



UNIVERSITÉ
CAEN
NORMANDIE

UNIVERSITÉ DE CAEN NORMANDIE

Régression à l'aide du modèle de Markov caché à effets mixtes

*Auteurs : S. Blin A. Bourjal
C. Champarou
M2 Statistiques Appliquées et
Analyse Décisionnelle*

Tuteur projet : M. F. CHAMROUKHI

Année universitaire 2018-2019

Introduction

Les modèles de Markov cachés (HMM) constituent une classe de modèles de données latentes appropriés aux données séquentielles. Ils sont largement utilisés dans de nombreux domaines d'application, notamment la reconnaissance vocale, l'analyse d'images, la prédiction de séries temporelles

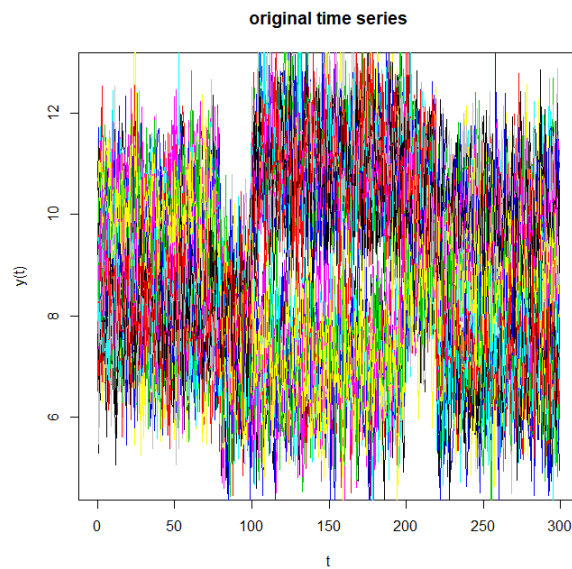


Table des matières

1	Paramètre	3
2	Estimation	3
2.1	Estimation par Maximum de vraisemblance	3
2.2	Estimation par l'Algorithme EM	4
2.2.1	Etape E	4
2.2.2	Etape M	6
3	Application	7

Modèle

Dans un HMM, l'observation la séquence (ou une série chronologique) $\mathbf{y}_i = (\mathbf{y}_{i1}, \dots, \mathbf{y}_{im})$ est supposée être régie par une séquence d'états cachés $\mathbf{z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{im})$ où la variable aléatoire discrète $z_{ij} = \{1, \dots, R\}$ représente l'état non observé associé à \mathbf{y}_{ij} à l'instant t_1 . La séquence d'état \mathbf{z} est généralement supposée être une chaîne de Markov homogène du premier ordre, c'est-à-dire que l'état actuel, car la séquence d'états précédente ne dépend que de l'état précédent. Formellement nous avons :

$$\mathbf{p}(\mathbf{z}_{ij} | \mathbf{z}_{i,j-1}, \mathbf{z}_{i,j-2}, \dots, \mathbf{z}_{i1}) = \mathbf{p}(\mathbf{z}_{ij} | \mathbf{z}_{i,j-1}) \forall j > 1 \quad (1)$$

Le modèle proposé suppose que chaque série chronologique y_i est issu d'un des \mathbf{K} clusters où, au sein de chaque cluster $\mathbf{k} (\mathbf{k} = 1, \dots, \mathbf{K})$, chaque série temporelle est générée par \mathbf{R} régimes polynomiaux non observés. La transition d'un régime à un autre est régi par un Markov homogène Chaîne de premier ordre. Formellement, la distribution d'une fois la série y_i est définie par le mélange conditionnel suivant densité :

$$f(\mathbf{y}_i | \mathbf{t}; \Psi) = \sum_{k=1}^K \alpha_k f_k(\mathbf{y}_i | \mathbf{t}; \Psi_k), \quad (2)$$

où chaque composante de densité $f_k(\cdot)$ associée à la \mathbf{k}_{th} cluster est un modèle de régression polynomial HMM. Dans ce contexte de clustering avec régression HMM, en fonction du cluster $\mathbf{h}_i = \mathbf{k}$, la série chronologique $\mathbf{y}_i = (\mathbf{y}_{i1}, \dots, \mathbf{y}_{im})$ est supposée être généré par le modèle de régression suivant :

$$\mathbf{y}_{ij} = \beta_{kx_{ij}}^T \mathbf{t}_j + \sigma_{kx_{ij}} \epsilon_{ij} \quad (j = 1, \dots, m) \quad (3)$$

