Université de Rennes 1
Master EEEA (parcours SISEA)

Examen du cours
“Filtre de Kalman et modèles de Markov cachés”
Jeudi 24 janvier 2019, 14:30 à 16:30

Le but de ce problème est de proposer une méthode algébrique, aussi appelée méthode *spectrale*, différente des formules de ré–estimation de Baum–Welsh vues en cours, pour identifier les caractéristiques locales d'un modèle de Markov caché dans le cas *symbolique*.

Dans le principe, il s'agit d'une *méthode de moments* : dans un modèle de Markov caché on sait que la distribution de probabilité jointe des hypothèses et des observations, et en particulier la distribution de probabilité des observations, peut s'exprimer en fonction des caractéristiques locales seulement, et on souhaite inverser ces relations, c'est–à–dire exprimer les caractéristiques locales du modèle de Markov caché en fonction de certains moments bien choisis. Il suffira ensuite de remplacer ces moments théoriques par leur approximation empirique, utilisant les observations disponibles, pour construire un estimateur statistique des caractéristiques locales. Ce but peut être atteint sous certaines conditions restrictives, listées ci–dessous (Conditions A, B et C).

Concrètement, on considère un modèle de Markov caché, avec un ensemble fini S d'hypothèses et des observations prenant leurs valeurs dans un autre ensemble fini O, caractérisé par

- la distribution de probabilité initiale $\pi = (\pi_i \, , \, i \in S)$ vue comme un vecteur–ligne $1 \times K$, avec
$$\pi_i = \mathbb{P}[X_0 = i] \qquad \text{pour tout } i \in S,$$
- la matrice des probabilités de transition $T = (T_{i,j} \, , \, i, j \in S)$ vue comme une matrice carrée $K \times K$, avec
$$T_{i,j} = \mathbb{P}[X_k = j \mid X_{k-1} = i] \qquad \text{pour tout } i, j \in S,$$
- la matrice des probabilités d'emission $B = (B_i^l \, , \, i \in S, l \in O)$ vue comme une matrice rectangulaire $K \times L$, avec
$$B_i^l = \mathbb{P}[Y_k = l \mid X_k = i] \qquad \text{pour tout } i \in S \text{ et tout } l \in O,$$

où les entiers K et L désignent le nombre d'éléments des ensembles finis S et O respectivement. Pour un usage ultérieur, on définit aussi la matrice $B^l = \mathrm{diag}(B_i^l\,,\, i \in S)$ vue comme une matrice diagonale $K \times K$, pour tout $l \in O$.

On introduit les conditions restrictives suivantes

Condition A la matrice T des probabilités de transition laisse invariante la distribution de probabilité initiale π c'est–à–dire que $\pi\, T = \pi$.

C'est une condition implicite : on ne connaît ni π ni T, mais on sait que π et T vérifient la contrainte $\pi\, T = \pi$.

Condition B toutes les composantes de la distribution de probabilité initiale π sont strictement positives, c'est–à–dire que $\pi_i > 0$ pour tout $i \in S$.

Condition C la matrice T des probabilités de transition et la matrice B des probabilités d'emission sont de rang K (donc $K \leq L$ nécessairement).

On rappelle (mais la preuve n'est pas demandée) que si la Condition C est vérifiée, alors

- la matrice T est une matrice carrée $K \times K$ de rang K, donc inversible,
- la matrice B est une matrice rectangulaire $K \times L$ de rang K, donc la matrice $B\, B^*$ est une matrice carrée $K \times K$ de rang K, donc inversible,
- si U est une matrice rectangulaire $L \times K$ de rang K, alors la matrice $B\, U$ est une matrice carrée $K \times K$ de rang K, donc inversible.

Les moments considérés ici sont les probabilités jointes d'une, deux ou trois observations successives, soit

$$P_1^l = \mathbb{P}[Y_k = l] \qquad \text{et} \qquad P_2^{l,q} = \mathbb{P}[Y_k = q, Y_{k-1} = l] \ ,$$

et

$$P_3^{p,l,q} = \mathbb{P}[Y_{k+1} = q, Y_k = l, Y_{k-1} = p] \ ,$$

respectivement, pour tout $p, l, q \in O$.

(i) Montrer (par exemple par récurrence) que si la Condition A est vérifiée, alors la distribution de probabilité de l'hypothèse X_k est π, à tout instant.

(ii) Montrer que si la Condition A est vérifiée, alors

$$P_1^l = \sum_{i \in S} B_i^l\, \pi_i \qquad \textbf{et} \qquad P_2^{l,q} = \sum_{i,j \in S} B_i^l\, B_j^q\, \pi_i\, T_{i,j} \ ,$$

et

$$P_3^{p,l,q} = \sum_{h,i,j \in S} B_h^p \, B_i^l \, B_j^q \, \pi_h \, T_{h,i} \, T_{i,j} \ ,$$

pour tout $p, l, q \in O$**.**

On définit $P_1 = (P_1^l \, , \, l \in O)$ vu comme un vecteur–ligne $1 \times L$, et on définit aussi $P_2 = (P_2^{l,q} \, , \, l, q \in O)$, $P_3^l = (P_3^{p,l,q} \, , \, l, q \in O)$ pour tout $l \in O$, vues comme des matrices $L \times L$.

