Université de Rennes 1
Master Recherche SISEA

Examen du cours UE S3-2
“Filtre de Kalman et modèles de Markov cachés”
Jeudi 11 février 2016, 14:00 à 16:00

Problème

Le but de ce problème est d'établir des formules de ré–estimation pour les paramètres d'un système linéaire gaussien, en utilisant le même *principe* que celui utilisé pour établir les formules de ré–estimation de Baum–Welsh pour les paramètres d'un modèle de Markov caché.

On considère une suite d'états cachés $(X_0, X_1, \cdots, X_n)$ à valeurs dans $\mathbb{R}^m$, vérifiant

$$X_k = F\, X_{k-1} + W_k \ ,$$

et une suite d'observations $(Y_0, Y_1, \cdots, Y_n)$ à valeurs dans $\mathbb{R}^d$, vérifiant

$$Y_k = H\, X_k + V_k \ ,$$

et on suppose que

- la condition initiale X_0 est un vecteur aléatoire gaussien, de moyenne μ et de matrice de covariance Σ *inversible*,
- la suite $(W_1, \cdots, W_n)$ est un bruit blanc gaussien, de matrice de covariance Q *inversible*,
- la suite $(V_0, V_1, \cdots, V_n)$ est un bruit blanc gaussien, de matrice de covariance R *inversible*,
- la condition initiale X_0 et les suites $(W_1, \cdots, W_n)$ et $(V_0, V_1, \cdots, V_n)$ sont mutuellement indépendants.

Il s'agit d'estimer le vecteur μ, les matrices F et H, et les matrices de covariance Σ, Q et R, au vu des observations $(Y_0, Y_1, \cdots, Y_n)$

La première étape consiste à obtenir une expression (possiblement peu efficace d'un point de vue numérique) de la vraisemblance L_n du modèle $\boldsymbol{M} = (\mu, \Sigma, F, Q, H, R)$.

(i) On admet que la suite $(X_0, X_1, \cdots, X_n)$ forme une chaîne de Markov à valeurs dans $\mathbb{R}^m$. Montrer que la loi initiale $\mathbb{P}[X_0 \in dx]$ possède une densité $p_0(x)$ dont on donnera l'expression. Montrer que le noyau de transition $\mathbb{P}[X_k \in dx' \mid X_{k-1} = x]$ possède une densité $p(x' \mid x)$, paramétrée par x, dont on donnera l'expression.

En déduire l'expression de la densité jointe $p_n(x_0, x_1, \cdots, x_n)$ de la suite d'états cachés $(X_0, X_1, \cdots, X_n)$.

(ii) On admet que la suite $(Y_0, \cdots, Y_n)$ est recueillie à travers un canal *sans mémoire*. Montrer que le noyau d'émission $\mathbb{P}[Y_k \in dy \mid X_k = x]$ possède une densité $q(y \mid x)$, paramétrée par x, dont on donnera l'expression.

En déduire l'expression de la densité conditionnelle jointe $p_n(y_0, \cdots, y_n \mid x_0, x_1, \cdots, x_n)$, paramétrée par $(x_0, x_1, \cdots, x_n)$, des observations $(Y_0, \cdots, Y_n)$ sachant les états cachés $(X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n)$.

(iii) En utilisant la formule de Bayes et les réponses aux questions (i) et (ii), donner l'expression de la densité jointe $p_n(x_0, x_1, \cdots, x_n, y_0, \cdots, y_n)$ des états cachés $(X_0, X_1, \cdots, X_n)$ et des observations $(Y_0, \cdots, Y_n)$.

En intégrant par rapport aux variables $(x_0, x_1, \cdots, x_n)$, en déduire une expression de la densité jointe $q_n(y_0, \cdots, y_n)$ des observations $(Y_0, \cdots, Y_n)$.

En utilisant la formule de Bayes, en déduire une expression de la densité conditionnelle jointe $p_n(x_0, x_1, \cdots, x_n \mid y_0, \cdots, y_n)$, paramétrée par $(y_0, \cdots, y_n)$, des états cachés $(X_0, X_1, \cdots, X_n)$ sachant les observations $(Y_0 = y_0, \cdots, Y_n = y_n)$.

Expliquer pourquoi ces expressions sont difficilement utilisables en pratique.

On définit la fonction

$$z_n(x_0, x_1, \cdots, x_n) = p_n(x_0, x_1, \cdots, x_n, Y_0, Y_1, \cdots, Y_n) \ ,$$

obtenue en remplacant les variables muettes $(y_0, \cdots, y_n)$ par les observations $(Y_0, \cdots, Y_n)$, dans l'expression de la densité jointe des états cachés $(X_0, X_1, \cdots, X_n)$ et des observations $(Y_0, \cdots, Y_n)$. En utilisant les réponses obtenues à la question (iii), on en déduit que la densité conditionnelle jointe $p_n(x_0, x_1, \cdots, x_n \mid Y_0, \cdots, Y_n)$ des états cachés $(X_0, X_1, \cdots, X_n)$ sachant les observations $(Y_0, \cdots, Y_n)$ et la vraisemblance L_n peuvent s'exprimer comme

$$p_n(x_0, x_1, \cdots, x_n \mid Y_0, \cdots, Y_n) = \frac{z_n(x_0, x_1, \cdots, x_n)}{\displaystyle\int_{\mathbb{R}^m} \cdots \int_{\mathbb{R}^m} z_n(x_0, x_1, \cdots, x_n) \, dx_0 \, dx_1 \cdots dx_n} \ ,$$

et

$$L_n = \int_{\mathbb{R}^m} \cdots \int_{\mathbb{R}^m} z_n(x_0, x_1, \cdots, x_n)\; dx_0\, dx_1 \cdots dx_n \ ,$$

respectivement.