1 Paramètre

β_{kr} est le vecteur des coefficients dimensionnels $(\mathbf{p} + 1)$ du r ième modèle de régression polynomiale du groupe \mathbf{k}

σ_{kr}^2 est sa variance de bruit associée et les σ_{ij} sont des variables aléatoires indépendantes réparties selon une distribution gaussienne à moyenne nulle et à variance unitaire. La séquence d'état cachée $\mathbf{z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{im})$ est supposée être une chaîne de Markov de paramètres $(\mathbf{p}_k, \mathbf{A}_k)$.

Le modèle proposé est illustré par la représentation graphique de la figure 1. Chaque composant la densité est donc paramétrée par le vecteur paramètre

$$\phi_k = (\mathbf{p}_k, \mathbf{A}_k, \beta_{k1}, \dots, \beta_{kR}, \sigma_{k1}^2, \dots, \sigma_{kR}^2) \quad (4)$$

2 Estimation

2.1 Estimation par Maximum de vraisemblance

L'estimation des paramètres est réalisée en maximisant la log des données observées - probabilité de Ψ :

$$\begin{aligned}
\mathcal{L}(\Psi) &= \log p(\mathbf{y}_1, \dots, \mathbf{y}_n | \mathbf{t}; \Psi) \\
&= \log \prod_{i=1}^n p(\mathbf{y}_i | \mathbf{t}; \Psi) \\
&= \sum_{i=1}^n \log \sum_{k=1}^K \alpha_k \sum_{x_i} p(z_{i1} \pi_k) \prod_{j=2}^m p(\mathbf{z}_{ij} | \mathbf{z}_{i,j-1}; \mathbf{A}_k) \prod_{j=1}^m \mathcal{N}(\mathbf{y}_{ij}; \beta_{kx_{ij}}^T \mathbf{t}_j, \sigma_{kx_{ij}}^2)
\end{aligned} \tag{5}$$

La maximisation de ce log-vraisemblance ne peut pas être effectuée sous une forme fermée. Nous le maximisons de manière itérative en utilisant un algorithme EM dédié. Avec cette spécification de l'algorithme EM, les données complètes pour le modèle proposé sont constituées de l'ensemble de courbes observées $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$, de leurs étiquettes de grappes correspondantes $\mathbf{h} = (h_1, \dots, h_n)$, et la matrice des étiquettes de régime (état) $\mathbf{h} = (\mathbf{z}_1, \dots, \mathbf{z}_n)$, \mathbf{z}_i étant la séquence d'état cachée associée à \mathbf{y}_i . Les données complètes La probabilité de Ψ est donc donnée par :

$$\begin{aligned}
p(\mathbf{Y}, \mathbf{h}, \mathbf{Z} | \mathbf{t}; \Psi) &= p(\mathbf{h}) p(\mathbf{Y}, \mathbf{Z} | \mathbf{h}, \mathbf{t}; \Psi) \\
&= p(\mathbf{h}) p(\mathbf{Z} | \mathbf{h}, \mathbf{t}; \Psi) p(\mathbf{Y} | \mathbf{h}, \mathbf{Z}, \mathbf{t}; \Psi)
\end{aligned} \tag{6}$$

