Utilisation de la programmation dynamique pour la segmentation de series temporelles

Aline Canard*

11 mars 2019

Abstract

Complex data analysis is ...

*Master2 SAAD, Université de Caen Normandie

1 Introduction

Les séries temporelles interviennent dans de nombreux domaines tels que la finance, l'économie, ingénierie, la bioinformatique pour représenter les variations d'une mesure au cours du temps. Des techniques de modélisation vont permettre de donner une représentation synthétique de tels phénomènes. Ce document présentera une méthode de segmentation de série temporelle : la régression par morceaux basée sur un modèle polynomial hétéroscédastique. La méthode utilisée est la programmation dynamique. Les paramètres du modèle seront estimés par maximum de vraisemblance. En général, cette méthode est plutôt utilisée dans le cas de changement abruptes de régimes.

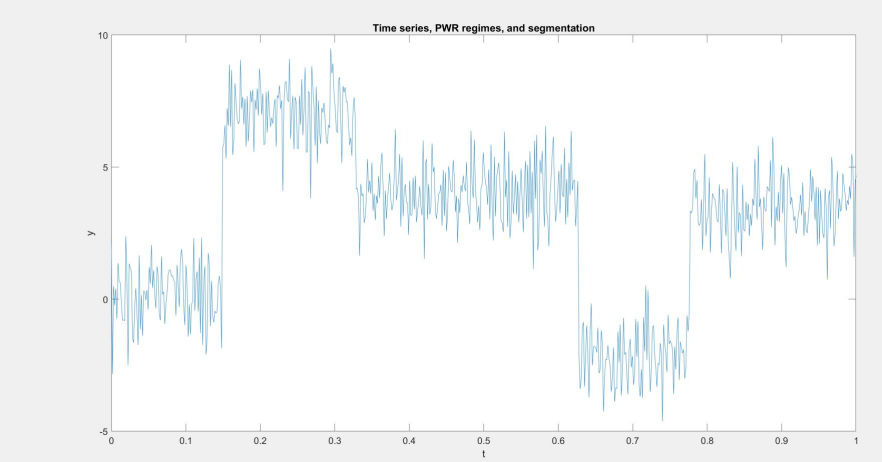

FIGURE 1 – Exemple de série temporelle à segmenter

Ce document a pour but de présenter la méthode mathématique et le code R réalisé pour effectuer la segmentation de telle série temporelle. Il est basé sur la publication : (**???**). Citer la publi

2 Modèle de mélange pour des données fonctionnelles

2.1 Définition du modèle

Soit $\mathbf{x} = (\mathcal{X}_1, ..., \mathcal{X}_n)$, une observation d'une série temporelle où les x_i sont observés au temps t_i. Le modèle de régression polynomiale par morceaux suppose que la série temporelle

soit composée de K régimes sur K intervalles dont les indices de bornes peuvent être notés : $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{K+1})$ avec $\gamma_1 = 0$ et $\gamma_{K+1} = n$. Cela définit une partition de la série temporelle en K segments polynomiaux $(\mathbf{x}_1, ..., \mathbf{x}_K)$, de longueur $n_1, ..., n_K$ où $\mathbf{x}_K = \{\mathbf{x}_i | i \in I_K\}$ est l'ensemble des éléments d'un segment k dont les indices sont : $I_K = (\boldsymbol{\gamma}_K, \boldsymbol{\gamma}_{K+1}]$. Les modèles standards de régression polynomiale sont des modèles homoscédastiques puisqu'ils supposent que les différents modèles de régression polynomiale ont la même de bruit. Dans notre cas, nous considérons le cadre plus général d'un modèle hétéroscédastique ; ce qui permet d'avoir une variance bruit entre les différents modèles de régression polynomiale.On peut le définir ainsi :

$$\forall i = 1, ..n, \mathbf{x}_i = \boldsymbol{\beta}_k^T \mathbf{r}_i + \sigma_k \epsilon_i; \epsilon_i \sim \mathcal{N}(0, 1), \tag{1}$$

où k est tel que $i \in I_k, \boldsymbol{\beta}_k$ est le vecteur des coefficients de la $(p+1)^{eme}$ dimension d'un polynôme de degré p associé au k^{eme} segment, avec $k \in \{1, .., K\}$, $\mathbf{r}_i = (1, t_i, t_i^2, ..., t_i^p)^T$ est le vecteur temporel de dimension $(p+1)$ associé au paramètre $\boldsymbol{\beta}_k$, et les ϵ_i sont des variables aléatoires, indépendantes et distribuées selon une loi normale. Elles représentent un bruit additionnel dans chaque segment k.

