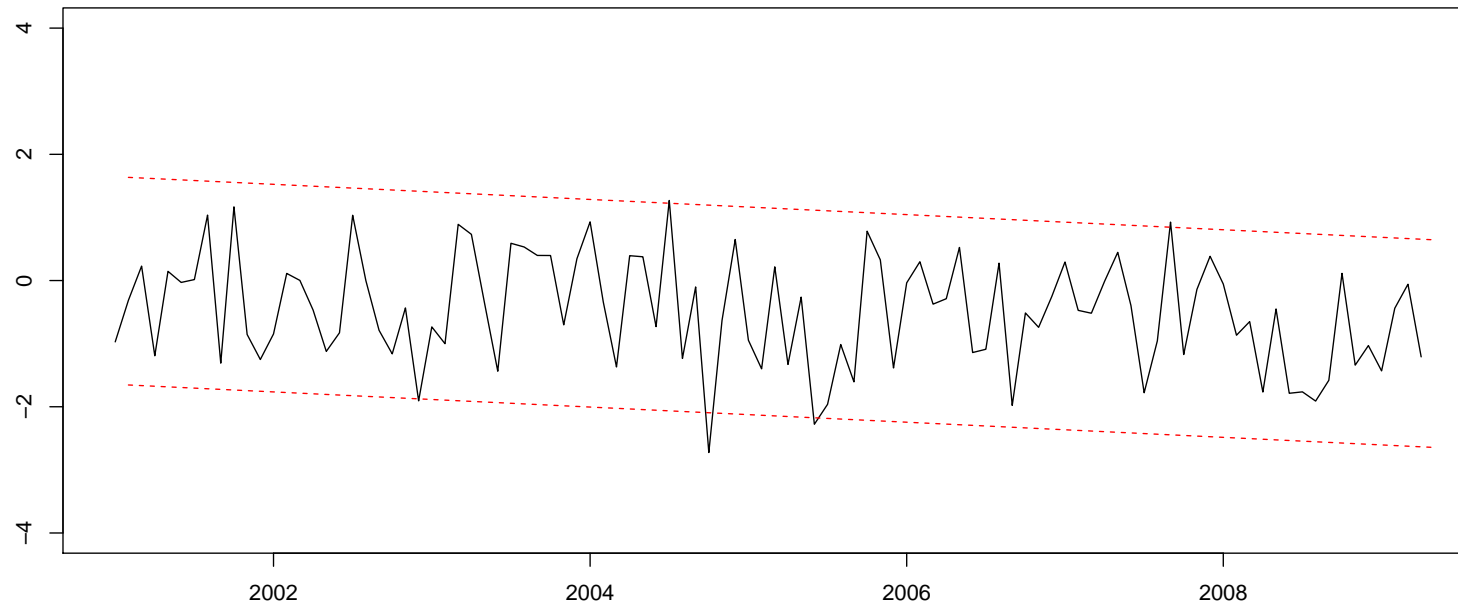


$$X_t = S(t) + \varepsilon_t \text{ où } \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

