# Chapitre 1

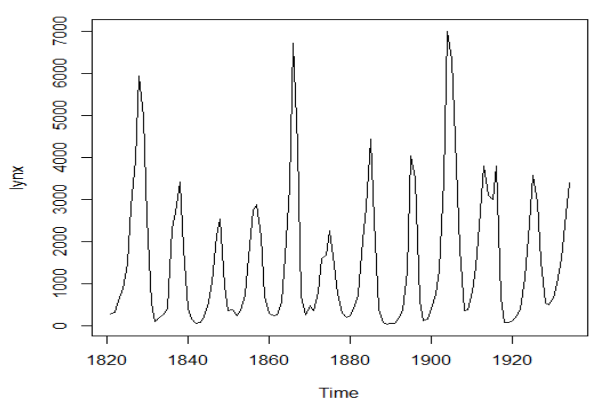
# Introduction et Analyse descriptive

## 1.1 Introduction

L'objectif de ce cours est de modéliser l'évolution d'un phénomène aléatoire dans le temps qui est supposé discret. On parle de série temporelle.

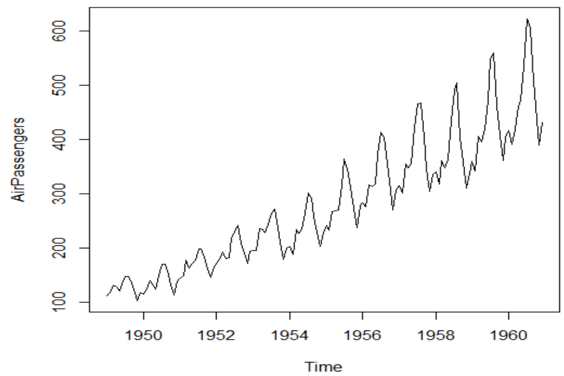
### 1.1.1 Exemples de séries temporelles.

Les séries temporelles sont présentes dans de nombreux domaines d'application. Par exemple, en écologie, on peut citer la célébre série du nombre de lynx capturés au Canada de 1821 à 1934 et dont la représentation est donnée par la Figure 1 (Source : datastes R : pour le graphe : plot(lynx))

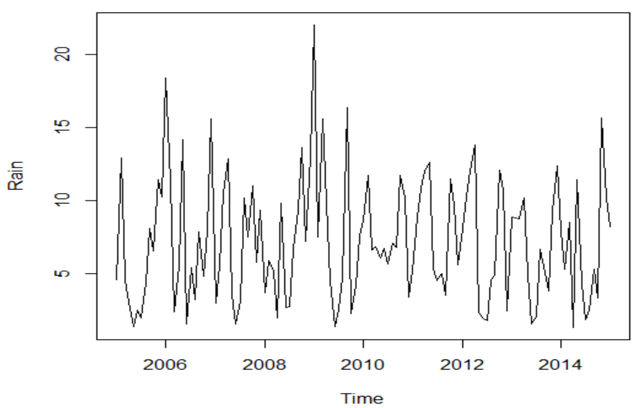



En économie les séries temporelles sont très utilisées. Un exemple est donné par la série :

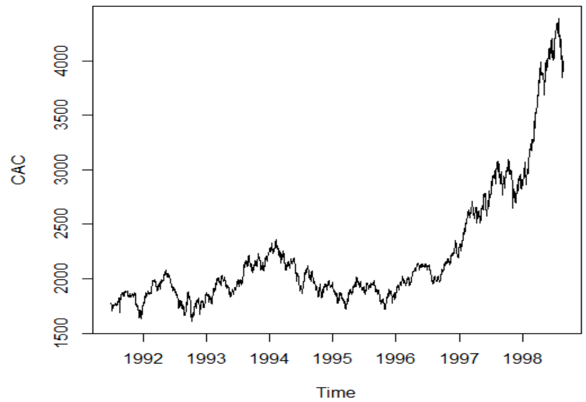
AirPassengers, du nombre de passagers dans les lignes aériennes entre les années 1949 et 1960 et est présente dans la base de donnée du R : plot(AirPassengers) Figure 2.




En écologie, on pourrait citer la série de pluie mensuelle de 2005 à 2015 en Algérie, figure 3. (Source : https ://climateknowledgeportal.worldbank.org/download-data)




En finance, le cours des actions est suivi avec beaucoup d'attention, la figure 4 représente l'indice boursier français CAC 40. (Source : EuStockMarkets, R).