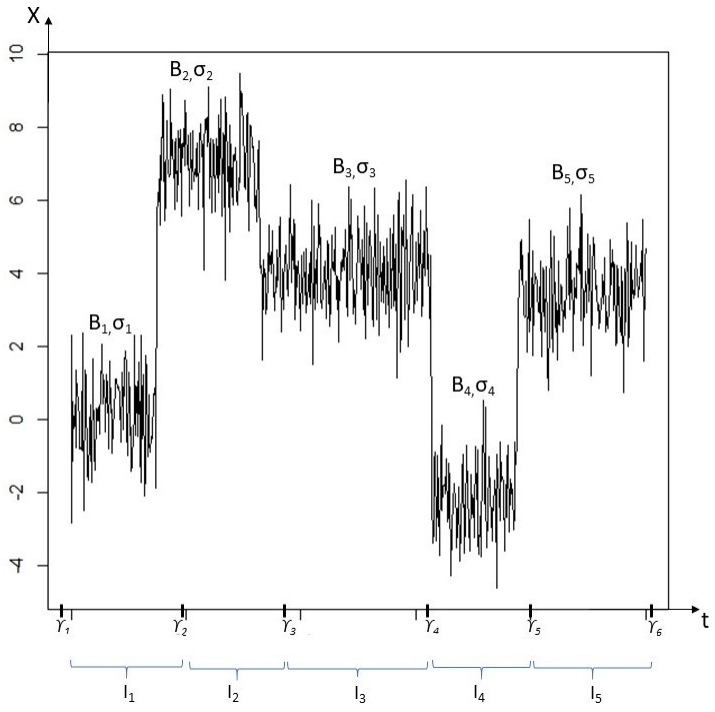

FIGURE 2 – Série temporelle avec les notations

L'objectif de la segmentation sera de positionner les $\boldsymbol{\gamma}_k$ de façon optimale.

$$\text{A l'instant } t, \text{ on peut écrire : } \mathbf{x}_t = \begin{cases} \boldsymbol{\beta}_1^T \mathbf{r}_t + \sigma_1 \epsilon_t, \; si \; \mathbf{x}_t \in I_1 \\ \boldsymbol{\beta}_2^T \mathbf{r}_t + \sigma_2 \epsilon_t, \; si \; \mathbf{x}_t \in I_2 \qquad \text{où } \epsilon_t \; \sim \mathcal{N}(0,1) \\ ... \\ \boldsymbol{\beta}_K^T \mathbf{r}_t + \sigma_K \epsilon_t, \; si \; \mathbf{x}_t \in I_K \end{cases}$$

2.2 Estimation des paramètres du modèle par maximum de vraisemblance

Avec ce modèle, les paramètres peuvent s'écrire de la forme : $(\boldsymbol{\psi}, \boldsymbol{\gamma})$ où $\boldsymbol{\psi} = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_k, \boldsymbol{\sigma}_1^2, .., \boldsymbol{\sigma}_k^2)$ est l"ensemble des coefficients polynomiaux et des variances de bruit, et $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{K+1})$ l'ensemble des points de rupture. L'estimation des paramètres se fera par maximum de vraisemblance. L'indépendance conditionnelle des données est supposée. Ainsi, d'après le modèle défini dans l'équation (1), on peut montrer que dans chaque segment k, $\mathbf{x}_i$ a une distribution gaussienne de moyenne $\boldsymbol{\beta}_k^T \mathbf{r}_i$ et de variance σ_k^2 et, par conséquent, que la log-vraisemblance du vecteur paramètre $(\boldsymbol{\psi}, \boldsymbol{\gamma})$ caractérisant le modèle de régression polynomiale par morceaux est la somme des log-vraisemblances locales sur les K segments comme suit :

