Université de Rennes 1
DEA STIR
Signal — Télécommunications — Images — Radar
Option Signal

Examen du cours
“Filtre de Kalman et modèles de Markov cachés”
Jeudi 9 janvier 1997, 8:00 à 9:30

Exercice :

Soit $\{X_k\}$ une chaîne de Markov à espace d'état fini $E = \{1, \cdots, N\}$, caractérisée par les données suivantes :

- *loi initiale* $\nu = (\nu_i)$

$$\nu_i = \mathbf{P}[X_0 = i] \ , \qquad \text{pour tout } i \in E,$$

- *matrice de transition* $\pi = (\pi_{i,j})$

$$\pi_{i,j} = \mathbf{P}[X_{k+1} = j \mid X_k = i] \ , \qquad \text{pour tout } i, j \in E.$$

Soit $\{Y_k\}$ une suite d'observations, telle que :

$$Y_k = h(X_k) + V_k \ ,$$

où la suite $\{V_k\}$ est un bruit blanc gaussien de dimension d, de moyenne nulle et de matrice de covariance identité, indépendant de la chaîne de Markov $\{X_k\}$. La fonction h définie sur E à valeurs dans $\mathbf{R}^d$ est caractérisée par la donnée d'une famille $h = (h_i)$ de N vecteurs de $\mathbf{R}^d$, et on a

$$\psi_i(y)\, dy = \mathbf{P}[Y_k \in dy \mid X_k = i] = \frac{1}{(\sqrt{2\pi})^d} \exp\Big\{ -\tfrac{1}{2}\, |y - h_i|^2 \Big\}\, dy \ ,$$

pour tout $i \in E$, et tout $y \in \mathbf{R}^d$.

Le but de cet exercice est de proposer un algorithme d'estimation pour le vecteur des moyennes (h_i).

On considère dans un premier temps le cas où $\{X_0, \cdots, X_n\}$ et $\{Y_0, \cdots, Y_n\}$ sont observés. La fonction de log–vraisemblance pour l'estimation de $h = (h_i)$, à partir de l'observation de $\{X_0, \cdots, X_n\}$ et $\{Y_0, \cdots, Y_n\}$, s'écrit

$$\ell(h) = -\tfrac{1}{2} \sum_{k=0}^{n} |Y_k - h(X_k)|^2 \ .$$

(i) **En utilisant la décomposition**

$$h(X_k) = \sum_{i \in E} h_i \, \mathbf{1}_{[X_k = i]} \ ,$$

montrer que

$$\ell(h) = -\tfrac{1}{2} \sum_{i \in E} \sum_{k=0}^{n} \mathbf{1}_{[X_k = i]} \, |Y_k - h_i|^2 \ .$$

(ii) **En maximisant $\ell(h)$ par rapport au vecteur h_i, montrer que l'estimateur du maximum de vraisemblance pour la moyenne h_i, à partir de l'observation de $\{X_0, \cdots, X_n\}$ et $\{Y_0, \cdots, Y_n\}$, est donné par**

$$\widehat{h}_i = \frac{\displaystyle\sum_{k=0}^{n} Y_k \, \mathbf{1}_{[X_k = i]}}{\displaystyle\sum_{k=0}^{n} \mathbf{1}_{[X_k = i]}} \ .$$

On considère ensuite le cas où seulement $\{Y_0, \cdots, Y_n\}$ est observé. On pose

$$N_i = \sum_{k=0}^{n} \mathbf{1}_{[X_k = i]} \qquad \text{et} \qquad O_i = \sum_{k=0}^{n} Y_k \, \mathbf{1}_{[X_k = i]} \ .$$

(iii) **En supposant le modèle $\mathbf{M} = (\nu, \pi, h)$ connu, donner l'expression des estimateurs $\widehat{N}_i = \mathbf{E}[N_i \mid \mathcal{Y}_n]$ et $\widehat{O}_i = \mathbf{E}[O_i \mid \mathcal{Y}_n]$ à l'aide des variables forward et backward correspondantes.**

(iv) **Proposer un algorithme itératif pour l'estimation de la moyenne h_i à partir de l'observation de $\{Y_0, \cdots, Y_n\}$ seulement.**

Problème :

Le but de ce problème est d'établir les équations de l'estimateur du maximum a posteriori dans le cas linéaire gaussien, et de montrer que cet estimateur coïncide avec le filtre de Kalman.

