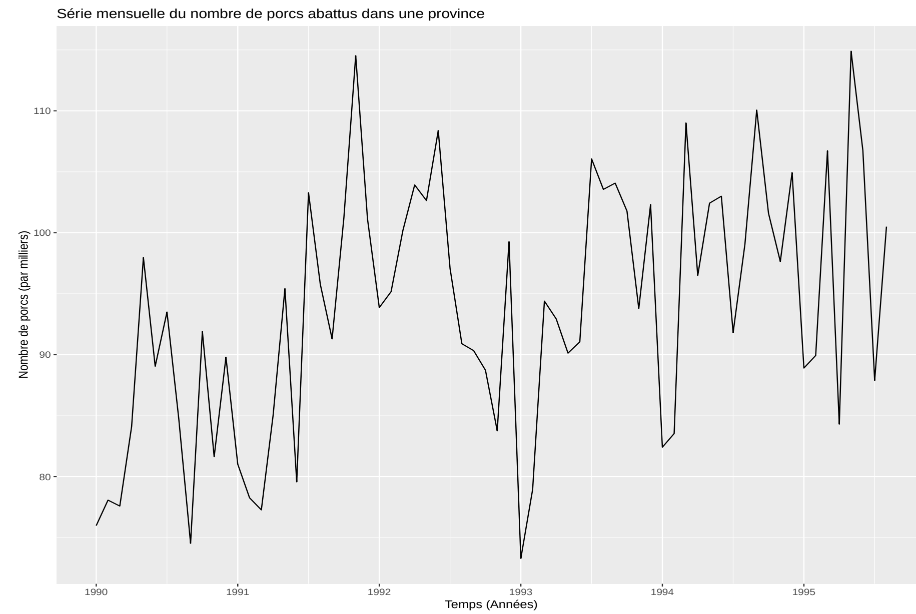
R et la prévision des séries temporelles Chapitre 2 - Quelques outils pour le prévisionniste

Michel CARBON

Université Laval de Québec

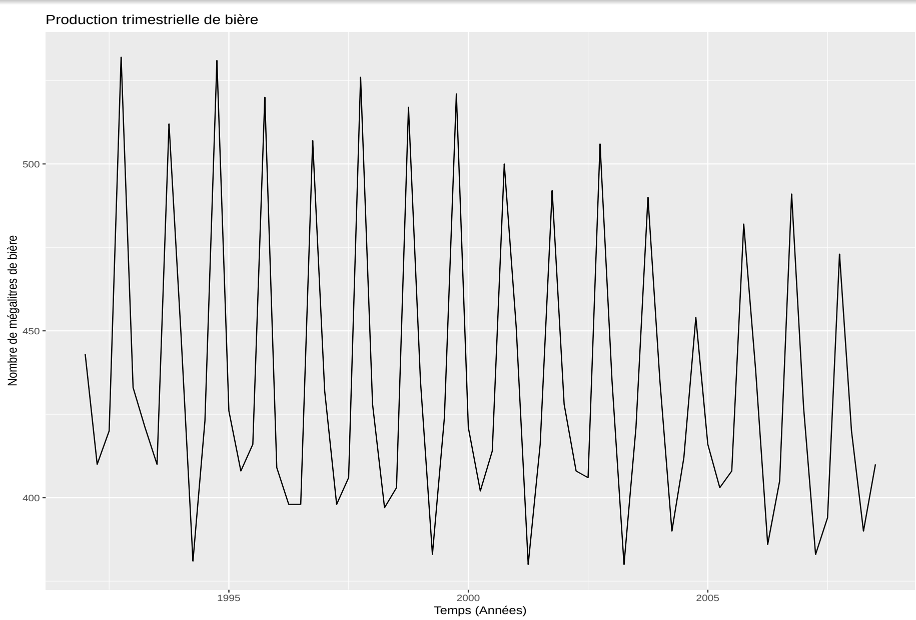
13 mai 2019

Quelques méthodes simples de prévision



Comment feriez-vous des prévisions à partir de cette série ?