Stationnarité, série # 4

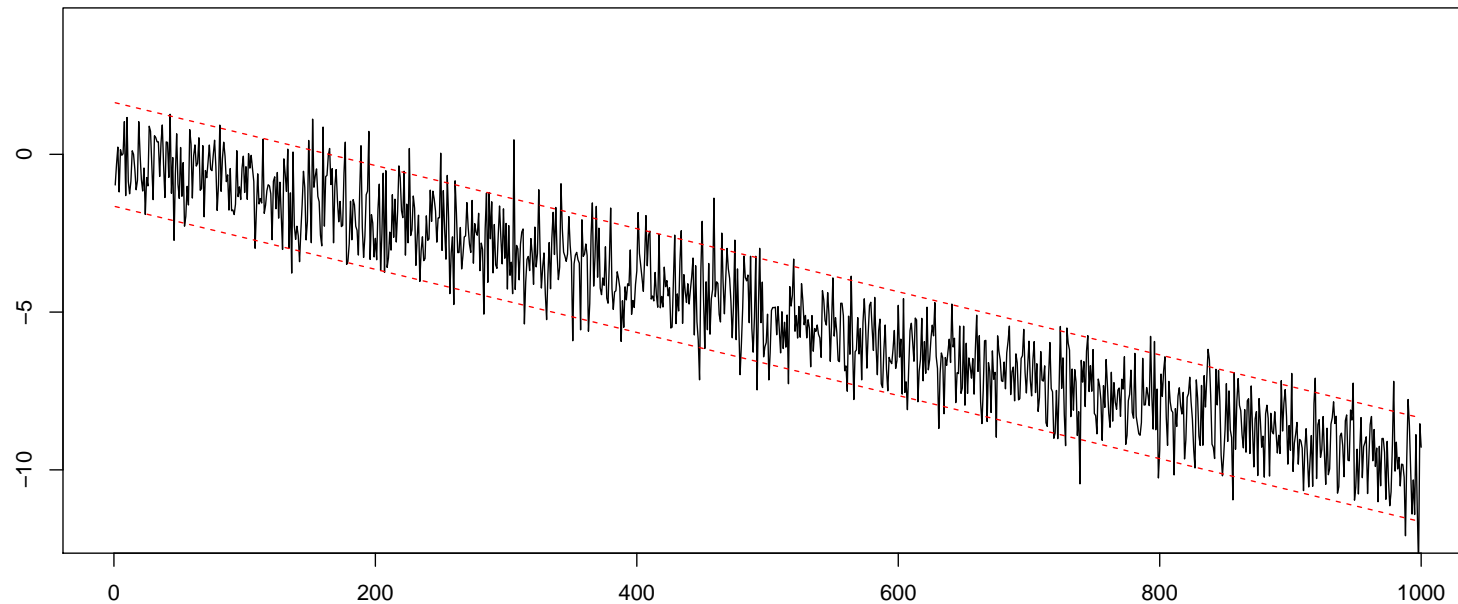


Stationnarité, série # 4



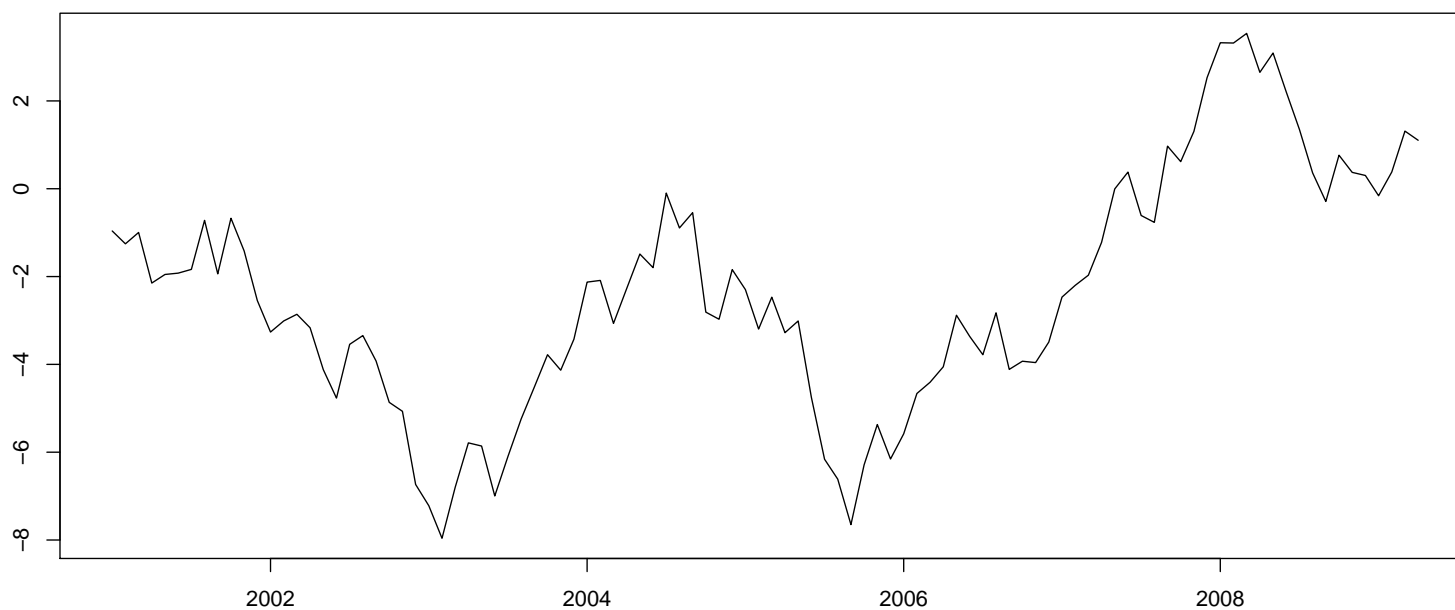
$$X_t = \varphi(t) \cdot \varepsilon_t \text{ où } \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

Stationnarité, série # 4

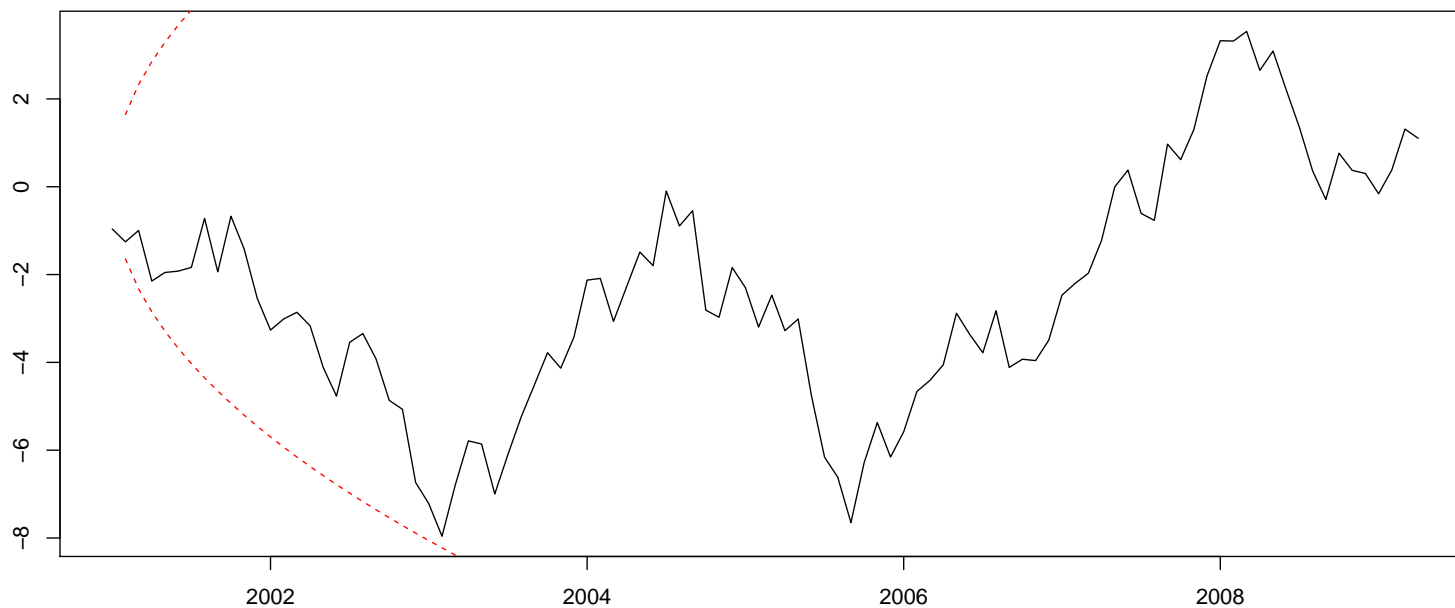


$$X_t = \varphi(t) \cdot \varepsilon_t \text{ où } \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

Stationnarité, série # 5

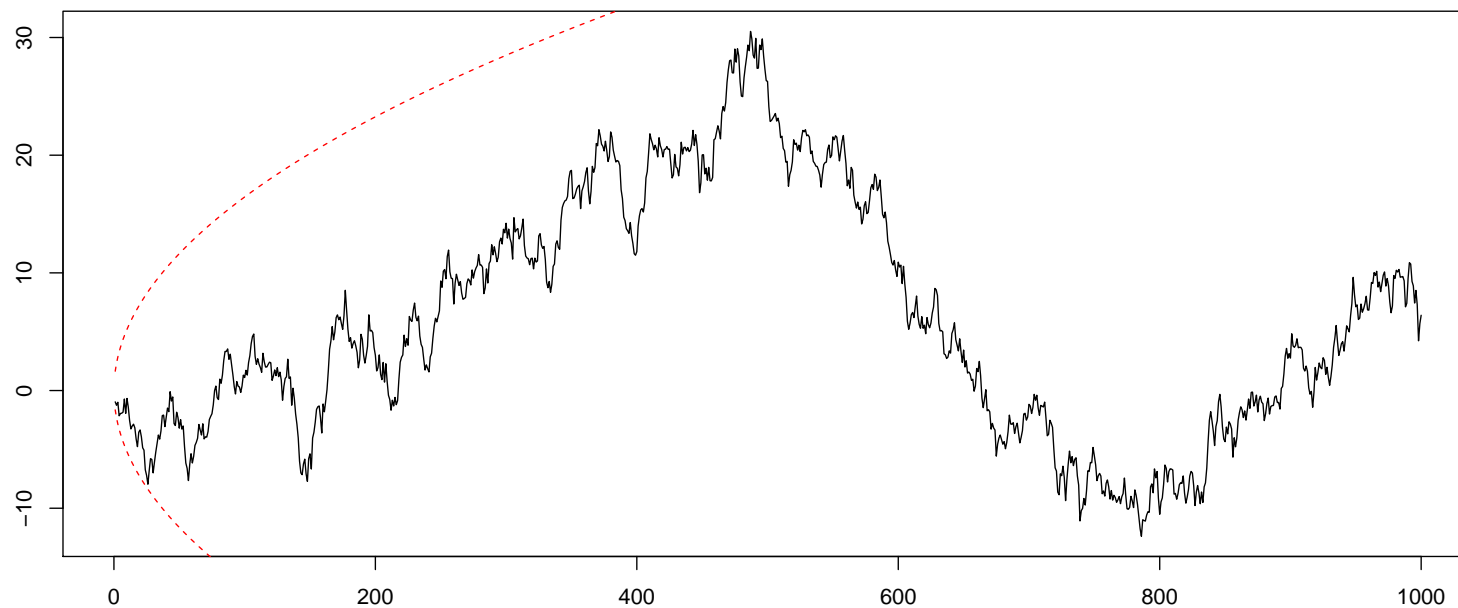


Stationnarité, série # 5



$$X_t = X_{t-1} + \varepsilon_t = X_0 + \sum_{t=1}^T \varepsilon_t \text{ où } \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