On considère le système linéaire suivant :

$$X_{k+1} = F\, X_k + W_k \ , \qquad (\text{dans } \mathbf{R}^m)$$

$$Y_k = H\, X_k + V_k \ , \qquad (\text{dans } \mathbf{R}^d)$$

avec les hypothèses habituelles :

- X_0 *variable aléatoire gaussienne, de moyenne* μ_0 *et de matrice de covariance* Σ_0*.*
- $\{W_k\}$ *bruit blanc gaussien centré de matrice de covariance* Q*.*
- $\{V_k\}$ *bruit blanc gaussien centré de matrice de covariance* R *inversible.*
- X_0*,* $\{W_k\}$ *et* $\{V_k\}$ *mutuellement indépendants.*

Pour simplifier la dérivation, on suppose en outre que les matrices de covariance Σ_0 et Q sont *inversibles.*

(i) **Montrer que la loi conditionnelle de X_k sachant $(X_0, \cdots, X_{k-1})$ est donnée par**

$$\mathbf{P}[X_k \in dx \mid X_0 = x_0, \cdots, X_{k-1} = x_{k-1}] =$$

$$= \frac{1}{(\sqrt{2\pi})^m \sqrt{\det Q}} \ \exp\Big\{ -\tfrac{1}{2}\, (x - F\, x_{k-1})^* \, Q^{-1}\, (x - F\, x_{k-1}) \Big\} \ dx \ .$$

En déduire que la suite $\{X_k\}$ est une chaîne de Markov à valeur dans $\mathbf{R}^m$.

(ii) **Déduire de (i) que la loi jointe de $(X_0, \cdots, X_n)$ est donnée par**

$$\mathbf{P}[X_0 \in dx_0, \cdots, X_n \in dx_n] =$$

$$= \frac{1}{(\sqrt{2\pi})^m \sqrt{\det \Sigma_0}} \ \Big[\frac{1}{(\sqrt{2\pi})^m \sqrt{\det Q}} \Big]^n$$

$$\exp\Big\{ -\tfrac{1}{2}\, (x_0 - \mu_0)^* \, \Sigma_0^{-1}\, (x_0 - \mu_0) - \tfrac{1}{2} \sum_{k=1}^{n} (x_k - F\, x_{k-1})^* \, Q^{-1}\, (x_k - F\, x_{k-1}) \Big\}$$

$$dx_0 \; dx_1 \cdots dx_n \ .$$

(iii) **Montrer que la propriété de canal sans mémoire est vérifiée, et que**

$$\mathbf{P}[Y_k \in dy \mid X_k = x] = \frac{1}{(\sqrt{2\pi})^d \sqrt{\det R}} \; \exp\Big\{ -\tfrac{1}{2}\,(y - H\,x)^*\, R^{-1}\,(y - H\,x) \Big\} \; dy \; .$$

(iv) **Déduire de (ii) et (iii) que la loi jointe de** $(X_0, \cdots, X_n, Y_0, \cdots, Y_n)$ **est donnée par**

$$\mathbf{P}[X_0 \in dx_0, \cdots, X_n \in dx_n, Y_0 \in dy_0, \cdots, Y_n \in dy_n] =$$

$$= \frac{1}{(\sqrt{2\pi})^m \sqrt{\det \Sigma_0}} \; \Big[\frac{1}{(\sqrt{2\pi})^m \sqrt{\det Q}} \Big]^n \; \Big[\frac{1}{(\sqrt{2\pi})^d \sqrt{\det R}} \Big]^{n+1}$$

$$\exp\Big\{ -\tfrac{1}{2}\,(x_0 - \mu_0)^*\, \Sigma_0^{-1}\,(x_0 - \mu_0) - \tfrac{1}{2}\, \sum_{k=1}^{n} (x_k - F\,x_{k-1})^*\, Q^{-1}\,(x_k - F\,x_{k-1})$$

$$- \tfrac{1}{2}\, \sum_{k=0}^{n} (y_k - H\,x_k)^*\, R^{-1}\,(y_k - H\,x_k) \Big\}$$

$$dx_0\; dx_1 \cdots dx_n\; dy_0\; dy_1 \cdots dy_n \; .$$

Par définition, l'estimateur du maximum a posteriori pour la suite $\{X_0, \cdots, X_n\}$ est la suite d'états $(x_0, \cdots, x_n)$ qui maximise la densité conditionnelle de $(X_0, \cdots, X_n)$ sachant $(Y_0, \cdots, Y_n)$.