Quelques méthodes simples de prévision - suite



Comment feriez-vous des prévisions à partir de cette série ?

Quelques méthodes simples de prévision - suite

1. Méthode de la moyenne :
   - Toutes les prévisions des valeurs futures sont égales à la moyenne des données historiques.
   - $\hat{x}_{T+h\,|\,T} = \dfrac{1}{T}(x_1 + \cdots + x_T)$
2. Méthode naïve :
   - Toutes les prévisions sont égales à la dernière observation.
   - $\hat{x}_{T+h\,|\,T} = x_T$

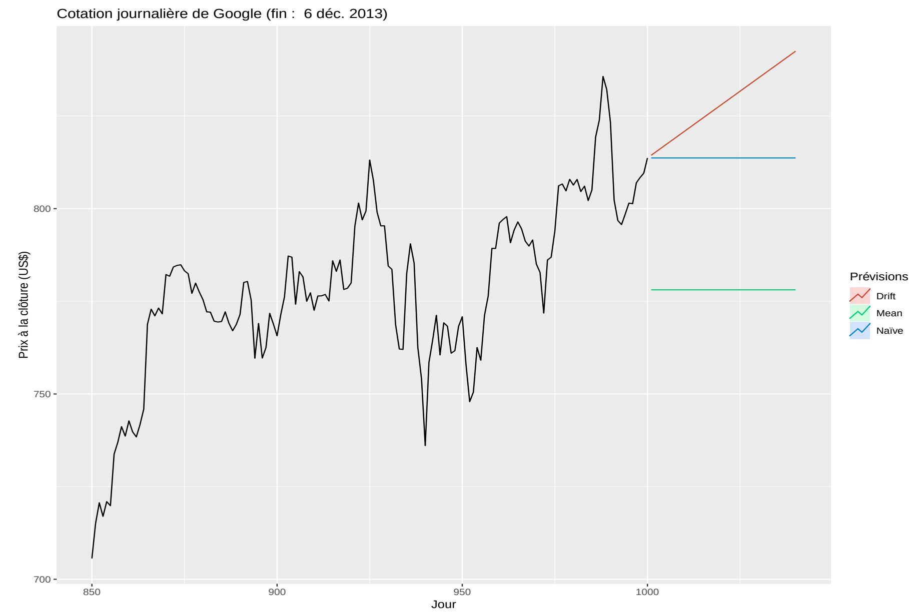
Quelques méthodes simples de prévision - suite

3. Méthode naïve saisonnière :
   - Toutes les prévisions sont égales à la dernière valeur de la précédente saison.
   - $\hat{x}_{T+h\,|\,T} = x_{T+h-m(k+1)}$ où $m =$ longueur de la saison et h est la partie entière de $(h-1)/m$.
4. Méthode de la dérive :
   - Les prévisions sont calculées via la formule :

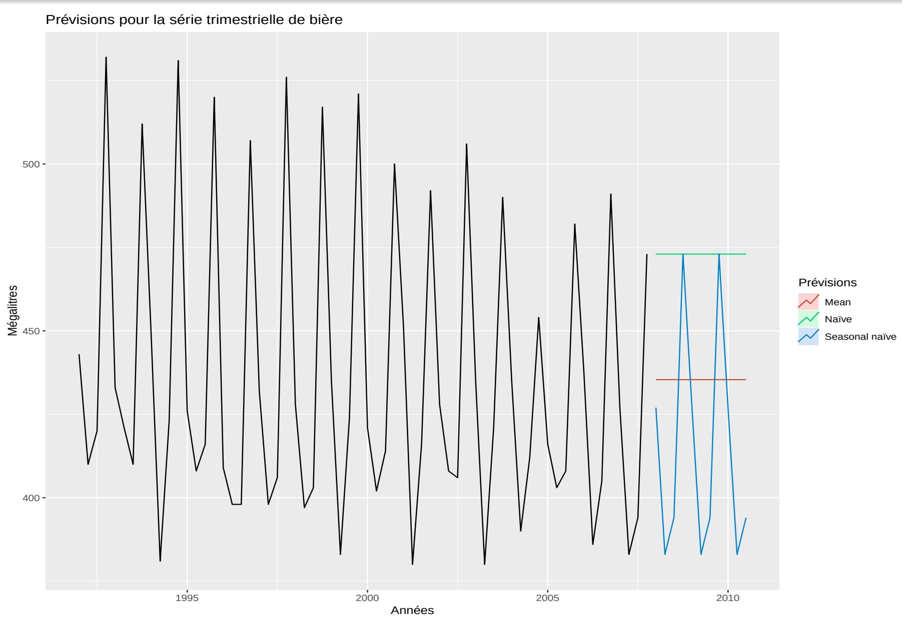
$$\hat{x}_{T+h\,|\,T} = x_T + \frac{h}{h-1}(x_T - x_1)$$

