Université de Caen **M2 SAAD**

UFR des Sciences TD/TP Par F. Chamroukhi

TD/TP : Modèle de Markov caché régressif pour une série temporelle

On considère une série temporelle $(Y_1, \ldots, Y_n)$ régie par un processus discret latent $(Z_1, \ldots, Z_n)$ où Z_t représente l'état (discret) à l'instant t ($t = 1, \ldots, n$). $Z_t \in \{1, 2, \ldots, K\}$. On dispose d'une série observée $(y_1, \ldots, y_n)$. L'objectif est de prédire et de segmenter la série sur la base du modèle de Markov Caché régressif suivant :

$$Z_1 \sim \mathcal{M}(1, \boldsymbol{p}) \quad (1)$$

$$Z_t|Z_{t-1} \sim \mathcal{M}(1, \mathbf{A}_{(Z_{t-1},.)}) \quad (2)$$

$$Y_t|Z_t \sim \mathcal{N}(\beta_{z_t0} + \beta_{z_t1}t, \sigma^2_{z_t}) \quad (3)$$

où $\boldsymbol{p}$ et $\mathbf{A}$ sont respectivement la loi initiale et la matrice des transitions de la chaine de Markov cachée, et $\boldsymbol{\beta}_k = (\beta_{k0} + \beta_{k1})^T$ et σ^2_k les paramètres de chaque état (coefficients de régression et variance du bruit). $\mathbf{A}_{(Z_{t-1},.)}$ étant la ligne Z_{t-1} de la matrice $\mathbf{A}$.

Travail demandé :

Partie 1 : Apprentissage

1. Utiliser l'algorithme EM fournie pour l'estimation des paramètres et le tester sur les séries temporelles fournies.
2. Commenter chacun des résultats
3. Remarquer les différences entre les probabilités à posteriori et les probabilités de filtrage
4. Afficher, sur le même graphique, la série temporelle et celle estimée
5. Remarquer les différences entre la série lissée (pondération par les probabilités à posteriori) et la série filtrée (pondération par les probabilités de filtrage)
6. Estimer le nombre d'état optimal en utilisant le critère BIC
7. Comparer à chaque fois qualitativement les résultats de ce modèle et ceux obtenus avec le modèle de régression à processus logistique latent

Partie 2 : Inférence et évaluation

1. On considère maintenant une série temporelle qu'on souhaite segmenter en se basant sur un HMM déjà appris. Utiliser l'algorithme de Viterbi pour segmenter chacune des deux séries `s1_asegmenter` et `s2_asegmenter`.
2. Afficher la série segmentée en colorant les observations de chaque état avec une couleur différente de celles des autres
3. Calculer la log-vraisemblance pour la série donnée

Partie 2 : HMM et régression polynomiale :

Implémenter l'algorithme d'estimation du HMM dans le cas d'une régression d'ordre p pour le modèle défini par l'équation (3).