Stationnarité, série # 5



$$X_t = X_t + X_{t-1} + \varepsilon_t = X_0 + \sum_{t=1}^T \varepsilon_t \text{ où } \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

Autocovariance et autocorrélation

Soit (X_t) une série stationnaire. Alors pour tout t et pour tout h ,
 $\text{cov}(X_t, X_{t+h}) = \mathbb{E}([X_t - \mu][X_{t+h} - \mu]) = \gamma(h)$, indépendante de t .

La fonction $\gamma(\cdot)$ sera appelée **fonction d'autocovariance**

On notera que $\text{Var}(X_t) = \gamma(0)$.

On peut poser

$$\rho(h) = \text{cor}(X_t, X_{t+h}) = \frac{\gamma(h)}{\gamma(0)}.$$

La fonction $\rho(\cdot)$ sera appelée **fonction d'autocorrélation**

Matrice d'autocorrélations

On appellera **matrice d'autocorrélation** du vecteur $(X_t, X_{t-1}, \dots, X_{t-h+1})$

$$\mathcal{R}(h) = \begin{bmatrix} 1 & \rho(1) & \rho(2) & & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & & \rho(h-2) \\ \rho(2) & \rho(1) & 1 & \ddots & \rho(h-3) \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & 1 & \rho(1) \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & & \rho(1) & 1 \end{bmatrix}$$

i.e.

$$\mathcal{R}(h) = \begin{bmatrix} \mathcal{R}(h-1) & \begin{bmatrix} \rho(h-1) \\ \vdots \\ \rho(1) \\ 1 \end{bmatrix} \\ \begin{bmatrix} \rho(h-1) & \cdots & \rho(1) \end{bmatrix} & 1 \end{bmatrix}$$

Notons que $\det \mathcal{R}(h) \geq 0$ pour tout $h \in \mathbb{N} \setminus \{0\}$

Relations sur les fonctions d'autocorrélation

$\det \mathcal{R}(2) \geq 0$ implique que pour $(\rho(1), \rho(2))$,

$$[1 - \rho(2)] [1 + \rho(2) - 2\rho(1)^2] \geq 0$$

en plus des conditions $-1 \leq \rho(1), \rho(2) \leq 1$, i.e.

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

Relations sur les fonctions d'autocorrélation

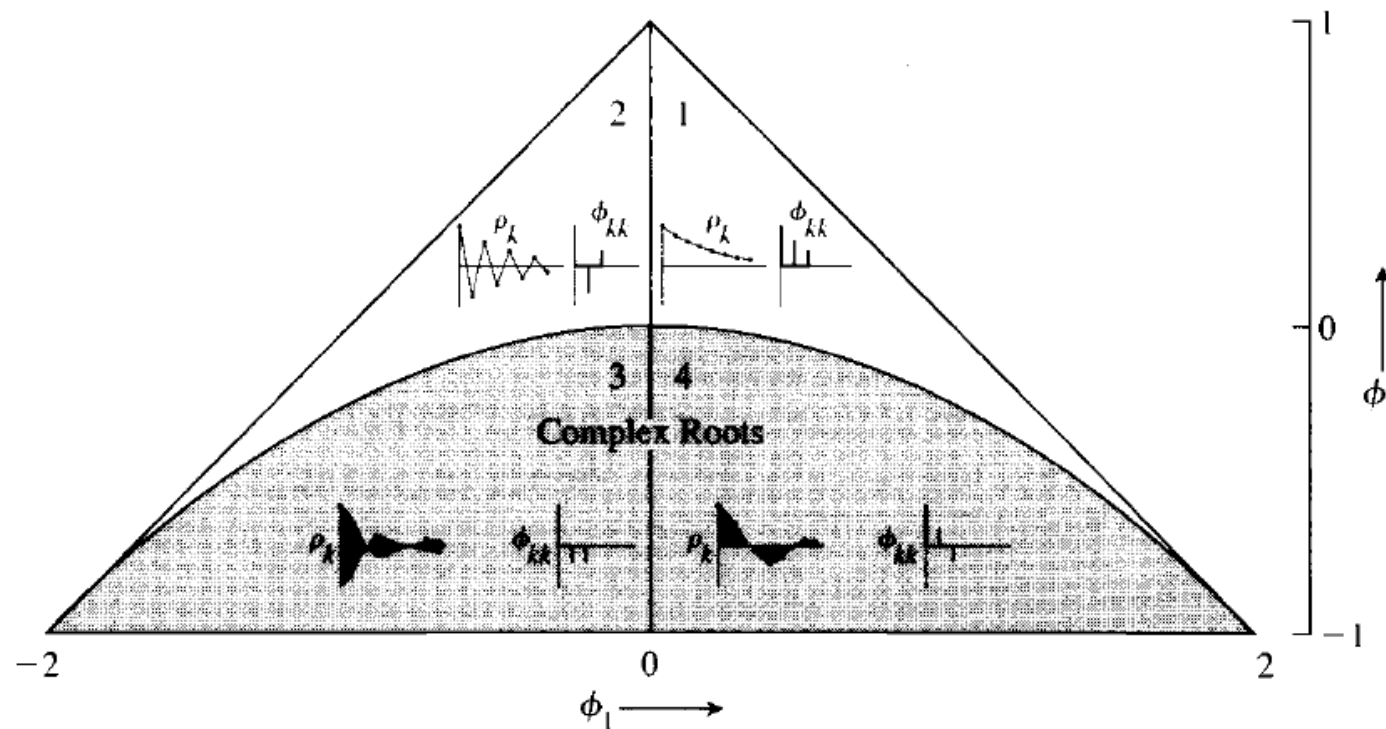


Figure 3.2 Typical autocorrelation and partial autocorrelation functions ρ_k and ϕ_{kk} for various stationary AR(2) models. (From [183].)

(via [Box, Jenkins & Reinsel \(1994\)](#)).

Estimation de ρ et de γ

Considérons un ensemble d'observations X_1, \dots, X_T .