2.2 Estimation par l'Algorithme EM

Dans cette etape on Calcule la log-vraisemblance complète attendue pour les données complètes :

$$\begin{aligned}
\mathcal{L}_c(\Psi) &= \log p(\mathbf{Y}, \mathbf{h}, \mathbf{Z} | \mathbf{t}; \Psi) = \sum_{k=1}^K \left[\sum_{i=1}^n h_{ik} \log \alpha_k + \sum_{i=1}^n \sum_{r=1}^R h_{ik} \mathbf{z}_{i1kr} \log \pi_{kr} \right. \\
&\quad + \sum_{i=1}^n \sum_{j=2}^m \sum_{l=1}^R h_{ik} \mathbf{z}_{ijkr} \mathbf{z}_{i(j-1)kr} \log \mathbf{A}_{klr} \\
&\quad \left. + \sum_{i=1}^n \sum_{j=1}^m \sum_{r=1}^R h_{ik} \mathbf{z}_{ijkr} \mathcal{N}(\mathbf{y}_{ij}; \beta_{kr}^T \mathbf{t}_j, \sigma_{kr}^2) \right]
\end{aligned} \tag{7}$$

2.2.1 Etape E

Calculer la loglikistence attendue pour les données complètes en fonction de la série temporelle \mathbf{Y} , du vecteur temporel \mathbf{t} et de la valeur actuelle du paramètre Ψ , notée $\Psi^{(q)}$:

$$Q(\Psi, \Psi^{(q)}) = \mathbb{E}[\mathcal{L}_c(\Psi) | \mathbf{Y}, \mathbf{t}; \Psi^{(q)}] \tag{8}$$

Ce qui nous donne ensuite :

$$Q(\Psi, \Psi^{(q)}) = Q_1(\alpha_k) + \sum_{k=1}^K [Q_2(\pi_k, \mathbf{A}_k) + Q_3(\beta_{kr}, \sigma_{kr}^2)], \tag{9}$$

où

$$Q_1(\alpha_k) = \sum_{k=1}^K \sum_{i=1}^n \tau_{ik}^{(q)} \log \alpha_k,$$

$$Q_2(\pi_k, \mathbf{A}_k) = \sum_{r=1}^R \sum_{i=1}^n \tau_{ik}^{(q)} [\gamma_{ijk}^{(q)} \log \pi_{kr} + \sum_{j=2}^m \sum_{l=1}^R \epsilon_{ijkl}^{(q)} \log \mathbf{A}_{klr}], \quad (10)$$

$$Q_3(\beta_{kr}, \sigma_{kr}^2) = \sum_{r=1}^R \sum_{i=1}^n \sum_j^m \tau_{ik}^{(q)} \gamma_{ijk}^{(q)} \log \mathcal{N}(\mathbf{y}_{ij}; \beta_{kr}^T \mathbf{t}_j$$

où

- $\tau_{ik}^{(q)} = \mathbf{p}(\mathbf{h}_i = \mathbf{k} | \mathbf{y}_i, \mathbf{t}; \Psi^q)$ qui est la probabilité à postériori de la classe k ;
- $\gamma_{ijk}^{(q)} = \mathbf{p}(z_{ijk} = r | \mathbf{y}_i, \mathbf{t}; \Psi_k^q)$ qui est la probabilité à postériori du k th régime polynomiale pour la k th classe,
- $\epsilon_{ijkl}^{(q)} = \mathbf{p}(z_{ijk} = r, z_{i(j-1)k} = l | \mathbf{p}_i, \mathbf{t}; \Psi_k^q)$ qui est la probabilité jointe d'avoir le régime r à un temps \mathbf{t}_j et le régime l à un temps \mathbf{t}_{j-1} dans la classe k

Ainsi, dans l'équation ci-dessus, il nous faut calculer les probabilités à postériori de $\epsilon_{ijkl}^{(q)}$, $\gamma_{ijk}^{(q)}$ et $\tau_{ik}^{(q)}$. Ces dernier calculer par les récursions du forward-backward.

Forward-Backward

Le processus du forward calcul recursivement les probabilités :

$$a_{ijk} = \mathbf{p}(\mathbf{y}_{i1}, \dots, \mathbf{y}_{ij}, z_{ijk} = r | \mathbf{t}; \Psi_k) \quad (11)$$

Dans un autre temps, le processus du backward calcul les probabilités à partir des temps $(\mathbf{t} + 1, \dots, n)$

$$b_{ijk} = \mathbf{p}(\mathbf{y}_{i,j+1}, \dots, \mathbf{y}_{im} | z_{ijk} = r | \mathbf{t}; \Psi_k) \quad (12)$$