   - Cela équivaut à extrapoler une ligne droite entre la première et la dernière observation.

Quelques méthodes simples de prévision - suite



Quelques méthodes simples de prévision - suite



Quelques méthodes simples de prévision - Code "R"

1. Moyenne : meanf(x,h=15)
2. Naïve : naive(x,h=15)
3. Naïve saisonnière : snaive(x,h=24)
4. Dérive : drift(x,drift=TRUE,h=20)

Stabilisation de la variance

Si la variabilité autour de la moyenne change au cours du temps, il peut alors être utile d'utiliser une transformation du type Box et Cox qui dépend d'un paramètre λ :

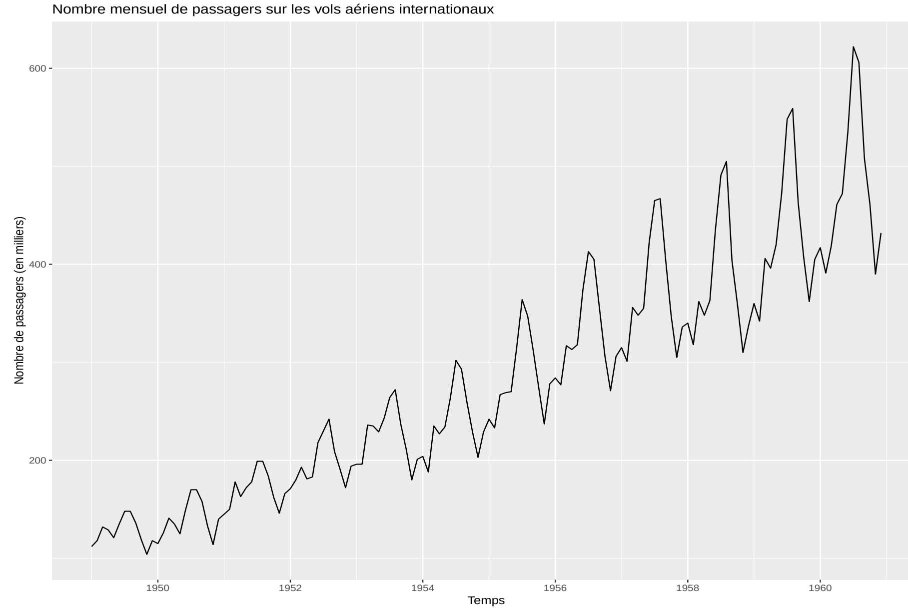
$$W_t = \begin{cases} \ln X_t & \text{si} \quad \lambda = 0 \\ \dfrac{X_t^\lambda - 1}{\lambda} & \text{si} \quad \lambda \neq 0\,. \end{cases}$$

La valeur de λ peut être calculée via "R" sur l'exemple de la série "airline" :

```
(lambda <- BoxCox.lambda(airline))
[1] -0.2947156
```