La **moyenne empirique** est donnée par

$$\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t.$$

La **fonction d'autocovariance empirique** est donnée par

$$\hat{\gamma}_T(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (X_t - \bar{X}_T) (X_{t+h} - \bar{X}_T),$$

et la **fonction d'autocorrélation empirique** est donnée par

$$\hat{\rho}_T(h) = \frac{\hat{\gamma}_T(h)}{\hat{\gamma}_T(0)}.$$

Estimation de ρ et de γ

Proposition 1. *Si (X_t) est un processus linéaire, au sens où il satisfait $X_t = \sum_{j \in \mathbb{Z}} \phi_j \varepsilon_{t-j}$ où (ε_t) est une suite de variables i.i.d. centrées, telle que $\mathbb{E}(\varepsilon_t^4) = \eta \mathbb{E}(\varepsilon_t^2)^2 < +\infty$, où les ϕ_j définissent une série absolument convergente, et où η est une constante positive, alors, on a la formule dite de Bartlett,*

$$\lim_{T \rightarrow \infty} T \text{cov}(\hat{\gamma}_T(h), \hat{\gamma}_T(k)) = \eta \gamma(h) \gamma(k) + \sum_{i=-\infty}^{+\infty} \gamma(i) \gamma(i+k-h) + \gamma(i+k) \gamma(i-h)$$

Estimation de ρ et de γ

Proposition 2. *Si (X_t) est un processus linéaire, au sens où il satisfait $X_t = \sum_{j \in \mathbb{Z}} \phi_j \varepsilon_{t-j}$ où (ε_t) est une suite de variables i.i.d. centrées, telle que $\mathbb{E}(\varepsilon_t^4) = \eta \mathbb{E}(\varepsilon_t^2)^2 < +\infty$, et $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, et où les ϕ_j définissent une série absolument convergente, et où η est une constante positive, alors, on a, pour tout $p \geq 0$,*

$$\sqrt{n} \begin{pmatrix} \hat{\gamma}_T(0) \\ \vdots \\ \hat{\gamma}_T(p) \end{pmatrix} \rightarrow \mathcal{N} \left(\begin{pmatrix} \gamma(0) \\ \vdots \\ \gamma(p) \end{pmatrix}, V \right),$$

où V est la matrice de variance-covariance définie par

$$V = \left[\eta \gamma(h) \gamma(k) + \sum_{i=-\infty}^{+\infty} \gamma(i) \gamma(i+k-h) + \gamma(i+k) \gamma(i-h) \right]_{h,k=0,\dots,p}.$$

Les bruits blancs

Un processus (ε_t) est un **bruit blanc fort** si les variables ε_t sont **indépendantes**, **identiquement distribuées**, et centrées (i.e $\mathbb{E}(\varepsilon_t) = 0$).

Un processus (ε_t) est un **bruit blanc faible** si les variables ε_t sont **orthogonales** (ou non-corrélées), **de même variance** et centrées.

Les tests de bruits blancs (faibles)

(ε_t) est un bruit blanc (faible) s'il est centré, et si $\rho(h) = 0$ pour tout $h \neq 0$.

On peut utiliser la statistique de Box & Pierce (Q) ou la statistique de Ljung & Box (Q') définies - à partir des $r_i = \hat{\rho}(i)$ - par

$$Q(k) = n \sum_{i=1}^k r_i^2 \text{ et } Q'(k) = n(n+2) \sum_{i=1}^k \frac{r_i^2}{n-i},$$

qui sont à comparer aux quantiles du chi-deux à k degrés de liberté. Pour rappel,

$$H_0 : \rho(1) = \dots = \rho(k) = 0$$

Les tests de bruits blancs (faibles)

Soit (ε_t) un bruit blanc simulé

```
> E=rnorm(120)
> Box.test(E,lag=6,type="Box-Pierce")
```

Box-Pierce test

```
data: E
X-squared = 6.3143, df = 6, p-value = 0.3889
```

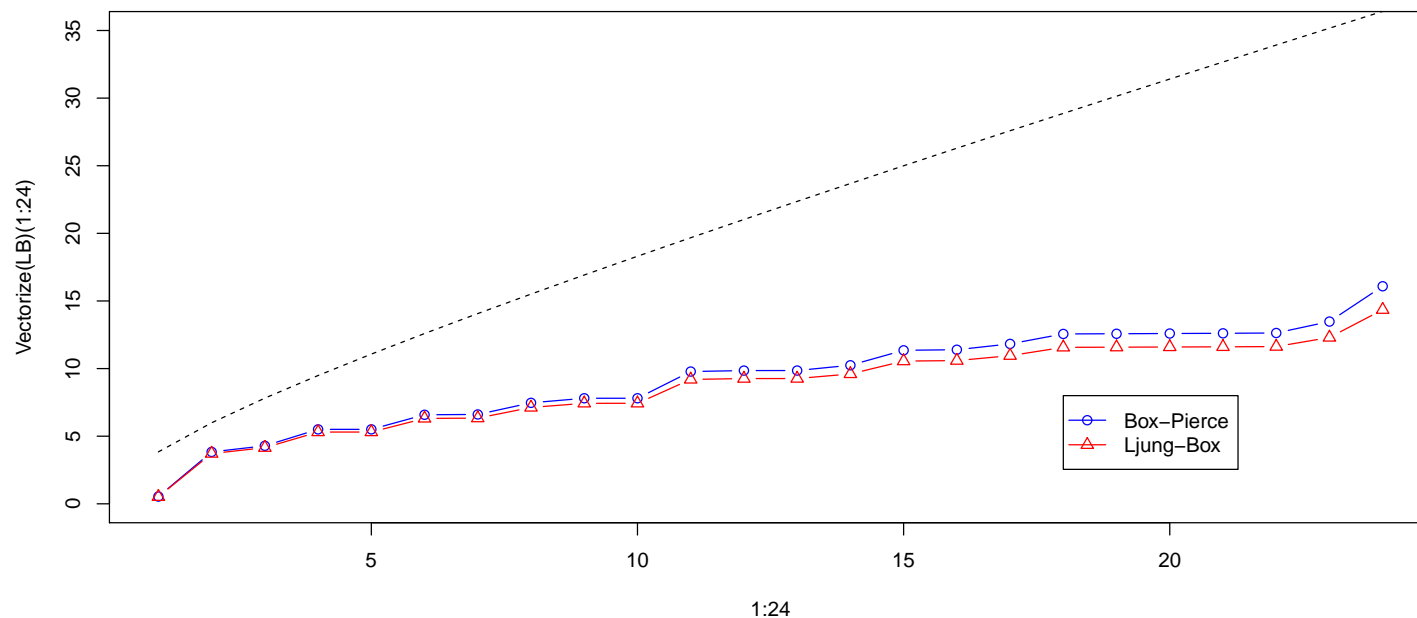
```
> Box.test(E,lag=6,type="Ljung-Box")
```