Ce qui nous permet de calculer les probabilités de $\gamma_{ijk}^{(q)}$ et $\epsilon_{ijkl}^{(q)}$.

$$\gamma_{ijk}^{(q)} = \frac{a_{ijk}^{(q)} b_{ijk}^{(q)}}{\sum_{l=1}^R a_{ijkl}^{(q)} b_{ijkl}^{(q)}} \quad (13)$$

et

$$\epsilon_{ijkl}^{(q)} = \frac{a_{i(j-1)l}^{(q)} \mathbf{A}_{klr}^{(q)} \mathcal{N}(\mathbf{y}_{ij}; \beta_{kr}^{(q)T} \mathbf{t}_j, \sigma_{kr}^{(q)2}) b_{ijk}^{(q)}}{\sum_{r,l=1}^R a_{i(j-1)lk}^{(q)} \mathbf{A}_{klr}^{(q)} \mathcal{N}(\mathbf{y}_{ij}; \beta_{kr}^{(q)T} \mathbf{t}_j, \sigma_{kr}^{(q)2}) b_{ijk}^{(q)}} \quad (14)$$

La probabilité à postériori de la classe $\tau_{ik}^{(q)}$ que la série temporelle \mathbf{y}_i appartient à la classe k est donné quant à elle par :

$$\tau_{ik}^{(q)} = \frac{\alpha_k^{(q)} f_k(\mathbf{y}_i | \mathbf{t}; \boldsymbol{\Psi}_k^{(q)})}{\sum_{k'=1}^K \alpha_{k'}^{(q)} f_{k'}(\mathbf{y}_i | \mathbf{t}; \boldsymbol{\Psi}_{k'}^{(q)})} \quad (15)$$

2.2.2 Etape M

Dans cette étape, la valeur du paramètre est mis à jour en maximisant la valeur loglikelihood des données complètes attendue par rapport à, à savoir :

$$\boldsymbol{\Psi}^{(q+1)} = \arg \max_{\boldsymbol{\Psi}} Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)}) \quad (16)$$

La maximisation de Q peut être réalisé en maximisant les fonctions Q_1 , Q_2 et Q_3 . La maximisation de Q_1 avec α_k est l'un des modèles mixte les plus standard. Les mises à jours étant donné par :

$$\alpha_k^{(q+1)} = \frac{\sum_{i=1}^n \tau_{ik}^{(q)}}{n} \quad (17)$$

La maximisation de Q_2 se fait grâce aux paramètres (π_k, \mathbf{A}) qui correspond à un version pondéré de la mise à jour des paramètres dans la chaîne de Markov dans le modèle standard de HMM. Ces paramètres deviennent ainsi :

$$\pi_{kr}^{(q+1)} = \frac{\sum_{i=1}^n \tau_{ik}^{(q)} \gamma_{i1kr}^{(q)}}{\sum_{i=1}^n \tau_{ik}^{(q)}} \quad (18)$$

et

$$\mathbf{A}_{klr}^{(q+1)} = \frac{\sum_{i=1}^n \sum_{j=2}^m \tau_{ik}^{(q)} \epsilon_{ijk}^{(q)}}{\sum_{i=1}^n \sum_{j=2}^m \tau_{ik}^{(q)} \gamma_{ijk}^{(q)}} \quad (19)$$

La maximisation de Q_3 se fait grâce aux paramètres de la regression β_{kr} pour $\mathbf{k} = 1, \dots, \mathbf{K}$ et $r = 1, \dots, R$. Le paramètre mis à jour est donné par :

$$\beta_{kr}^{(q+1)} = \left[\mathbf{X}^T \left(\sum_{i=1}^n \tau_{ik}^{(q)} \mathbf{W}_{ikr}^{(q)} \right) \mathbf{X} \right]^{-1} \mathbf{X}^T \left(\sum_{i=1}^n \tau_{ik}^{(q)} \mathbf{W}_{ikr}^{(q)} \mathbf{y}_i \right) \quad (20)$$

