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.

Solution

Soit p_k la distribution de probabilité de l'hypothèse X_k à l'intant k, vue comme vecteur–ligne $1 \times L$. Par définition, $p_0 = \pi$ de sorte que la propriété est vérifiée à l'instant $k = 0$.

Supposons que la propriété soit vérifiée à l'instant $(k-1)$. Si la Condition A est vérifiée, alors

$$p_k = p_{k-1}\, T = \pi\, T = \pi \ ,$$

de sorte que la propriété est vérifiée à l'instant k.

□

(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 $p, l, q \in O$**.**

Solution

Si la Condition A est vérifiée, alors

$$\begin{aligned} \mathbb{P}[X_k = i] &= \pi_i \ , \\ \mathbb{P}[X_k = j, X_{k-1} = i] &= \pi_i\, T_{i,j} \ , \\ \mathbb{P}[X_{k+1} = j, X_k = i, X_{k-1} = h] &= \pi_h\, T_{h,i}\, T_{i,j} \ , \end{aligned}$$

pour tout $h, i, j \in S$.

D'après la formule de Bayes, on a

$$\mathbb{P}[Y_k = l, X_k = i] = \mathbb{P}[Y_k = l \mid X_k = i]\, \mathbb{P}[X_k = i] = B_i^l\, \pi_i \ ,$$

pour tout $i \in S$ et tout $l \in O$, et en sommant par rapport à $i \in S$ on obtient

$$P_1^l = \mathbb{P}[Y_k = l] = \sum_{i \in S} \mathbb{P}[Y_k = l, X_k = i] = \sum_{i \in S} B_i^l\, \pi_i \ ,$$

pour tout $l \in O$.

D'après la formule de Bayes et l'hypothèse de canal sans mémoire, on a

$$\begin{aligned} &\mathbb{P}[Y_k = q, Y_{k-1} = l, X_k = j, X_{k-1} = i] \\ &\qquad = \mathbb{P}[Y_k = q, Y_{k-1} = l \mid X_k = j, X_{k-1} = i]\, \mathbb{P}[X_k = j, X_{k-1} = i] \\ &\qquad = \mathbb{P}[Y_k = q \mid X_k = j]\, \mathbb{P}[Y_{k-1} = l \mid X_{k-1} = i]\, \mathbb{P}[X_k = j, X_{k-1} = i] \\ &\qquad = B_i^l\, B_j^q\, \pi_i\, T_{i,j} \ , \end{aligned}$$

pour tout $i, j \in S$ et tout $l, q \in O$, et en sommant par rapport à $i, j \in S$ on obtient

$$\begin{aligned} P_2^{l,q} &= \mathbb{P}[Y_k = q, Y_{k-1} = l] \\ &= \sum_{i,j \in S} \mathbb{P}[Y_k = q, Y_{k-1} = l, X_k = j, X_{k-1} = i] \\ &= \sum_{i,j \in S} B_i^l \, B_j^q \, \pi_i \, T_{i,j} \ , \end{aligned}$$

pour tout $l, q \in O$.

D'après la formule de Bayes et l'hypothèse de canal sans mémoire, on a

$$\begin{aligned} &\mathbb{P}[Y_{k+1} = q, Y_k = l, Y_{k-1} = p, X_{k+1} = j, X_k = i, X_{k-1} = h] \\ &\quad = \mathbb{P}[Y_{k+1} = q, Y_k = l, Y_{k-1} = p \mid X_{k+1} = j, X_k = i, X_{k-1} = h] \\ &\qquad\quad \mathbb{P}[X_{k+1} = j, X_k = i, X_{k-1} = h] \\ &\quad = \mathbb{P}[Y_{k+1} = q \mid X_{k+1} = j] \, \mathbb{P}[Y_k = l \mid X_k = i] \, \mathbb{P}[Y_{k-1} = p \mid X_{k-1} = h] \\ &\qquad\quad \mathbb{P}[X_{k+1} = j, X_k = i, X_{k-1} = h] \\ &\quad = B_h^p \, B_i^l \, B_j^q \, \pi_h \, T_{h,i} \, T_{i,j} \ , \end{aligned}$$

pour tout $h, i, j \in S$ et tout $p, l, q \in O$, et en sommant par rapport à $h, i, j \in S$ on obtient

$$\begin{aligned} P_3^{p,l,q} &= \mathbb{P}[Y_{k+1} = q, Y_k = l, Y_{k-1} = p] \\ &= \sum_{h,i,j \in S} \mathbb{P}[Y_{k+1} = q, Y_k = l, Y_{k-1} = p, X_{k+1} = j, X_k = i, X_{k-1} = h] \\ &= \sum_{h,i,j \in S} B_h^p \, B_i^l \, B_j^q \, \pi_h \, T_{h,i} \, T_{i,j} \ , \end{aligned}$$