Dans la suite, on suppose que la série observée est une suite de variables aléatoires où l'ordre est important. On note $X_t$ la valeur observée au temps $t$, le processus $(X_t)_{t\in\mathbb{N}}$ est alors appelée série temporelle. Le but est de trouver un modèle qui représente au mieux les données, c'est le problème de la modélisation. L'objectif est de détecter la présence d'une tendance, d'une saisonnalité et d'extraire la structure du signal dans les données mais bien sur la mission principale reste la prévision des valeurs futures de la série.

## 1.2 Décomposition d'une série temporelle

### 1.2.1 Tendance et saisonnalité d'une série temporelle.

On peut décomposer une série temporelle en trois termes sous la forme :

$$Y_t = T_t + S_t + X_t$$

où :*$T_t$ est une fonction déterministe qui représent la variations à long terme appelée la **tendance**.

*$S_t$ :est une fonction déterministe périodique appelée **saisonnalité** de période $T$ telle que $\sum_{i=1}^{T} S_{t+i} = 0, \forall t \in \mathbb{N}$.

*$X_t$ est un bruit aléatoire stationnaire appellé **résidu**.

Le but est d'apprendre à modéliser et estimer les différentes composantes et de faire des prévisions sur les valeurs futures de la série temporelle initiale. (le modèle précédent est appelé additif, en remplaçant $+$ par $\times$ on aura le modèle multiplicatif).

**Une approche générale de la modélisation des séries temporelles.**

La méthode générale pour étudier une série temporelle est :

1) En premier, on trace la série des données et on repère ses principales composantes. On regarde en particulier si on détecte :

*une tendance, *une composante saisonnière, * des ruptures dans le comportement de la série et des observations aberrantes.

2) Deuxièment, on estime la tendance et la composante saisonnière pour obtenir une série de résidus stationnaires.

3) Choisir un modèle de série stationnaire pour les résidus

4) Prévoir les valeurs futures de la série en prévoyant d'abord celles des résidus puis "remonter" jusqu'à la série initiale en utilisant les transformations inverses.

**Indices descriptifs d'une série temporelle**

-**Indices de tendance centrale :** la moyenne empirique :

$$\overline{X} = \frac{1}{n} \sum_{t=1}^{n} X_t$$

-**Indices de dispersion :** la variance empirique

$$\widehat{\gamma}_0 = \frac{1}{n} \sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2$$

-**Indices de dépendance : L'auto-covariance** empirique d'ordre $h$ : c'est une notion très importante et qui renseigne sur la dépendance entre les données écartées de $h$ pas de temps.

$$\widehat{\gamma}_h = \frac{1}{n-h} \sum_{t=1}^{n-h} \left(X_t - \overline{X}\right) \left(X_{t+h} - \overline{X}\right)$$

On considère les covariances empiriques jusqu'à un ordre $h$ pas trop grand. Les **auto-corrélations** empiriques sont les quotients des covariances empiriques par la variance empirique :

$$\widehat{\rho}_h = \frac{\widehat{\gamma}_h}{\widehat{\gamma}_0}$$

En interprétant l'auto-corrélation, il sera possible de « deviner »si une série temporelle admet une tendance, l'auto-corrélation tend vers 1, ou une saisonnalité.

## 1.3 Lissages exponentiels

Les méthodes de lissages exponentiels sont un outil permettant de réaliser des prévisions à partir de l'observation d'une série temporelle. Ces méthodes sont simples en mise en oeuvre, elles sont souvent utilisées dans l'industrie, notamment lorsque le nombre de prévisions à réaliser est important (par exemple, prévisions des ventes de centaines de produits dans une grande surface).

On présente trois types de lissage exponentiel :

– le lissage exponentiel simple qui consiste à ajuster localement à la série temporelle une constante,

– le lissage exponentiel double qui ajuste une droite,

– le lissage exponentiel de Holt-Winters qui considère des fonctions plus complexes (polynomiales, périodiques...).

### 1.3.1 Lissage exponentiel simple (LES)

Soit la série temporelle $X_1, ..., X_n$ l'objectif du lissage exponentiel est d'estimer la valeur $X_{n+h}$ non encore observée. Nous noterons $\widehat{X}_n(h)$ cette prévision. Etant donnée une constante de lissage $0 < \alpha < 1$, on donne la formules récursives de mise à jour

$$\widehat{X}_n(h) = \alpha X_n + (1-\alpha)\widehat{X}_{n-1}(h)$$

L'utilisation de cette récurrence permet de réaliser des algorithmes très rapides d'estimation de la prévision par lissage exponentiel avec l'initialisation $\widehat{X}_1(h) = X_1$.