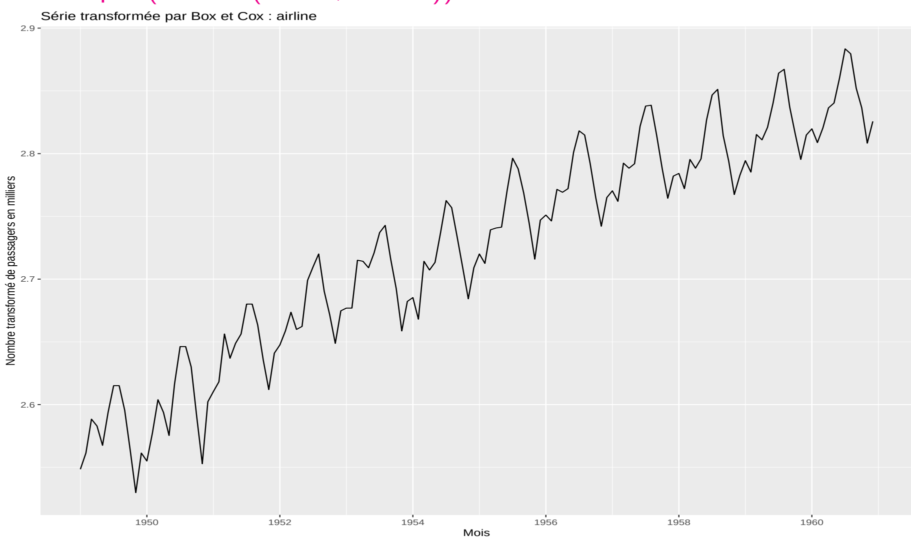
UNIVERSITÉ LAVAL

Série airline brute



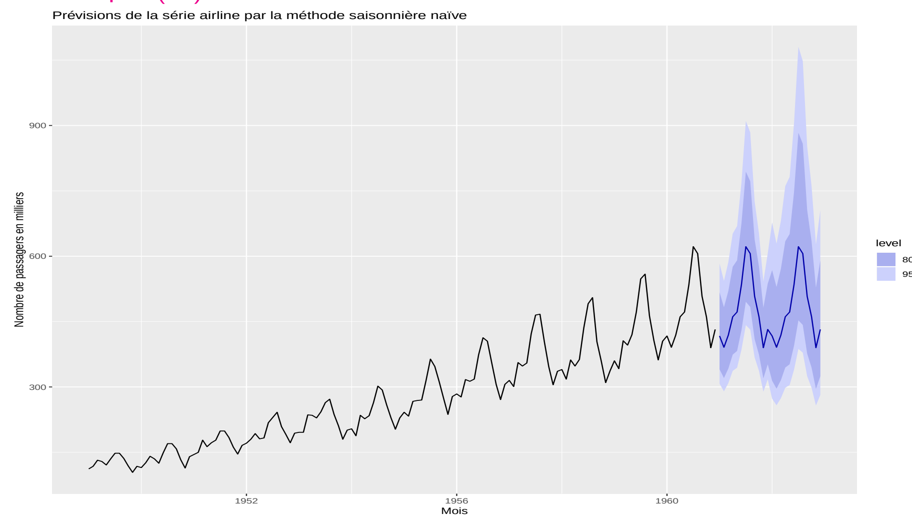
Série airline transformée

```
autoplot(BoxCox(airline,lambda))
```



Série airline et prévisions

```
fit<-snaive(airline,lambda=-0.3)
autoplot(fit)
```



Valeurs ajustées et résidus

- $\hat{x}_{t\,|\,t-1}$ est la prévision de x_t basée sur $x_1, \cdots, x_{t-1}$. On les appelle valeurs ajustées.
- On notera : $\hat{x}_t = \hat{x}_{t\,|\,t-1}$.