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

$\square$

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

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

Solution

On définit la matrice $A = B^* \operatorname{diag}(\pi)\, T\, B$, et on vérifie qu'il s'agit bien d'une matrice carrée $L \times L$, dont la composante à l'intersection de la ligne l et de la colonne q s'écrit

$$A^{l,q} = \sum_{i,j \in S} B_i^l\, \pi_i\, T_{i,j}\, B_j^q \ ,$$

et on vérifie ainsi que $A^{l,q} = P_2^{l,q}$ pour tout $l, q \in O$. En d'autres termes

$$P_2 = B^* \operatorname{diag}(\pi)\, T\, B \ .$$

Pour $l \in O$ fixé, on définit la matrice $A_l = B^* \operatorname{diag}(\pi)\, T\, B^l\, T\, B$, et on vérifie qu'il s'agit bien d'une matrice carrée $L \times L$, dont la composante à l'intersection de la ligne p et de la colonne q s'écrit

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

et on vérifie ainsi que $A_l^{p,q} = P_3^{p,l,q}$ pour tout $p, q \in O$. En d'autres termes

$$P_3^l = B^* \operatorname{diag}(\pi)\, T\, B^l\, T\, B \ ,$$

pour tout $l \in O$.

Par définition, la matrice B^l est une matrice diagonale $K \times K$, dont la composante à l'intersection de la ligne i et de la colonne i est

$$B_i^l = \mathbb{P}[Y_k = l \mid X_k = i] \ ,$$

pour tout $l \in O$, et on vérifie que

$$\sum_{l \in O} B_i^l = \sum_{l \in O} \mathbb{P}[Y_k = l \mid X_k = i] = 1 \ .$$

c'est–à–dire que la matrice $\sum_{l \in O} B^l$ est une matrice diagonale $K \times K$, dont la composante à l'intersection de la ligne i et de la colonne i est égale à 1 pour tout $i \in S$. On en déduit que

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

compte tenu que $\sum_{l \in O} B^l = I$.

$\square$

(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).

Solution

Clairement

$$P_2 = M\, B \ , \qquad P_3^l = M\, B^l\, T\, B \qquad \text{et} \qquad P_3^\bullet = M\, T\, B \ ,$$

pour tout $l \in O$.

$\square$

(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 $K \times L$.

Solution

On rappelle que $P_2 = M\, B$ et $P_3^\bullet = M\, T\, B$. Clairement $B\, B^+ = I$, et en multipliant à droite par la matrice B^+, on obtient $P_2\, B^+ = M$ et $P_3^\bullet\, B^+ = M\, T$.

$\square$

(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.

Solution

Si la Condition B est vérifiée, alors la matrice diagonale $\operatorname{diag}(\pi)$ est une matrice carrée $K \times K$ inversible, et si la Condition C est vérifiée, alors

- la matrice T est une matrice carrée $K \times K$ inversible,
- la matrice B est une matrice rectangulaire $K \times L$ de rang K, et il a été rappelé plus haut que dans ce cas, 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.

Compte tenu que les matrices $\text{diag}(\pi)$, T et $B\,U$ sont des matrices carrées $K \times K$ inversibles, on en déduit que la matrice $\text{diag}(\pi)\,T\,T\,B\,U$ est une matrice carrée $K \times K$ inversible. On obtient donc la factorisation

$$P_3^\bullet\, U = B^*\,\text{diag}(\pi)\,T\,T\,B\,U = B^*\,A \ ,$$

où la matrice B^* est une matrice rectangulaire $L \times K$ de rang K, et la matrice A est une matrice carrée $K \times K$ inversible. On en déduit que la matrice $P_3^\bullet\, U$ est une matrice rectangulaire $L \times K$ de rang K : En effet, si v est un vecteur colonne de dimension K tel que $P_3^\bullet\, U\,v = B^*\,A\,v = 0$, alors nécessairement $A\,v = 0$ compte tenu que la matrice B^* est une matrice rectangulaire $L \times K$ de rang K de sorte que les K colonnes de la matrice B^* sont linérairement indépendantes, et nécessairement $v = 0$ compte tenu que la matrice A est une matrice carrée $K \times K$ inversible. En d'autres termes, les K colonnes de la matrice $P_3^\bullet\, U$ sont linérairement indépendantes, c'est–à–dire que la matrice $P_3^\bullet\, U$ est une matrice rectangulaire $L \times K$ de rang K.