### 1.3.2 Lissage exponentiel double (LED)

On ajuste au voisinage de l'instant $n$ une droite d'équation $Y_t = a_1 + a_2 (t - n)$ . La prévision par lissage exponentiel double est :

$$\widehat{X}_n (h) = \widehat{a}_1 (n) + \widehat{a}_2 (n) h$$

avec les formules récursives de mise à jour

$$\begin{aligned} \widehat{a}_1 (n) &= \widehat{a}_1 (n-1) + \widehat{a}_2 (n-1) + \alpha (2-\alpha) \left( X_n - \widehat{X}_{n-1} (1) \right) \\ \widehat{a}_2 (n) &= \widehat{a}_2 (n-1) + \alpha (2-\alpha) \left( X_n - \widehat{X}_{n-1} (1) \right) \end{aligned}$$

Les valeurs initiales $\widehat{a}_1 (0) = X_1, \widehat{a}_2 (0) = X_2 - X_1$

### 1.3.3 Méthode de Holt-Winters

**Méthode saisonnière additive (HWA)**

On veut ajuster au voisinage de l'instant $n$ une droite d'équation

$$Y_t = a_1 + a_2 (t - n) + S$$

où $S_t$ est une composante périodique de période $T$ . Les formules récursives de mise à jour sont

$$\begin{aligned} \widehat{a}_1 (n) &= \alpha \left( X_n - \widehat{S}_{n-T} \right) + (1-\alpha) (\widehat{a}_1 (n-1) - \widehat{a}_2 (n-1)) \\ \widehat{a}_2 (n) &= \beta (\widehat{a}_1 (n) - \widehat{a}_1 (n-1)) + (1-\beta) \widehat{a}_2 (n-1) \\ \widehat{S}_n &= \gamma (X_n - \widehat{a}_1 (n)) + (1-\gamma) \widehat{S}_{n-T} \end{aligned}$$

Les prévisions sont de la forme :

$$\begin{aligned} \widehat{X}_n (h) &= \widehat{a}_1 + \widehat{a}_2 h + \widehat{S}_{n+h-T}, 1 \leq h \leq T \\ \widehat{X}_n (h) &= \widehat{a}_1 + \widehat{a}_2 h + \widehat{S}_{n+h-2T}, T+1 \leq h \leq 2T, ... \end{aligned}$$

L'effet des trois constantes de lissages, $\alpha, \beta$ et $\gamma$ est que plus elles sont petites plus l'importance des données éloignées est significative.

**Méthode saisonnière multiplicative (HWM)**

On ajuste au voisinage de l'instant $n$ une droite d'équation

$$Y_t = (a_1 + a_2 (t - n)) \times S_{t.}$$

Les formules récursives de mise à jour sont

$$\begin{aligned}
\widehat{a}_1 (n) &= \alpha \left( \frac{X_n}{\widehat{S}_{n-T}} \right) + (1 - \alpha) (\widehat{a}_1 (n - 1) - \widehat{a}_2 (n - 1)) \\
\widehat{a}_2 (n) &= \beta (\widehat{a}_1 (n) - \widehat{a}_1 (n - 1)) + (1 - \beta) \widehat{a}_2 (n - 1) \\
\widehat{S}_n &= \gamma \left( \frac{X_n}{\widehat{a}_1 (n)} \right) + (1 - \gamma) \widehat{S}_{n-T}
\end{aligned}$$

Les prévisions sont de la forme :

$$\begin{aligned}
\widehat{X}_n (h) &= (\widehat{a}_1 + \widehat{a}_2 h) \widehat{S}_{n+h-T}, 1 \leq h \leq T \\
\widehat{X}_n (h) &= (\widehat{a}_1 + \widehat{a}_2 h) \widehat{S}_{n+h-2T}, T + 1 \leq h \leq 2T
\end{aligned}$$

**Commandes sous R**

Soit une série temporelle $x$, les lissages sous R se font avec :

– pour LES : $xlis < -HoltWinters(x, alpha = \alpha, beta = FALSE, gamma = FALSE)$,

– pour LED : $xlis < -HoltWinters(x, alpha = \alpha, beta = \beta, gamma = FALSE)$,

– un HWA : $xlis < -HoltWinters(x, alpha = \alpha, beta = \beta, gamma = \gamma, seasonal = \prime add \prime)$,