Box-Ljung test

```
data: E
X-squared = 6.5849, df = 6, p-value = 0.3609
```

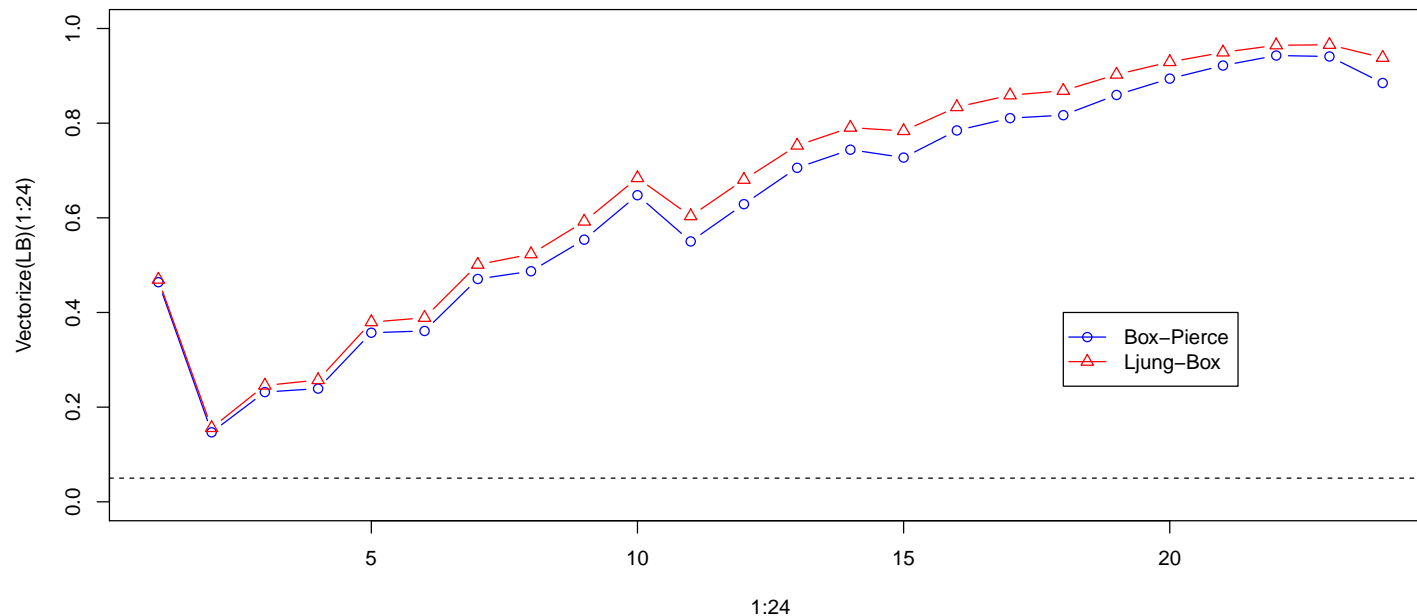
Les tests de bruits blancs (faibles)

```
> BP=function(h) Box.test(E,lag=h,type="Box-Pierce")$statistic  
> LB=function(h) Box.test(E,lag=h,type="Ljung-Box")$statistic  
> plot(1:24,Vectorize(LB)(1:24),type="b",col="blue",ylim=c(0,35))  
> points(1:24,Vectorize(BP)(1:24),type="b",col="red",pch=2)  
> lines(1:24,qchisq(.95,1:24),lty=2)
```



Les tests de bruits blancs (faibles)

```
> BP=function(h) Box.test(E,lag=h,type="Box-Pierce")$p.value  
> LB=function(h) Box.test(E,lag=h,type="Ljung-Box")$p.value  
> plot(1:24,Vectorize(LB)(1:24),ylim=c(0,1),type="b",col="blue")  
> points(1:24,Vectorize(BP)(1:24),ylim=c(0,1),type="b",col="red",pch=2)  
> abline(h=.05,lty=2)
```



Les tests de bruits blancs (faibles)

Soit (X_t) la moyenne mobile du bruit blanc, $X_t = \alpha\varepsilon_t + [1 - \alpha]\varepsilon_{t-1}$

```
> X=.8*E[2:120]+.2*E[1:119]
> Box.test(X,lag=6,type="Box-Pierce")
```

Box-Pierce test

```
data: X
X-squared = 11.3128, df = 6, p-value = 0.07918
```

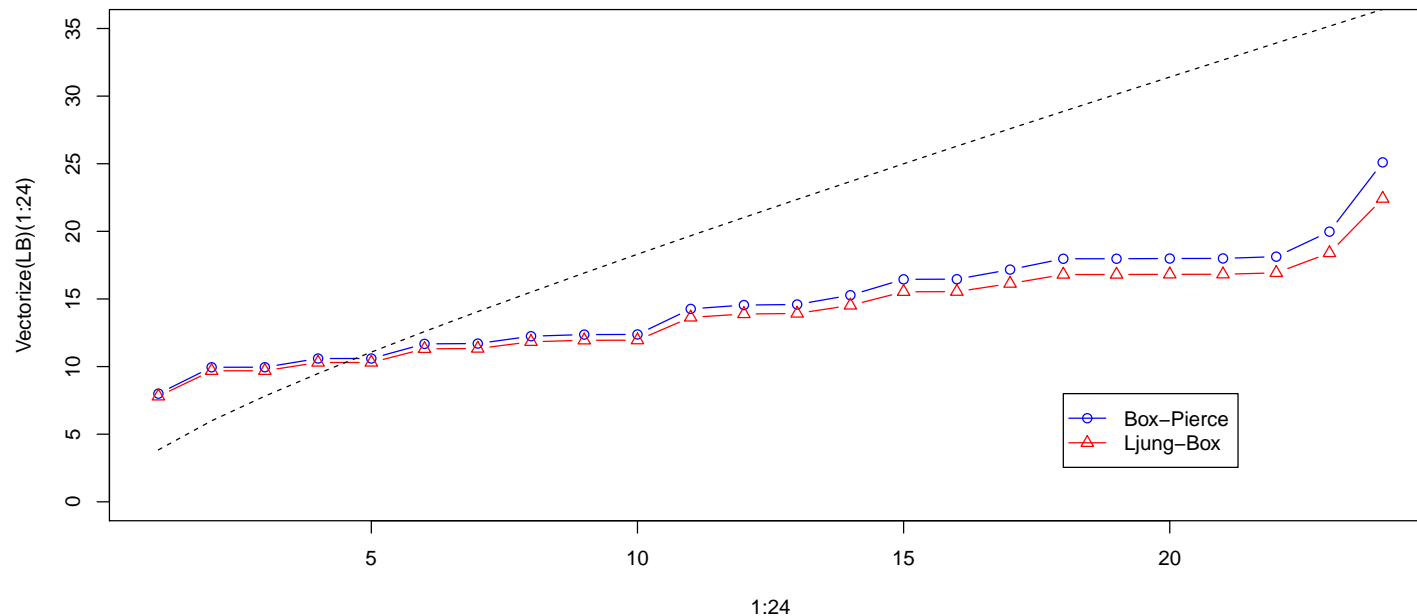
```
> Box.test(X,lag=6,type="Ljung-Box")
```