(v) **Montrer que maximiser par rapport à** $(x_0, \cdots, x_n)$ **la densité conditionnelle de** $(X_0, \cdots, X_n)$ **sachant** $(Y_0, \cdots, Y_n)$ **donne le même résultat que maximiser par rapport à** $(x_0, \cdots, x_n)$ **la densité jointe de** $(X_0, \cdots, X_n, Y_0, \cdots, Y_n)$**.**

On se propose donc de minimiser par rapport à $(x_0, \cdots, x_n)$ la fonction suivante

$$\begin{aligned} J_n(x_0, \cdots, x_n) \;=\; & \tfrac{1}{2}\,(x_0 - \mu_0)^*\, \Sigma_0^{-1}\,(x_0 - \mu_0) \\ & + \tfrac{1}{2}\, \sum_{k=1}^{n} (x_k - F\,x_{k-1})^*\, Q^{-1}\,(x_k - F\,x_{k-1}) \\ & + \tfrac{1}{2}\, \sum_{k=0}^{n} (Y_k - H\,x_k)^*\, R^{-1}\,(Y_k - H\,x_k) \; . \end{aligned}$$

On définit

$$V_n(x) = \min_{x_0, \cdots, x_{n-1}} \; J_n(x_0, \cdots, x_{n-1}, x) \; .$$

L'objectif de la fin du problème est de montrer par récurrence que

$$V_n(x) = \tfrac{1}{2}\,(x - \mu_n)^*\, \Sigma_n^{-1}\,(x - \mu_n) + c_n \; , \qquad (\star)$$

et de caractériser la matrice Σ_n, le vecteur μ_n, et le scalaire c_n.

(vi) **Déduire de ($\star$) que μ_n est l'estimateur du maximum a posteriori pour l'état X_n.**

(vii) **Montrer que**

$$J_{k+1}(x_0, \cdots, x_k, x) = J_k(x_0, \cdots, x_k) + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k)$$

$$+ \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x)\ .$$

(viii) **En prenant le minimum par rapport à $(x_0, \cdots, x_k)$ et en utilisant l'hypothèse de récurrence ($\star$) supposée vraie à l'instant k, montrer que**

$$V_{k+1}(x) = \min_{x_k} \Big[\ \tfrac{1}{2}\,(x_k - \mu_k)^*\,\Sigma_k^{-1}\,(x_k - \mu_k) + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k)\ \Big]$$

$$+ \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x) + c_k\ .$$

(ix) **Effectuer la minimisation par rapport à x_k, et montrer que**

$$V_{k+1}(x) = \tfrac{1}{2}\,(x - F\,\mu_k)^*\,(F\,\Sigma_k\,F^* + Q)^{-1}\,(x - F\,\mu_k)$$

$$+ \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x) + c_k\ .$$

On introduit les notations suivantes :

$$\mu_{k+1}^- = F\,\mu_k \qquad \text{et} \qquad \Sigma_{k+1}^- = F\,\Sigma_k\,F^* + Q\ ,$$

ce qui donne

$$V_{k+1}(x) = \tfrac{1}{2}\,(x - \mu_{k+1}^-)^*\,(\Sigma_{k+1}^-)^{-1}\,(x - \mu_{k+1}^-)$$

$$+ \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x) + c_k\ .$$

(x) **Montrer que**

$$V_{k+1}(x) = \tfrac{1}{2}\,(x - \mu_{k+1})^*\,\Sigma_{k+1}^{-1}\,(x - \mu_{k+1}) + c_{k+1}\ ,$$

c'est–à–dire que la propriété ($\star$) est vérifiée, avec en particulier

$$\mu_{k+1} = \mu_{k+1}^- + \Sigma_{k+1}^-\,H^*\,(H\,\Sigma_{k+1}^-\,H^* + R)^{-1}\,(Y_{k+1} - H\,\mu_{k+1}^-)\ .$$

Comparer ces équations avec celles du filtre de Kalman, et conclure.