□

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 $l \in O$.

Solution

Si la Condition C est vérifiée, alors

- la matrice T est une matrice carrée $K \times K$ inversible,

- la matrice B est une matrice rectangulaire $K \times L$ de rang K, et il a été rappelé plus haut que dans ce cas, 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.

Compte tenu que les matrices T et $B\,U$ sont des matrices carrées $K \times K$ inversibles, on en déduit que la matrice $T\,B\,U$ est une matrice carrée $K \times K$ inversible.

En utilisant les expressions matricielles obtenues aux questions (iii) et (iv), on a

$$\begin{aligned} P_3^l U &= M\,B^l\,T\,B\,U \\ &= M\,T\,B\,U\,(T\,B\,U)^{-1}\,B^l\,T\,B\,U \\ &= P_3^{\bullet}\,U\,(T\,B\,U)^{-1}\,B^l\,T\,B\,U \ , \end{aligned}$$

de sorte que

$$\begin{aligned} C_3^l &= (P_3^{\bullet}\,U)^+\,P_3^l\,U \\ &= (P_3^{\bullet}\,U)^+\,P_3^{\bullet}\,U\,(T\,B\,U)^{-1}\,B^l\,T\,B\,U \\ &= (T\,B\,U)^{-1}\,B^l\,T\,B\,U \ , \end{aligned}$$

pour tout $l \in O$, compte tenu que $(P_3^{\bullet}\,U)^+\,P_3^{\bullet}\,U = I$.

$\square$

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

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

Solution

On rappelle que $P_1 = \pi \, B$, et en multipliant à droite par la matrice B^+, inverse généralisée de Moore–Penrose de la matrice B, on obtient $\pi = P_1 \, B^+$.

On rappelle que $P_2 = B^* \, \mathrm{diag}(\pi) \, T \, B$, et en multipliant à droite par la matrice B^+, inverse généralisée de Moore–Penrose de la matrice B, et à gauche par la matrice transposée $(B^+)^*$, on obtient $\mathrm{diag}(\pi) \, T = (B^+)^* \, P_2 \, B^+$. Si la Condition B est vérifiée, alors la matrice diagonale $\mathrm{diag}(\pi)$ est inversible, et on en déduit que $T = \mathrm{diag}(\pi)^{-1} \, (B^+)^* \, P_2 \, B^+$.

□

(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.

Solution

La méthode specrale proposée :

- est une méthode directe dont le temps de calcul est faible et contrôlé,
- requiert que des conditions restrictives soient vérifiées par les caractéristiques locales inconnues :
  - condition de stationnarité (Condition A),
  - condition de positivité stricte (Condition B),
  - condition de rang (Condition C),
- n'offre aucune garantie que les estimations obtenues pour les caractréristiques locales vérifient les contraintes structurelles :

    - composantes toutes positives, et somme des composantes égale à 1, pour l'estimation de la distribution de probabilité initiale,
    - composantes toutes positives, et somme des composantes égale à 1 sur chaque ligne, pour l'estimation de la matrice des probabilités de transition, et pour l'estimation de la matrice des probabilités d'emission,

- n'offre aucune garantie que les estimations obtenues pour la distribution de probabilité initiale et pour la matrice des probabilités de transition vérifient la condition de stationnarité (Condition A).

Pour sa part, la méthode statistique vue en cours :

- est une méthode itérative dont la convergence peut nécessiter un grand nombre, inconnu à l'avance, d'itérations,
- ne requiert aucune condition restrictive sur les caractéristiques locales,
- garantit par construction que les estimations obtenues pour les caractréristiques locales vérifient les contraintes structurelles, à chaque itération.

□

Ce problème est inspiré de l'article suivant :

Daniel Hsu, Sham M. Kakade, and Tong Zhang. A spectral algorithm for learning hidden Markov models. *Journal of Computer and System Sciences*, 78(5 (Special issue on Learning Theory)):1460–1480, September 2012.