Box-Ljung test

```
data: X
X-squared = 11.6793, df = 6, p-value = 0.06952
```

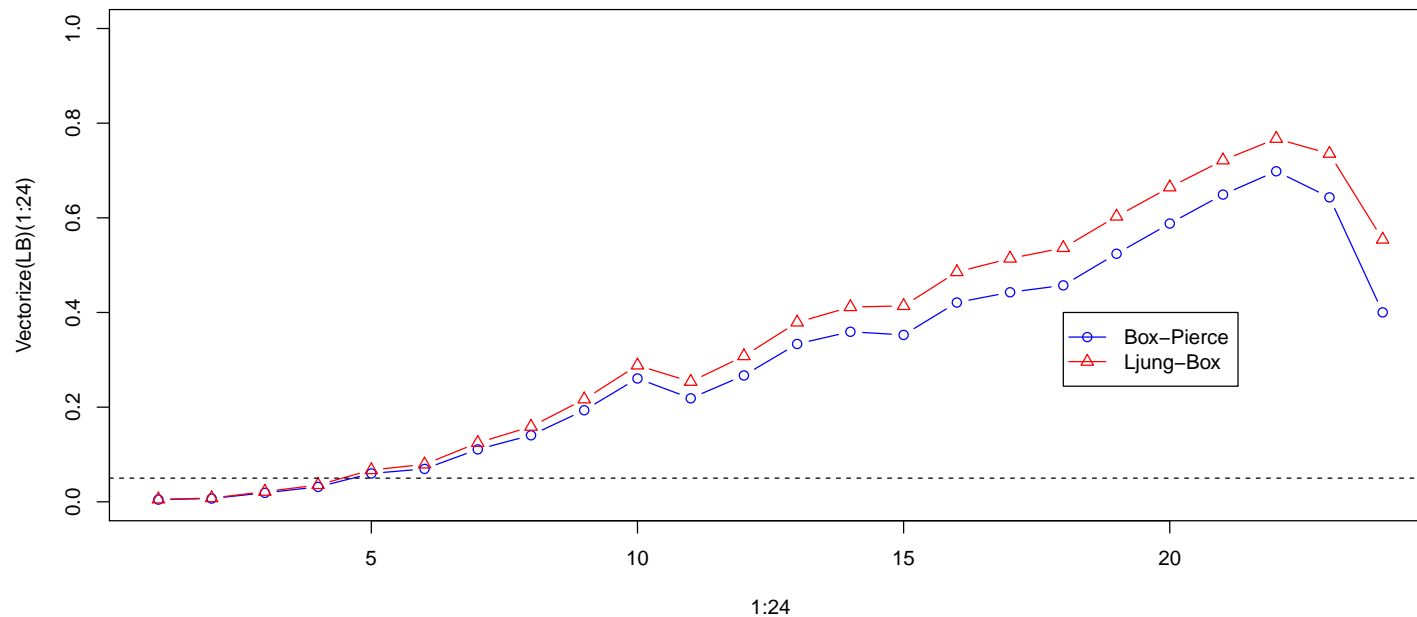
Les tests de bruits blancs (faibles)

```
> BP=function(h) Box.test(X,lag=h,type="Box-Pierce")$statistic  
> LB=function(h) Box.test(X,lag=h,type="Ljung-Box")$statistic  
> plot(1:24,Vectorize(LB)(1:24),type="b",col="blue",ylim=c(0,35))  
> points(1:24,Vectorize(BP)(1:24),type="b",col="red",pch=2)  
> lines(1:24,qchisq(.95,1:24),lty=2)
```



Les tests de bruits blancs (faibles)

```
> BP=function(h) Box.test(X,lag=h,type="Box-Pierce")$p.value  
> LB=function(h) Box.test(X,lag=h,type="Ljung-Box")$p.value  
> plot(1:24,Vectorize(LB)(1:24),ylim=c(0,1),type="b",col="blue")  
> points(1:24,Vectorize(BP)(1:24),ylim=c(0,1),type="b",col="red",pch=2)  
> abline(h=.05,lty=2)
```



Les processus Gaussiens

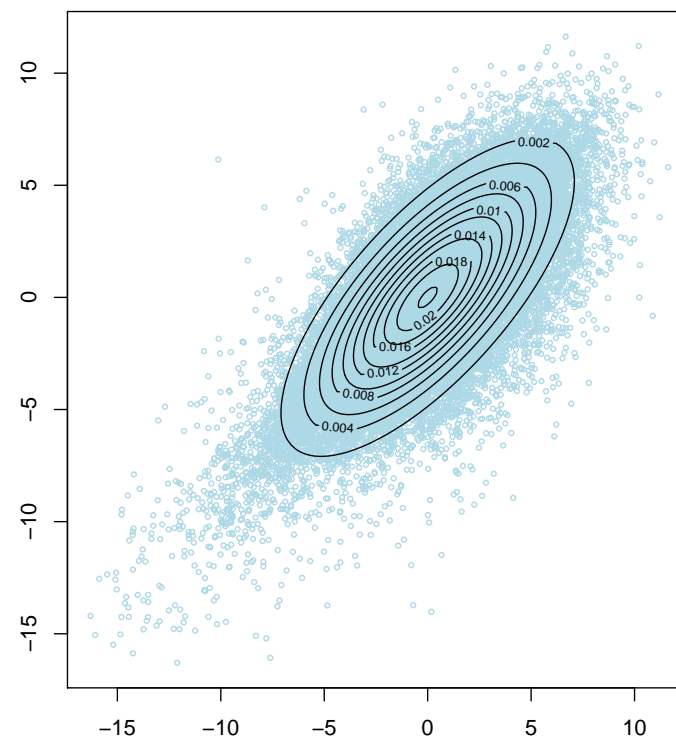
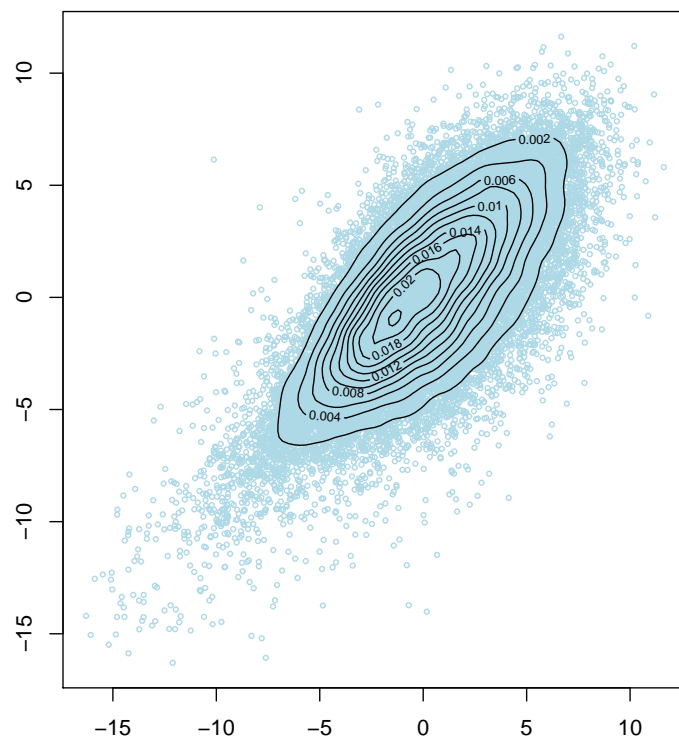
Le vecteur $\mathbf{X} = (X_1, \dots, X_d)$ est un **vecteur gaussien** si toute combinaison linéaire des X_i est une variable gaussienne, i.e. pour tout $\mathbf{a} \in \mathbb{R}^d$, $\mathbf{a}'\mathbf{X}$ est une variable gaussienne. Sa densité s'écrit alors