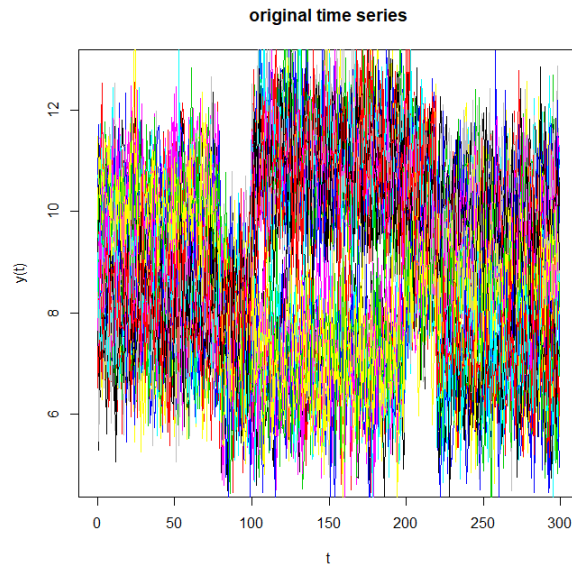
où $\mathbf{W}_{ikr}^{(q)}$ est une matrice diagonale $m \times m$ avec pour éléments diagonals les poids $\{\gamma_{ijk}^{(q)}; j = 1, \dots, m\}$

Ansi, la maximisation de Q_3 en respectant le bruit des variances $\sigma_{kr}^{2(q+1)}$ consiste à la formule :

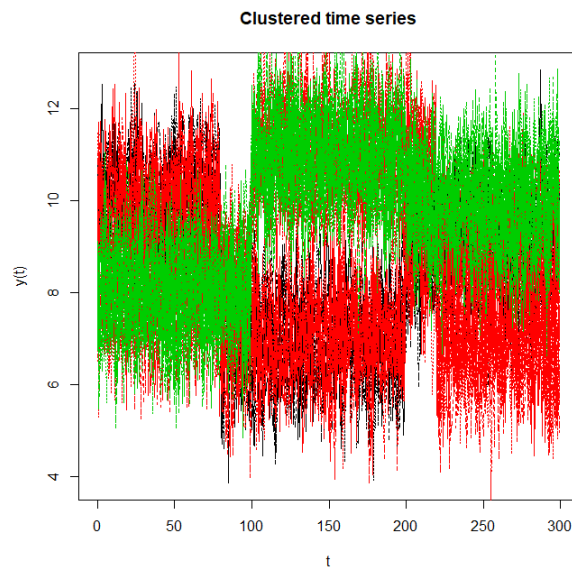
$$\sigma_{kr}^{2(q+1)} = \frac{\sum_{i=1}^n \tau_{ik}^{(q)} \|\sqrt{\mathbf{W}_{ikr}^{(q)}} (\mathbf{y}_i - \mathbf{X} \beta_{kr}^{(q+1)})\|^2}{\sum_{i=1}^n \tau_{ik}^{(q)} \text{trace}(\mathbf{W}_{ikr}^{(q)})} \quad (21)$$

3 Application

Dans cette section, nous présenterons l'utilisation de la régression à l'aide du modèle de Markov à effets mixtes sur des données simulées. Celle-ci sont représenté sur le graphe suivant :