(iii) Montrer que les expressions composante–par–composante obtenues à la question (ii) peuvent se ré–écrire sous forme matricielle comme

$$P_1 = \pi \, B \ , \qquad P_2 = B^* \operatorname{diag}(\pi) \, T \, B \qquad \textbf{et} \qquad P_3^l = B^* \operatorname{diag}(\pi) \, T \, B^l \, T \, B \ ,$$

pour tout $l \in O$**, et en déduire que**

$$P_3^\bullet = \sum_{l \in O} P_3^l = B^* \operatorname{diag}(\pi) \, T \, T \, B \ .$$

(iv) Montrer que

- **la matrice** B **apparaît comme postfacteur, de dimension** $K \times L$**,**
- **et la matrice** $M = B^* \operatorname{diag}(\pi) \, T$ **apparaît comme préfacteur, de dimension** $L \times K$**,**

dans chacune des expressions matricielles obtenues à la question (iii).

(v) Montrer que

$$P_2 \, B^+ = M \qquad \textbf{et} \qquad P_3^\bullet \, B^+ = M \, T \ ,$$

où la matrice rectangulaire $B^+ = B^* \, (B \, B^*)^{-1}$ **de dimension** $L \times K$ **désigne la matrice inverse généralisée de Moore–Penrose de la matrice rectangulaire** B **de dimension** $K \times L$**.**

(vi) Montrer que si les Conditions B et C sont vérifiées, alors la matrice $P_3^\bullet$ **est une matrice carrée** $L \times L$ **de rang** K**. En déduire que si** U **est une matrice rectangulaire** $L \times K$ **quelconque de rang** K**, alors la matrice** $P_3^\bullet \, U$ **est une matrice rectangulaire** $L \times K$ **de rang** K**, donc la matrice** $(P_3^\bullet \, U)^* \, P_3^\bullet \, U$ **est une matrice carrée** $K \times K$ **de rang** K**, donc inversible.**

On définit la matrice carrée $K \times K$

$$C_3^l = (P_3^\bullet \, U)^+ \, P_3^l \, U \ ,$$

pour tout $l \in O$, où la matrice rectangulaire $(P_3^\bullet \, U)^+ = ((P_3^\bullet \, U)^* \, P_3^\bullet \, U)^{-1} \, (P_3^\bullet \, U)^*$ de dimension $K \times L$ désigne la matrice inverse généralisée de Moore–Penrose de la matrice rectangulaire $P_3^\bullet \, U$ de dimension $L \times K$.

(vii) Montrer que la matrice carrée $T\,B\,U$ de dimension $K \times K$ est inversible. En déduire que

$$P_3^l\,U = P_3^\bullet\,U\,(T\,B\,U)^{-1}\,B^l\,T\,B\,U \ ,$$

et

$$C_3^l = (T\,B\,U)^{-1}\,B^l\,T\,B\,U \ ,$$

pour tout $l \in O$.

On en déduit que la matrice B^l peut être obtenue par diagonalisation de la matrice C_3^l, et qu'une seule et même matrice permet de diagonaliser toutes les matrices C_3^l pour tout $l \in O$.

Le principe de la méthode est donc le suivant : on suppose que les matrices P_3^l sont connues pour tout $l \in O$ (exactement ou sous la forme d'une approximation empirique utilisant les observations disponibles, par exemple

$$\widehat{P}_3^{p,l,q} = \frac{1}{n-1}\,\sum_{k=1}^{n-1} 1(Y_{k+1} = q, Y_k = l, Y_{k-1} = p) \ ,$$

pour tout $p, l, q \in O$), on diagonalise une des matrices C_3^l pour un certain $l \in O$, ou bien une combinaison linéaire des matrices C_3^l pour $l \in O$, ce qui permet de diagonaliser simultanément toutes les matrices C_3^l pour $l \in O$. On obtient ainsi toutes les matrices diagonales B^l pour tout $l \in O$, c'est–à–dire toutes les colonnes de la matrice B et au final la matrice B complète.

On peut désormais supposer que la matrice B est connue (exactement ou en fonction d'approximations empiriques utilisant les observations disponibles). On suppose aussi que le vecteur P_1 et la matrice P_2 sont connus (exactement ou sous la forme d'une approximation empirique utilisant les observations disponibles, par exemple

$$\widehat{P}_1^l = \frac{1}{n+1}\,\sum_{k=0}^{n} 1(Y_k = l) \qquad \text{et} \qquad \widehat{P}_2^{l,q} = \frac{1}{n}\,\sum_{k=1}^{n} 1(Y_k = q, Y_{k-1} = l) \ ,$$

pour tout $l, q \in O$).

(viii) Montrer que la distribution de probabilité initiale π et la matrice T des probabilités de transition, peuvent être obtenues comme

$$\pi = P_1\,B^+ \qquad \textbf{et} \qquad T = (\mathrm{diag}(\pi))^{-1}\,(B^+)^*\,P_2\,B^+ \ .$$

(ix) Discuter les avantages et inconvénients (en terme de conditions nécessaires, de satisfaction des contraintes, de temps de calcul, etc.) de la méthode *spectrale* proposée et de la méthode statistique utilisant les formules de ré–estimation de Baum–Welsh vues en cours.