Exemple

- $\hat{x}_t = \overline{x}$ *par la méthode de la moyenne.*
- $\hat{x}_t = x_{t-1} + (x_T - x_1)/(T-1)$ *par la méthode de la dérive.*

Résidus

Un résidu à la date t est la différence entre la valeur observée et la valeur ajustée :

$$\varepsilon_t = x_t - \hat{x}_{t\,|\,t-1}\,.$$

Hypothèses

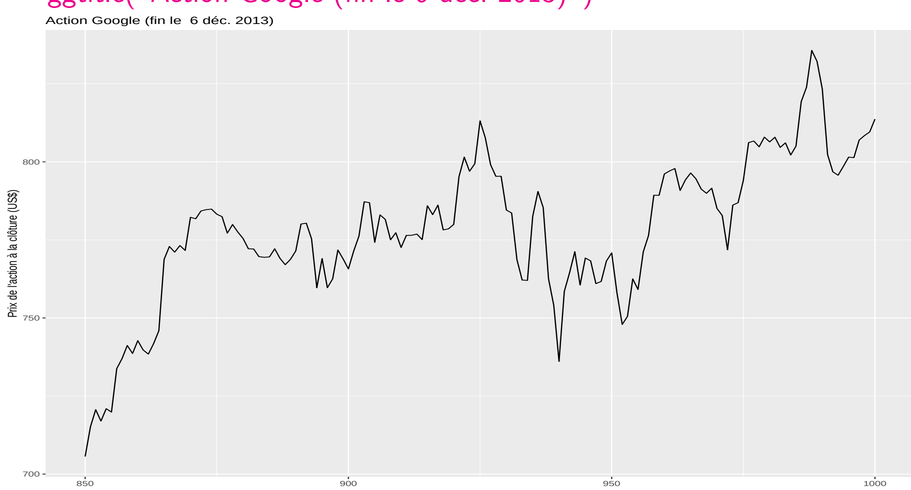
- *Les* ε_t *sont de moyenne nulle.*
- *Les* ε_t *sont deux à deux non corrélés.*

Propriétés utiles

- *Les* ε_t *ont une variance constante.*
- *Les* ε_t *suivent une loi normale..*

Exemple : série google

```
autoplot(goog150) +
xlab("Jour") + ylab("Prix de l'action à la clôture (dollars US)") +
ggtitle("Action Google (fin le 6 déc. 2013)")
```



Exemple : série google - suite

La prévision naïve est :

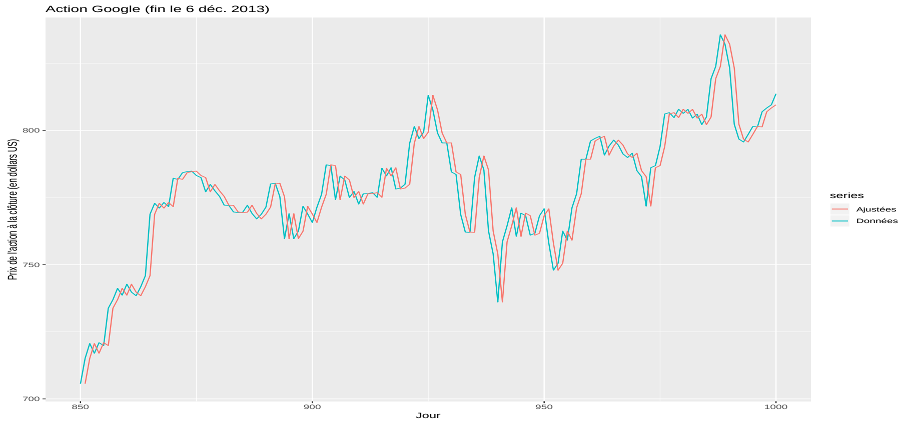
$$x_t - \hat{x}_{t\,|\,t-1} = x_{t-1}\,.$$