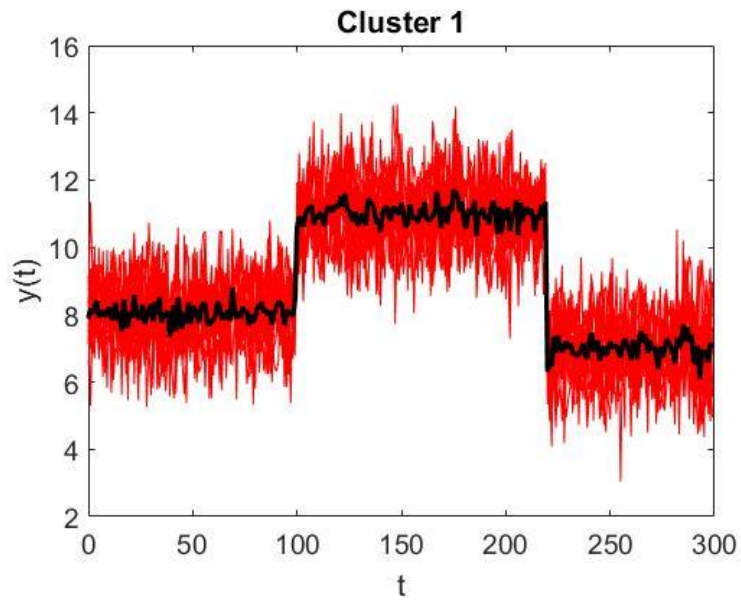


Nous supposons avec ce graphique que les données peuvent être séparées en 3 séries temporelles distinctes (voir graphique ci-dessous).

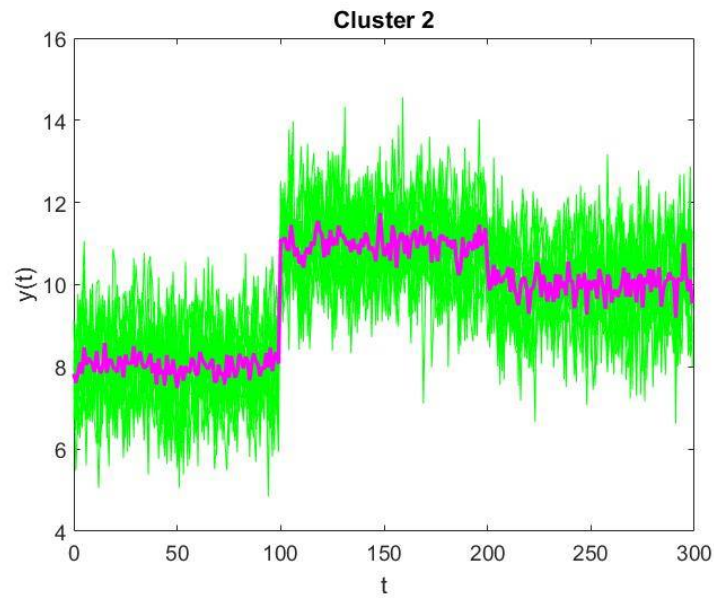


Ensuite, nous séparons les 3 séries temporelles que nous ajusterons séparément grâce à un modèle

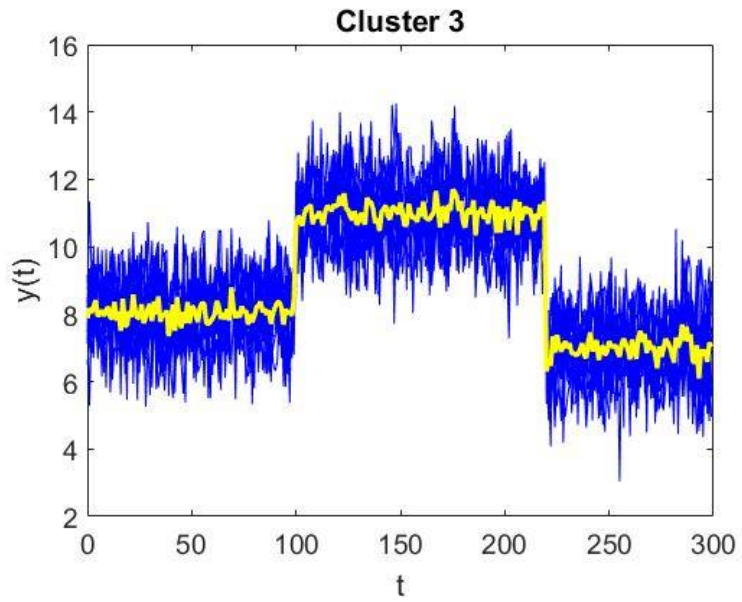
de Markov caché.



Le graphique ci-dessus montre une première série temporelle avec un ajustement avec une régression à l'aide du modèle de Markov caché.



Le graphique ci-dessus montre une seconde série temporelle avec un ajustement avec une régression à l'aide du modèle de Markov caché.



Le graphique ci-dessus montre une troisième série temporelle avec un ajustement avec une régression à l'aide du modèle de Markov caché.