

Statistiques : notes du cours 6
Lissage de séries temporelles

Une série temporelle est une suite de données indéxée par le temps $(x_t)_{1 \le t \le n}$, comme par exemple la production mensuelle d'une entreprise, le nombre d'acheteurs par jour sur un site web, le nombre d'individus atteint chaque année par une épidemie, le niveau d'une pollution, le prix d'une matière première ou celui d'un indice boursier. On cherche le plus souvent à prévoir une valeur future mais aussi parfois à reconstituer une valeur manquante et toujours à comprendre et expliquer les variations observées.

Les techniques mathématiques d'étude de séries temporelles vont des plus simples comme le lissage aux plus élaborées comme la modèlisation de Box et Jenkins qui est à la base de l'économétrie. Ces techniques sont aujourd'hui implémentées dans la plupart des logiciels de traitement statistique des données.

1 Lissage par régression linéaire (rappel)

Si la suite des points $(t, x_t)_{1 \le t \le n}$ représentés dans un plan ressemble à une droite, on a vu comment calculer la *droite des moindre carrés* encore appelée *droite de régression linéaire*. On pose :

$$(\hat{a}, \hat{b}) = ArgMin_{(a,b)} \sum_{t=1}^{n} (x_t - (at + b))^2.$$

La valeur $x^* = x_{t^*}$ prédite par ce modèle à un instant $t^* > t$ est alors donnée par $x^* = \hat{a}t^* + \hat{b}$.

Si les données présentent une croissance exponentielle, on fait de même en remplaçant (t, x_t) par les données (t, z_t) où $z_t = \ln x_t$ et la valeur prédite sera alors $x^* = \exp(\hat{a}t^* + \hat{b})$.

On peut aussi ajuster les données, par la méthode des *moindres carrés ordinaires* (MCO, ou OLS en anglais), à une fonction $f(t)$ non nécessairement linéaire : par exemple pour les ajuster à une fonction polynome de degré deux, on pose :

$$(\hat{a}_2, \hat{a}_1, \hat{b}) = ArgMin_{(a_2,a_1,b)} \sum_{t=1}^{n} (x_t - (a_2 t^2 + a_1 t + b))^2.$$

Cependant, lorsque la série présente de grandes variations, un lissage peut se revéler plus adapté.

2 Lissage par moyenne mobile

La moyenne mobile (MA pour Moving Average en anglais) la plus simple de la série temporelle $(x_t)_{1 \le t \le n}$ est une série temporelle $(y_t)_{2 \le t \le n}$ définie par $y_t = (x_t + x_{t-1})/2$. Les séries $y_t = (x_t + x_{t-1} + x_{t-2})/3$ ou $y_t = (x_{t+1} + x_t + x_{t-1})/3$ sont d'autres exemples, définis respectivement pour $3 \le t \le n$ et $2 \le t \le n-1$. Plus généralement, on dit qu'on applique à la série x_t un *filtre moyenne mobile d'ordre* $2m + 1$ si on la transforme en

$$y_t = \sum_{k=-m}^{k=+m} \alpha_k x_{t+k}$$

avec $\sum_{-m}^{m} \alpha_k = 1$, les α_k étant des réels positifs donnés ; si pour tout $k \in \{-m, \ldots, m\}$, on a $\alpha_k = \frac{1}{2m+1}$, le filtre est appelé filtre *optimal* et noté $y_t = F_{2m+1}(x_t)$.

Proposition 1 *Un filtre moyenne mobile optimal laisse passer sans distorsion une série temporelle linéaire* $x_t = at + b$.

Proposition 2 *Toute série temporelle périodique de période T (i.e. vérifiant, pour tout t, $S(t+T) = S(t)$ et aussi $\frac{1}{T}\sum_{k=1}^{T} S(t+k) = 1$) est arrêtée par un filtre moyenne mobile optimal d'ordre T.*

Pour faire des prévisions on ne prendra pas un filtre centré car il ne pourrait prendre en compte les dernières valeurs connues de x ce qui est regrettable. On prendra plutôt un filtre moyenne mobile de la forme $y_t = \sum_{k=-m}^{-1} \alpha_k x_{t+k}$. La valeur prédite à un instant $t + N$ si on connait les valeurs aux instants antérieurs, sera alors donnée par $x^*_{t+N} = \sum_{k=-m}^{-1} x_{t+k}$ pour tout $N \geq 0$. C'est une prévision constante dans le futur.

Notons enfin que pour filtrer une saisonnalité de période paire, $T = 2m$, tout en gardant le caractère symétrique du filtre, on utilise une moyenne mobile de la forme :

$$y_t = \frac{1}{4m}\left(\sum_{k=-m}^{m-1} x_{t+k} + \sum_{k=-m+1}^{m} x_{t+k}\right)$$

ce qui donne par exemple pour une série de données mensuelles présentant une saisonnalité de 12, $y_t = \frac{1}{24}(x_{t-6} + 2x_{t-5} + \ldots + 2x_t + \ldots + 2x_{t-5} + x_{t+6})$.