Le résidu associé est donc :

$$\varepsilon_t = x_t - x_{t-1}\,.$$

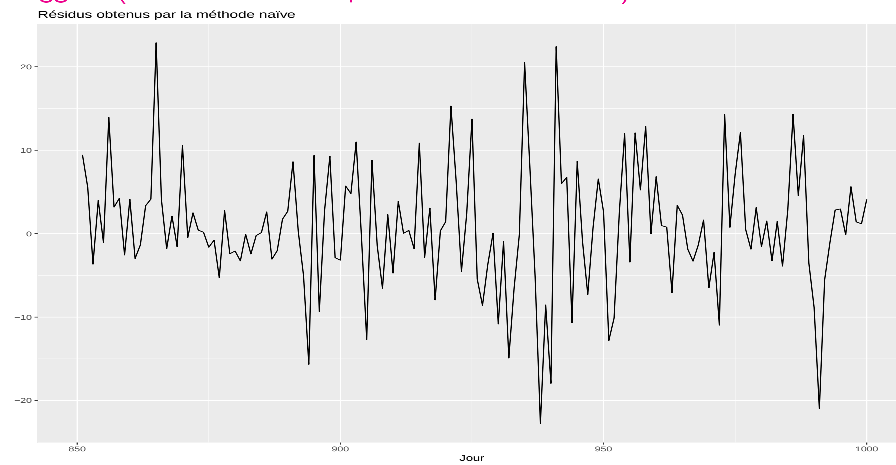
Exemple : série google - suite

```
fits <- fitted(naive(goog150))
autoplot(goog150, series="Données") +
autolayer(fits, series="Ajustées") +
xlab("Jour") + ylab("Prix de l'action à la clôture (en dollars US)")
+ ggtitle("Action Google (fin le 6 déc. 2013)")
```



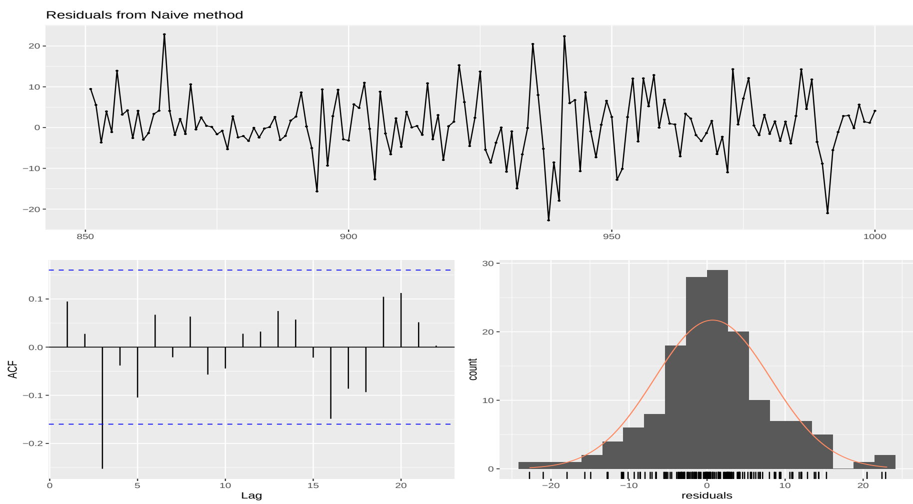
Exemple : série google - résidus

```
res <- residuals(naive(goog150))
autoplot(res) + xlab("Jour") + ylab("") +
ggtitle("Résidus obtenus par la méthode naïve")
```



Exemple : série google - étude des résidus

```
checkresiduals(naive(goog150))
```



Exemple : série google - étude des résidus

Box-Ljung test
data : Residuals from Naive method,
Q* = 15.623, df = 10, **p-value = 0.111**

On ne rejette pas l'hypothèse que les résidus sont issus d'un bruit blanc.

Mesures de précision des prévisions

Soient x_t une observation et f_t sa prévision, pour tout $t = 1, \cdots, T$.

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|x_t - f_t|$$

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(x_t - f_t)^2 \quad ; \quad RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - f_t)^2}$$

$$MAPE = \frac{1}{100\,T}\sum_{t=1}^{T}(|x_t - f_t|)/|x_t|$$

Elles dépendent toutes de l'échelle des observations, sauf le *MAPE*.

Mesures de précision des prévisions

Une autre mesure est le *MASE* (Mean Absolute Scale Error) :

$$MASE = \frac{1}{T}\sum_{t=1}^{T}|x_t - f_t|/q$$

où q est une mesure stable de l'échelle de la série (x_t).