(iv) Montrer que la fonction $z_n(x_0, x_1, \cdots, x_n)$ peut s'exprimer comme

$$\begin{aligned} z_n(x_0, x_1, \cdots, x_n) \;=\; & \frac{1}{\sqrt{\det(2\pi\Sigma)}}\; (\frac{1}{\sqrt{\det(2\pi Q)}})^n\; (\frac{1}{\sqrt{\det(2\pi R)}})^{n+1} \\ & \exp\{-\tfrac{1}{2}\,(x_0-\mu)^*\,\Sigma^{-1}\,(x_0-\mu)\} \\ & \exp\{-\tfrac{1}{2}\,\sum_{k=1}^{n}(x_k - F\,x_{k-1})^*\,Q^{-1}\,(x_k - F\,x_{k-1})\} \\ & \exp\{-\tfrac{1}{2}\,\sum_{k=0}^{n}(Y_k - H\,x_k)^*\,R^{-1}\,(Y_k - H\,x_k)\} \ . \end{aligned}$$

En utilisant l'identité $u^*\,A\,u = \mathrm{trace}(A\,u\,u^*)$, proposer une expression équivalente pour la fonction $z_n(x_0, x_1, \cdots, x_n)$.

La deuxième étape consiste à utiliser le même *principe* (mais les calculs sont différents, au moins en partie) que celui utilisé pour établir les formules de ré–estimation de Baum–Welsh pour les paramètres d'un modèle de Markov caché.

Soit $\boldsymbol{M}' = (\mu', \Sigma', F', Q', H', R')$ le modèle courant. On souhaite proposer un nouveau modèle $\boldsymbol{M} = (\mu, \Sigma, F, Q, H, R)$ dont la vraisemblance L_n est par construction supérieure à la vraisemblance L'_n du modèle courant.

(v) Montrer que le rapport des vraisemblances peut s'exprimer comme

$$\frac{L_n}{L'_n} = \mathbb{E}'[\frac{z_n(X_0, X_1, \cdots, X_n)}{z'_n(X_0, X_1, \cdots, X_n)} \mid Y_0, \cdots, Y_n] \ ,$$

c'est–à–dire comme une espérance conditionnelle évaluée pour le modèle courant $\boldsymbol{M}'$.

Au lieu de rechercher le maximum de cette expression par rapport aux paramètres du modèle $\boldsymbol{M}$, on préfère utiliser une méthode itérative dans laquelle on exploite la minoration

$$\begin{aligned} \log \frac{L_n}{L'_n} \;&\geq\; \mathbb{E}'[\log \frac{z_n(X_0, X_1, \cdots, X_n)}{z'_n(X_0, X_1, \cdots, X_n)} \mid Y_0, \cdots, Y_n] \\ &= \mathbb{E}'[\log z_n(X_0, X_1, \cdots, X_n) \mid Y_0, \cdots, Y_n] + \mathrm{cste} \ , \end{aligned}$$

où la constante additive ne dépend pas des paramètres du modèle $\boldsymbol{M}$, mais dépend seulement des paramètres du modèle courant $\boldsymbol{M}'$. On se propose donc de rechercher le maximum de l'expression

$$\mathbb{E}'[\log z_n(X_0, X_1, \cdots, X_n) \mid Y_0, \cdots, Y_n] \ ,$$

par rapport aux paramètres du modèle $\boldsymbol{M}$, ou de manière équivalente le minimum de l'expression

$$\Xi_n = -\mathbb{E}'[\log z_n(X_0, X_1, \cdots, X_n) \mid Y_0, \cdots, Y_n] \ ,$$

par rapport aux paramètres du modèle $\boldsymbol{M}$.

(vi) Donner une expression explicite de Ξ_n en fonction des matrices Σ, Q et R, et en fonction des espérances conditionnelles (à valeurs matrices)

$$\mathbb{E}'[(X_0 - \mu)\,(X_0 - \mu)^* \mid Y_0, \cdots, Y_n] \ ,$$

$$\sum_{k=1}^{n} \mathbb{E}'[(X_k - F\,X_{k-1})\,(X_k - F\,X_{k-1})^* \mid Y_0, \cdots, Y_n] \ ,$$

$$\sum_{k=0}^{n} \mathbb{E}'[(Y_k - H\,X_k)\,(Y_k - H\,X_k)^* \mid Y_0, \cdots, Y_n] \ ,$$

évaluées pour le modèle courant $\boldsymbol{M}'$.

(vii) Exprimer ces matrices en fonction du vecteur μ, des matrices F et H, et en fonction du lisseur de Kalman

$$\widehat{X}'^{n}_{k} = \mathbb{E}'[X_k \mid Y_0, \cdots, Y_n] \ ,$$

et des matrices de covariance et de corrélation des erreurs d'estimation

$$P'^{n}_{k} = \mathbb{E}'[(X_k - \widehat{X}'^{n}_{k})\,(X_k - \widehat{X}'^{n}_{k})^*] \quad \textbf{et} \quad C'^{n}_{k} = \mathbb{E}'[(X_k - \widehat{X}'^{n}_{k})\,(X_{k-1} - \widehat{X}'^{n}_{k-1})^*] \ ,$$

évalués pour le modèle courant $\boldsymbol{M}'$.

(viii) En déduire des formules explicites pour les paramètres $(\mu, \Sigma, F, Q, H, R)$ du modèle $\boldsymbol{M}$ qui atteignent le minimum de l'expression Ξ_n.