$$\begin{aligned} L(\boldsymbol{\psi}, \boldsymbol{\gamma}, \mathbf{x}) &= \prod_{k=1}^{K} p(\mathbf{x}_k, \boldsymbol{\psi}, \boldsymbol{\gamma}, \mathbf{x}) \\ \ln(L(\boldsymbol{\psi}, \boldsymbol{\gamma}, \mathbf{x})) &= \sum_{k=1}^{K} ln(p(\mathbf{x}_k, \boldsymbol{\psi}, \boldsymbol{\gamma}, \mathbf{x})) \\ &= \sum_{k=1}^{K} ln(\prod_{t=\gamma_K+1}^{\gamma_{K+1}} \mathcal{N}(\mathbf{x}_t, \boldsymbol{\beta}_k^T \mathbf{r}_i, \sigma_k^2)) \\ &= \sum_{k=1}^{K} ln(\prod_{t \in I_k} \mathcal{N}(\mathbf{x}_t, \boldsymbol{\beta}_k^T \mathbf{r}_i, \sigma_k^2)) \\ &= \sum_{k=1}^{K} \sum_{t \in I_k} ln(\mathcal{N}(\mathbf{x}_t, \boldsymbol{\beta}_k^T \mathbf{r}_i, \sigma_k^2)) \\ &= \sum_{k=1}^{K} \sum_{t \in I_k} ((-\frac{1}{2}))(\frac{\mathbf{x}_t - \boldsymbol{\beta}_k^T \mathbf{r}_i}{\sigma_k})^2 + \ln(\sigma_k^2)) + Constante \end{aligned} \tag{2}$$

Maximiser la log-vraisemblance revient à minimiser le critère ci-dessous selon les paramètres $\boldsymbol{\psi}$ et $\boldsymbol{\gamma}$:

$$J(\boldsymbol{\psi}, \boldsymbol{\gamma}) = \sum_{k=1}^{K} \left[\frac{1}{\sigma_k^2} \sum_{t \in I_k} (\mathbf{x}_t - \boldsymbol{\beta}_k^T \mathbf{r}_i)^2 + n_k \ln(\sigma_k^2) \right] \tag{3}$$

où n_k est le nombre d'éléments dans le segment k.

Comme le critère J est additif sur les K segments, on peut utiliser l'algorithme de Fisher (Fisher, 1958 ; Lechevallier, 1990), procédé de programmation dynamique (Bellman, 1961 ; Brailovsky et Kempner, 1992) pour réaliser la minimisation globale. Ce procédé dynamique a une complexité en temps en $O(Kp^2n^2)$; ce qui peut être coûteux en temps machine dans le cadre de jeux de données volumineux.

2.3 Approximation et segmentation d'une série temporelle par la méthode de régression par morceaux

Une fois les paramètres estimés, la segmentation d'une série temporelle, qui peut être représentée par le vecteur des classes : $\widehat{\mathbf{z}} = (\widehat{z}_1, ..., \widehat{z}_n)$, où $\widehat{z}_i \in \{1, ..K\}$, peut être déduit en posant $\widehat{z}_i = k$ si $i \in (\widehat{\boldsymbol{\gamma}}_K, \widehat{\boldsymbol{\gamma}}_{K+1}]$; les paramètres $(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\gamma}})$ ayant été obtenus par le procédé de programmation dynamique. Puis une approximation de la série temporelle est donnée par $\widehat{\mathbf{x}}_i = \sum_{k=1}^{K} \widehat{z}_{ik} \widehat{\boldsymbol{\beta}}_k^T \mathbf{r}_i$ où $\widehat{z}_{ik} = 1$ si $\widehat{z}_i = k$ et $\widehat{z}_{ik} = 0$ sinon. La formulation vectorielle de la série temporelle approchée $\widehat{\mathbf{x}}$ peut s'écrire :

$$\widehat{\mathbf{x}} = \sum_{k=1}^{K} \widehat{\mathbf{z}}_k \widehat{\boldsymbol{\beta}}_k \tag{4}$$

où $\widehat{\mathbf{z}}_k$ est une matrice diagonale dont les éléments diagonaux sont $(\widehat{z}_1, ..., \widehat{z}_n)$ et

$$\mathbf{T} = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^p \\ 1 & t_2 & t_2^2 & \cdots & t_2^p \\ \vdots & \vdots & \vdots & \vdots & \\ & t_n & t_n^2 & \cdots & t_n^p \end{pmatrix}$$

est la matrice de régression de dimension $[n \times (p+1)]$.