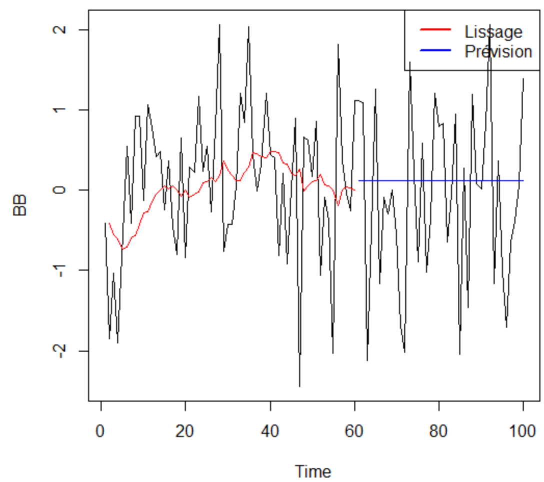
– un HWM : $xlis < -HoltWinters(x, alpha = \alpha, beta = \beta, gamma = \gamma, seasonal = \prime mul \prime)$.

La représentation de l'ajustement de la série se fait avec : $plot(xlis\$fitted[, 1])$

Les prévisions à l'horizon $h$ sont réalisées à l'aide de la fonction $predict$ :

$p < -predict(xlis, n.ahead = h)$.

Le graphe suivant représente le lissage et prévision par LES d'un bruit blanc gaussien.




## 1.4 Estimation et élimination de la tendance et de la saisonnalité

### 1.4.1 Bruit blanc

**Définition**. Un processus de bruit blanc est une suite de variables aléatoires $(\varepsilon_t)$ indépendantes, d'espérance 0 et de variance constante. Si les variables aléatoires sont gaussiennes, le bruit blanc est gaussien.

**Processus stationnaire**

Un processus aléatoire $(X_t)$ est stationnaire s'il est d'espérance constante et si les covariances sont stables par translation dans le temps

$$E(X_t) = \mu, \forall t \text{ et } cov(X_t, X_{t-h}) = \gamma_h, \forall t$$

### 1.4.2 Estimation paramétrique de la tendance

Supposons que la série soit la réalisation du processus stochastique

$$Y_t = T_t + \varepsilon_t \text{ où } T_t = a + bt$$

la tendance est linéaire, on estime les paramètres a et b par moindres carrés. On peut supposer que la tendance soit de forme polynomiale et l'estimation se fera toujours par moindres carrés.

Mais il est parfois difficile d'estimer le degré du polynôme, et lorsque le degré est trop important, le nombre de paramètres à estimer devient grand et les calculs fastidieux. Dans cette situation, on a recourt à une méthode d'estimation non paramétrique.

### 1.4.3 Estimation non paramétrique : moyenne mobile

**Tendance sans saisonnalité** un bon estimateur de la tendance $T_t$ est la moyenne

$$\widehat{T}_t = \frac{1}{2q+1} \sum_{k=-q}^{q} Y_{t+k}$$

la tendance est estimé à chaque temps t en calculant la moyenne sur les observations étant dans une fenêtre de largeur 2q + 1 autour de t : l'estimation par moyenne mobile (q est supposé impaire, on verra le cas pair en TD).

**Tendance et saisonnalité** Soit le modèle

$$Y_t = T_t + S_t + \varepsilon_t$$

avec $\mathrm{S}_t$ une fonction T-périodique. Le principe d'estimation est : on estime la tendance moyenne sur une période, puis on estime la composante saisonnière en moyennant sur toutes les périodes les écarts à la tendance moyenne de la période. (Voir TD, fonction R : decompose)

### 1.4.4 Tester le bruit blanc

-**Les autocorrélations empiriques** Lorsque $n$ est assez grand, les autocorrélations d'un bruit blanc sont approximativement indépendantes et de loi N $(0,\frac{1}{n})$. Ainsi, 95% des autocorrélations devraient se trouver dans l'intervalle [-$\frac{1.96}{\sqrt{n}}$, $\frac{1.96}{\sqrt{n}}$], et en traçant les 40 premières autocorrélations il ne devrait pas y en avoir plus de 2 ou 3 en dehors de ces limites. fonction R : acf.

- **Test Ljung et Box** La statistique du test est

$$Q = n\,(n+2) \sum_{j=1}^{h} \frac{\widehat{\rho}_j^2}{n-j}$$

qui suit une loi du $\chi^2$à $h$ degrés de liberté. L'hypothèse nulle est que la série est un bruit blanc. fonction R : Box.test.