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \right),$$

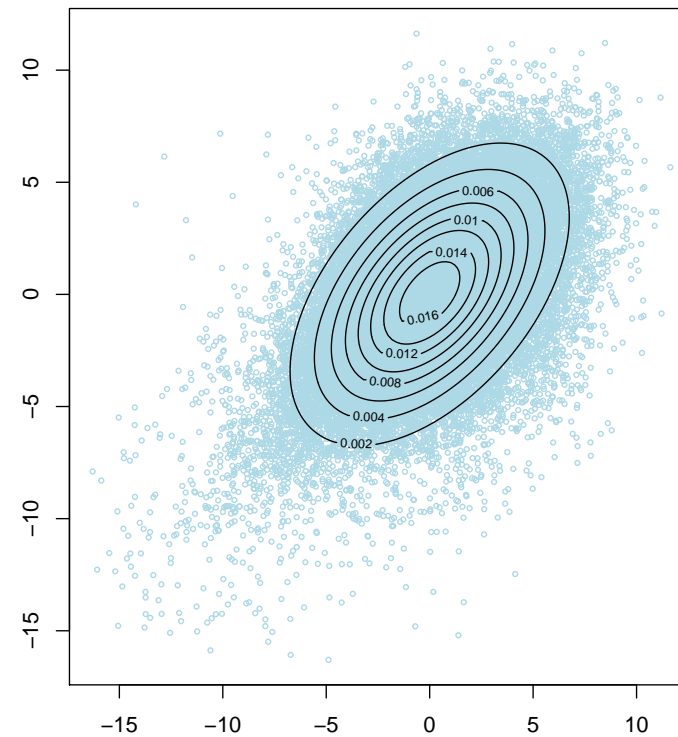
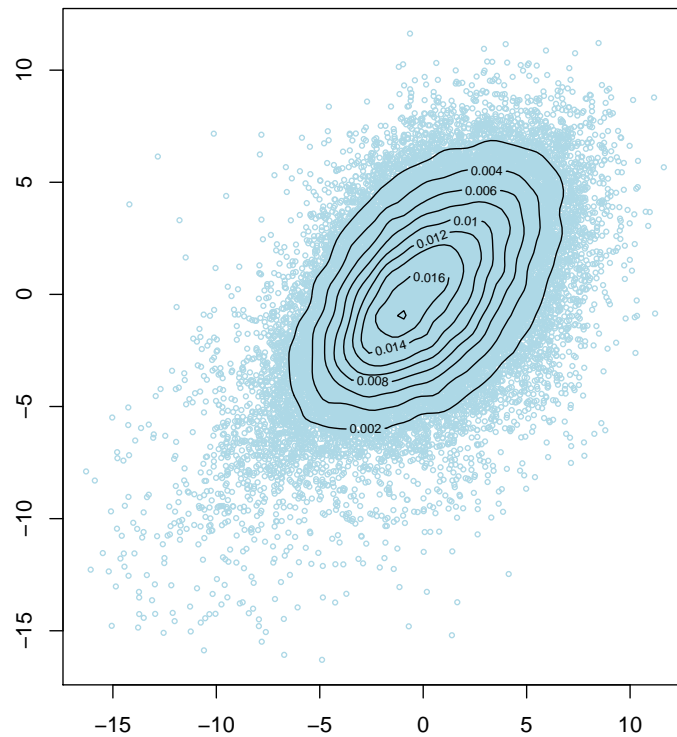
où $\mu \in \mathbb{R}^d$ et Σ est une matrice hermitienne positive $d \times d$.

Le processus (X_t) est un **processus gaussien** si tout système fini extrait est un vecteur aléatoire gaussien, i.e. pour tout n , pour tout t_1, \dots, t_n , $(X_{t_1}, \dots, X_{t_n})$ est un vecteur gaussien.

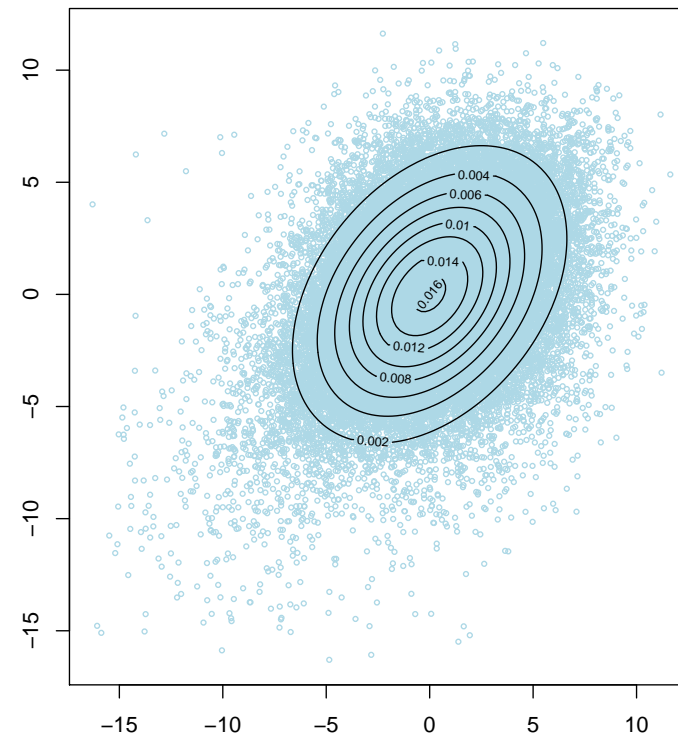
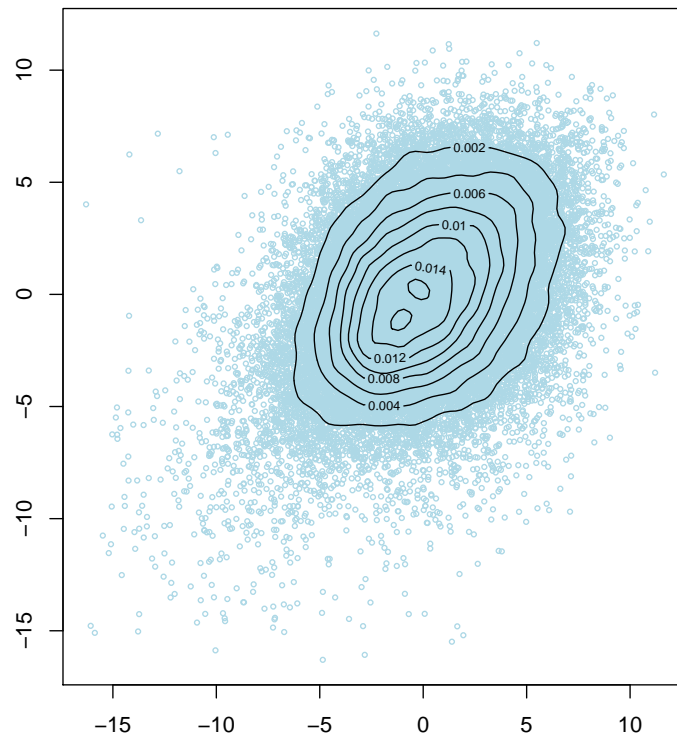
Température journalière (X_{t-1}, X_t)



Température journalière (X_{t-2}, X_t)



Température journalière (X_{t-3}, X_t)



La propriété de Markov

(X_t) est un **Markovien** à l'ordre k si

$$\mathcal{L}(X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots) = \mathcal{L}(X_t | X_{t-1}, \dots, X_{t-k}),$$

qui peut se réécrire, à l'ordre 1,

$$(X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots) \stackrel{\mathcal{L}}{=} (X_t | X_{t-1}).$$

Il est possible de montrer que cette relation est équivalente à

$$X_t = g(X_{t-1}, \varepsilon_t), \text{ où } (\varepsilon_t) \text{ est un bruit blanc.}$$

Par exemple, les processus $AR(1)$: $X_t = \alpha + \beta X_{t-1} + \varepsilon_t$, où (ε_t) est un bruit blanc indépendant du passé du processus, est markovien.