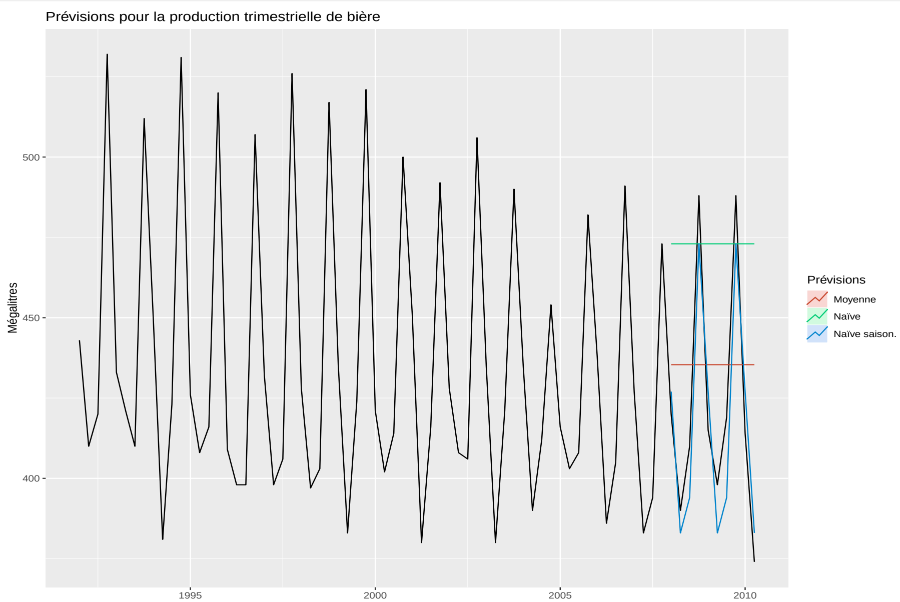
Exemple

Pour une série non saisonnière :

$$q = \frac{1}{T-1}\sum_{t=2}^{T}|x_t - x_{t-1}|$$

fonctionne bien. Alors le MASE équivaut au MAE pour la méthode naïve.

Mesures de précision des prévisions



Mesures de précision des prévisions

Le code "R" est le suivant :

```
biere2 <- window(biere,start=1992,end=c(2007,4))
bierefit1 <- meanf(biere2,h=10)
bierefit2 <- rwf(biere2,h=10)
bierefit3 <- snaive(biere2,h=10)
autoplot(window(biere, start=1992)) +
  autolayer(bierefit1, series="Moyenne", PI=FALSE) +
  autolayer(bierefit2, series="Naïve", PI=FALSE) +
  autolayer(bierefit3, series="Naïve saison.", PI=FALSE) +
  xlab("Années") + ylab("Mégalitres") +
  ggtitle("Prévisions pour la production trimestrielle de bièr
  guides(colour=guide_legend(title="Prévisions"))
```

Mesures de précision des prévisions

Le code "R" est :

```
  biere3 <- window(biere, start=2008)
accuracy(bierefit1, biere3)
accuracy(bierefit2, biere3)
accuracy(bierefit3, biere3)
```

	MAE	MAPE	RMSE	MASE
Moyenne	34,83	8,28	38,45	2,44
Méth. naïve	57,40	14,18	62,69	4,01
Naïve saison.	13,40	3,17	14,31	0,94

Intervalles de prévision

- Un intervalle de confiance est un intervalle dans lequel on espère que se situe x_{T+h} avec une certaine probabilité.
- Si on suppose les erreurs gaussiennes, alors l'intervalle de prévision à 95 % sera de la forme :

$$\hat{x}_{T+h} \pm 1,96 \times \hat{\sigma}_h$$

où $\hat{\sigma}_h$ est l'écart-type associé à la loi pour h décalages.

- Quand $h = 1$, $\hat{\sigma}_1$ peut être estimé via les résidus.