3 Modèle stochastique

Plutôt qu'une simple suite de nombres, il est souvent utile de considérer une série temporelle $(x_t)_{1\leq t\leq n}$ comme les valeurs prises par une suite de variables aléatoires $X_t : \Omega \to \mathbb{R}$ indexée par le temps. L'ensemble Ω représente l'ensemble des *états du monde* et pour chaque t, $x_t = X_t(\omega)$ est la valeur prise par la v.a. X_t dans l'état ω dans lequel on se trouve. Par exemple on peut considérer que la série $x_t = f(t) + \varepsilon_t$ est la perturbation d'une fonction $f(t)$ déterministe par un *bruit* aléatoire ε_t. Une suite $(\varepsilon_t)_{1\leq t\leq n}$ de v.a. centrées ($\mathbb{E}(\varepsilon_t) = 0$), de variance constante ($\mathrm{Var}\,(\varepsilon_t) = \sigma^2$) et non corrélée ($\mathrm{Cov}\,(\varepsilon_t, \varepsilon_{t'}) = 0$ si $t \neq t'$) s'appelle un *bruit blanc*.

Proposition 3 *Le filtre moyenne mobile d'une série temporelle de la forme $x_t = f(t) + \varepsilon_t$ perturbation d'une fonction par un bruit blanc de variance σ^2, est une série de la forme $y_t = \tilde{f}(t) + \tilde{\varepsilon}_t$ où $\tilde{\varepsilon}_t$ est un bruit blanc de variance $\sigma^2 \sum_{k=-m}^{k=m} \alpha_k^2$ donc strictement inférieure à σ^2. De plus le filtre qui rend $Var(\tilde{\varepsilon}_t)$ minimal est le filtre optimal (i.e. le filtre pour lequel $\alpha_k = \frac{1}{2m+1}$ pour tout k).*

4 Lissage exponentiel

Soit $\beta \in]0, 1[$. On appelle *filtre exponentiel* de la série temporelle x_t la série

$$y_t = \beta \sum_{k=0}^{\infty} (1-\beta)^k x_{t-k}.$$

C'est une combinaison linéaire infinie de toute les observations passées avec la pondération $\alpha_k = \beta(1-\beta)^k$. Cette suite de coefficients est une suite géométrique de premier terme β, de raison $1-\beta$, décroissant très rapidement vers zéro et telle que $\sum_{k=0}^{\infty} \alpha_k = 1$. Plus β est proche de 1, moins le passé immédiat compte au regard du passé lointain et, au contraire, plus β est proche de zéro, plus il compte.

Dans un modèle stochastique où les x_t seraient les valeurs de variables aléatoires indépendantes de variance constante $\mathrm{Var}\,(x_t) = \sigma^2$, la variance de la série y_t obtenue par lissage exponentiel vaut $\mathrm{Var}\,(y_t) = \frac{\beta}{2-\beta}\sigma^2$. La variance de la série lissée est donc strictement inférieure à celle de la série d'origine.

Ce lissage fournit, comme les précédents, une prévision (constante) $x^*_{t+N} = \beta \sum_{k=0}^{\infty} (1-\beta)^k x_{t-k}$ pour toute date $t + N > t$. Si β est proche de 1, la prévision comporte une certaine rigidité et une grande dépendance au passé. Si au contraire β est petit, la prévision est plus souple et ne dépend que faiblement du passé. Pour ne pas être obligé de recalculer la prévision chaque fois qu'une nouvelle valeur de x est révélée, on dispose également de la *formule de mise à jour* suivante : $x^*_{t+1} = \beta x_t + (1-\beta)x^*_t$.

5 Lissage exponentiel double

La méthode de lissage exponentiel double combine la méthode de lissage par régression linéaire et le lissage exponentiel (que l'on appelle encore pour cette raison *lissage exponentiel simple*). Le principe est d'ajuster la série à une droite $at + b$, de façon à pouvoir faire une prévision linéaire et non constante, $x^*_{t+N} = aN + b$ mais on calcule cette fois des valeurs de a et b dépendant du temps en posant :

$$(\hat{a}(t), \hat{b}(t)) = ArgMin_{(a,b)} \sum_{k=0}^{t-1} (1-\beta)^k (x_{t-k} - (ak + b))^2$$

c'est-à-dire que l'on recherche les valeurs de a et b qui minimisent la somme des carrés des écarts *pondérés de façon exponentielle* par les $(1-\beta)^k$.