La propriété de martingale

Soit $\{\mathcal{F}_t, t \in \mathbb{N}\}$ la filtration naturelle associée au processus (X_t) .

(X_t) est une **martingale** si et seulement si, pour tout t ,

$$\mathbb{E}(X_{t+1}|\mathcal{F}_t) = \mathbb{E}(X_{t+1}|\underline{X}_t) = X_t$$

presque sûrement.

L'opérateur retard L

On définit l'opérateur retard L par $L : X_t \mapsto L(X_t) = LX_t = X_{t-1}$ et l'opérateur avance F par $F : X_t \mapsto F(X_t) = FX_t = X_{t+1}$. On notera alors

$$L^p = \underbrace{L \circ L \circ \dots \circ L}_{p \text{ fois}} \text{ où } p \in \mathbb{N},$$

avec la convention $L^0 = \mathbb{I}$ et $L^{-1} = F$. Et de façon analogue, $L^{-p} = F^p$ pour $p \in \mathbb{N}$.

Aussi, $L^p(X_t) = X_{t-p}$

L'opérateur retard L

On peut définir un **polynôme d'opérateurs retards** de la manière suivante : soit P un polynôme réel (de degré k),

$$P(z) = \alpha_0 + \alpha_1 z + \cdots + \alpha_k z^k.$$

L'opérateur $P(L)$ associe à une série (X_t) la série (Y_t) définie

$$Y_t = P(L)(X_t) = \alpha_0 X_t + \alpha_1 X_{t-1} + \cdots + \alpha_k X_{t-k}$$

Par exemple, on peut considérer une **moyenne glissante**

$$Y_t = \frac{1}{12} [X_t + X_{t-1} + \cdots + X_{t-12}].$$

Inversibilité de $P(L)$

Considérons $P(z) = 1 - \lambda z$ (de racine λ^{-1}).

Proposition 3. *Si $\|\lambda\| < 1$ (ou $\|\lambda^{-1}\| > 1$) alors $1 - \lambda L$ est inversible, et de plus,*

$$(1 - \lambda L)^{-1} = \sum_{k=0}^{\infty} \lambda^k L^k.$$

Si $\|\lambda\| = 1$, alors $1 - \lambda L$ n'est pas inversible.

Remark Tout polynôme $P(L) = 1 + a_1 L + \dots + a_n L^n$ peut s'écrire

$$P(z) = a_n (z - z_1) (z - z_2) \dots (z - z_n),$$

correspondant à la décomposition en éléments simples (z_i étant les *racines* du polynôme).

Inversibilité de $P(L)$

On peut écrire

$$P(L) = \prod_{i=1}^n (1 - \lambda_i L) \quad \text{où } \lambda_i = \frac{1}{z_i}$$

Si pour tout i , $|\lambda_i| \neq 1$, alors $P(L)$ est inversible, et

$$P(L) = \prod (1 - \lambda_i L) = \underbrace{\prod_{|\lambda_i| < 1} (1 - \lambda_i L)}_{P_1(L)} \underbrace{\prod_{|\lambda_i| > 1} \left(1 - \frac{1}{\lambda_i} L\right)}_{P_2(L)} \underbrace{\prod_{|\lambda_i| > 1} (-\lambda_i L)}_{P_3(L)},$$

Les autocorrélations partielles

On suppose ici que le processus (X_t) est stationnaire, et centré, $\mathbb{E}(X_t) = 0$ pour tout t .

Pour une série stationnaire (X_t) , on définit la fonction d'autocorrélation partielle $h \mapsto \psi(h)$ par

$$\psi(h) = \text{corr}(\hat{X}_t, \hat{X}_{t-h}),$$

où

$$\begin{cases} \hat{X}_{t-h} = X_{t-h} - EL(X_{t-h} | X_{t-1}, \dots, X_{t-h+1}) \\ \hat{X}_t = X_t - EL(X_t | X_{t-1}, \dots, X_{t-h+1}). \end{cases}$$

Il est équivalent de connaître ρ et ψ .

Les autocorrélations partielles

$$\mathcal{R}(h) = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \rho(h-3) & \rho(h-2) & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \rho(h-4) & \rho(h-3) & \rho(h-2) \\ \rho(2) & \rho(1) & 1 & \ddots & \rho(h-5) & \rho(h-4) & \rho(h-3) \\ & & \ddots & \ddots & \ddots & & \\ \rho(h-3) & \rho(h-4) & \rho(h-5) & \ddots & 1 & \rho(1) & \rho(2) \\ \rho(h-2) & \rho(h-3) & \rho(h-4) & & \rho(1) & 1 & \rho(1) \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & & \rho(2) & \rho(1) & 1 \end{bmatrix}$$

et on introduit de façon analogue la matrice $\mathcal{R}^*(h)$ obtenue...

Les autocorrélations partielles

...en remplaçant la dernière colonne de $\mathcal{R}(h)$ par le vecteur $[\rho(1), \dots, \rho(h)]'$,

$$\mathcal{R}^*(h) = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \rho(h-3) & \rho(h-2) & \rho(1) \\ \rho(1) & 1 & \rho(1) & \rho(h-4) & \rho(h-3) & \rho(2) \\ \rho(2) & \rho(1) & 1 & \ddots & \rho(h-5) & \rho(h-4) & \rho(3) \\ & & \ddots & \ddots & \ddots & & \\ \rho(h-3) & \rho(h-4) & \rho(h-5) & \ddots & 1 & \rho(1) & \rho(h-2) \\ \rho(h-2) & \rho(h-3) & \rho(h-4) & & \rho(1) & 1 & \rho(h-1) \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & & \rho(2) & \rho(1) & \rho(h) \end{bmatrix}$$

Il est alors possible de montrer que

$$\psi(h) = \frac{|\mathcal{R}^*(h)|}{|\mathcal{R}(h)|} \text{ pour tout } h.$$

Les autocorrélations partielles

On peut obtenir cette fonction via

```
> X=rnorm(100)
> as.vector(pacf(X))
```

Partial autocorrelations of series 'X', by lag

1	2	3	4	5	6	7	8	9
-0.004	-0.027	-0.108	-0.116	-0.105	-0.153	0.023	-0.002	-0.025

Remark pour une série stationnaire (X_t)

$$\psi_X(1) = \rho_X(1) \text{ et } \psi_X(2) = \frac{[\rho_X(2) - \rho_X(1)^2]}{[1 - \rho_X(1)^2]}$$