Intervalles de prévision

Prévisions naïves avec un intervalle de prévision :

```
res_sd <- sqrt(mean(res^2, na.rm=TRUE))
naive(goog150, level=95)

        Point Forecast   Lo 95      Hi 95
1001          813.67    799.0449   828.2951
1002          813.67    792.9869   834.3530
1003          813.67    788.3385   839.0015
1004          813.67    784.4197   842.9202
1005          813.67    780.9672   846.3728
1006          813.67    777.8459   849.4941
1007          813.67    774.9755   852.3644
1008          813.67    772.3039   855.0361
1009          813.67    769.7946   857.5454
1010          813.67    767.4213   859.9187
```


Intervalles de prévision

Remarque

- *Il est très important d'avoir des intervalles de prévision, ce qui donne une information très pertinente sur l'efficacité des dites prévisions.*
- *Les intervalles de confiance requièrent que le modèle sous-jacent soit stochastique.*
- *Le calcul des prévisions à l'horizon h nécessitent une approche plus sophistiquée, avec des intervalles de confiance qui grossissent avec l'horizon h de prévision.*

Intervalles de prévision

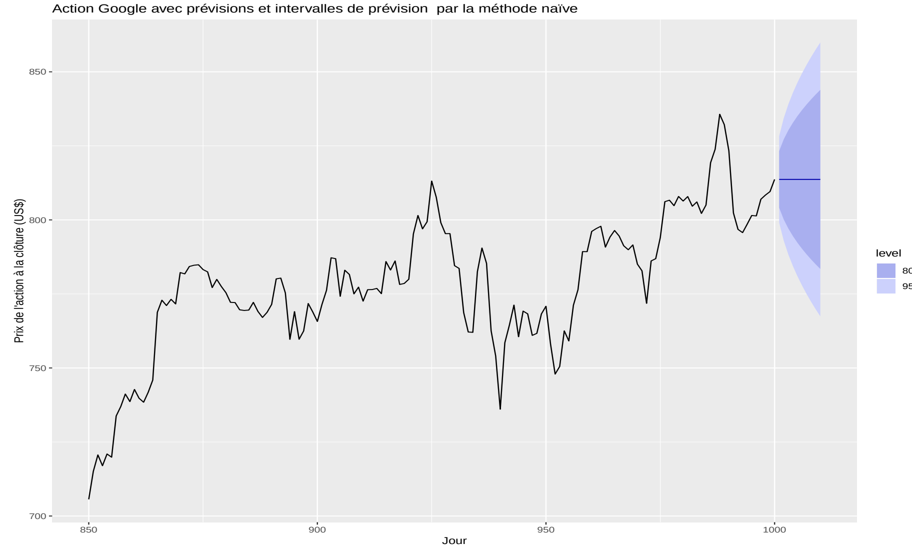
Si les résidus sont gaussiens, non corrélés deux à deux, et d'écart-type $\hat{\sigma}$, on a :

1. Méthode de la moyenne : $\hat{\sigma}_h = \hat{\sigma}\sqrt{1+1/T}$
2. Méthode naïve : $\hat{\sigma}_h = \hat{\sigma}\sqrt{h}$
3. Méthode naïve saison. : $\hat{\sigma}_h = \hat{\sigma}\sqrt{k+1}$
4. Méthode de la dérive : $\hat{\sigma}_h = \hat{\sigma}\sqrt{h(1+h/T)}$

où k est la partie entière de $(h-1)/m$.

Intervalles de prévision

autoplot(naive(goog150))



Exercice

À vous !

1. Importez la série mensuelle "lait.dat" démarrant en janvier 2005. Tracez-là.
2. Séparez les données en deux parties (la première "lait.train" allant de janvier 2005 à décembre 2017, et l'autre "lait.test", démarrant en janvier 2018). Tracez-les.
3. Calculez les prévisions saisonnières naïves sur 12 mois à partir de la série "lait.train". Faites un graphique.
4. Comparez la pertinence de vos prévisions vis-à-vis du réalisé a posteriori.
5. Examinez les résidus. Commentez.