Université de Rennes 1
Master Recherche SISEA

Examen du cours UE S3-2
“Filtre de Kalman et modèles de Markov cachés”
Jeudi 2 février 2012, 10:00 à 12:00

Problème

Le but de ce problème est d'étendre aux systèmes conditionnellement linéaires gaussiens (définis plus bas) les résultats obtenus dans le cours pour les systèmes linéaires gaussiens, en particulier les équations du filtre de Kalman.

Un système conditionnellement linéaire gaussien ressemble beaucoup à un système linéaire gaussien, à la différence que les coefficients sont autorisés à dépendre des observations passées. On considère ainsi une suite d'états cachés $\{X_k\}$ à valeurs dans $\mathbb{R}^m$, vérifiant

$$X_k = F_k(Y_{0:k-1})\, X_{k-1} + f_k(Y_{0:k-1}) + G_k(Y_{0:k-1})\, W_k \ , \tag{1}$$

où $\{X_k\}$ et $\{W_k\}$ prennent respectivement leurs valeurs dans $\mathbb{R}^m$ et $\mathbb{R}^p$, et une suite d'observations $\{Y_k\}$ à valeurs dans $\mathbb{R}^d$, vérifiant

$$Y_k = H_k(Y_{0:k-1})\, X_k + h_k(Y_{0:k-1}) + V_k \ , \tag{2}$$

où $\{V_k\}$ prend donc nécessairement ses valeurs dans $\mathbb{R}^d$ aussi, et on fait les hypothèses habituelles sur les variables aléatoires de base entrant dans la modélisation

- l'état initial X_0 est gaussien, de moyenne $\bar{X}_0$ et de matrice de covariance Q_0^X,
- la suite $\{W_k\}$ est un bruit blanc gaussien centré, de matrice de covariance Q_k^W,
- la suite $\{V_k\}$ est un bruit blanc gaussien centré, de matrice de covariance Q_k^V supposée inversible,
- les suites $\{W_k\}$ et $\{V_k\}$ et l'état initial X_0 sont mutuellement indépendants.

(i) **La suite $\{(X_k, Y_k)\}$ ainsi définie est–elle gaussienne ? À défaut, la suite $\{X_k\}$ est–elle gaussienne, la suite $\{Y_k\}$ est–elle gaussienne ?**

(ii) **En procédant comme dans le cours, c'est–à–dire en faisant un simple bilan des variables aléatoires de base mises en jeu, montrer que la variable W_k est indépendante des variables $(X_{k-1}, Y_0, \cdots, Y_{k-1})$ et que la variable V_k est indépendante des variables $(X_k, Y_0, \cdots, Y_{k-1})$.**

D'après la formule de Bayes, on rappelle que la distribution de probabilité conditionnelle jointe de l'état présent et de l'observation présente sachant les observations passées, peut se factoriser comme

$$\begin{aligned} &\mathbb{P}[X_k \in dx', Y_k \in dy_k \mid Y_{0:k-1} = y_{0:k-1}] \\ &\qquad = \mathbb{P}[Y_k \in dy_k \mid X_k = x', Y_{0:k-1} = y_{0:k-1}] \, \mathbb{P}[X_k \in dx' \mid Y_{0:k-1} = y_{0:k-1}] \ . \end{aligned} \tag{3}$$

Cette factorisation ne sera vraiment exploitée qu'à la question (viii).

(iii) **En utilisant la formule de Bayes, donner une factorisation de la distribution de probabilité conditionnelle jointe**

$$\mathbb{P}[X_k \in dx', X_{k-1} \in dx \mid Y_{0:k-1} = y_{0:k-1}] \ ,$$

de l'état présent et de l'état précédent sachant les observations passées, et en intégrant par rapport à la variable $x \in \mathbb{R}^m$, montrer que

$$\begin{aligned} &\mathbb{P}[X_k \in dx' \mid Y_{0:k-1} = y_{0:k-1}] \\ &\quad = \int_{\mathbb{R}^m} \mathbb{P}[X_k \in dx' \mid X_{k-1} = x, Y_{0:k-1} = y_{0:k-1}] \, \mathbb{P}[X_{k-1} \in dx \mid Y_{0:k-1} = y_{0:k-1}] \ . \end{aligned}$$

La suite du problème consiste à donner une expression explicite pour les distributions conditionnelles

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x, Y_{0:k-1} = y_{0:k-1}] \ ,$$

et

$$\mathbb{P}[Y_k \in dy_k \mid X_k = x', Y_{0:k-1} = y_{0:k-1}]$$

respectivement, et à effectuer ensuite un raisonnement par récurrence. On rappelle que pour étudier ou caractériser une distribution de probabilité, il est souvent utile de considérer l'intégrale d'une fonction arbitraire.

(iv) **À partir de l'équation d'état (1), montrer que la distribution de probabilité conditionnelle**

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x, Y_{0:k-1} = y_{0:k-1}] \ ,$$

de l'état présent sachant l'état précédent et les observation passées, coïncide avec la distribution de probabilité du vecteur aléatoire

$$X' = F_k(y_{0:k-1})\, x + f_k(y_{0:k-1}) + G_k(y_{0:k-1})\, W_k \ .$$

En déduire qu'il s'agit d'une distribution de probabilité gaussienne, dont on donnera l'expression de la moyenne et de la matrice de covariance.

(v) **À partir de l'équation d'observation (2), montrer que la distribution de probabilité conditionnelle**

$$\mathbb{P}[Y_k \in dy' \mid X_k = x', Y_{0:k-1} = y_{0:k-1}] \ ,$$

de l'observation présente sachant l'état présent et les observation passées, coïncide avec la distribution de probabilité du vecteur aléatoire

$$Y' = H_k(y_{0:k-1})\, x' + h_k(y_{0:k-1}) + V_k \ .$$

En déduire qu'il s'agit d'une distribution de probabilité gaussienne, dont on donnera l'expression de la moyenne et de la matrice de covariance.

Comme annoncé, on introduit l'hypothèse de récurrence suivante :

> La distribution de probabilité conditionnelle $\mathbb{P}[X_k \in dx \mid Y_{0:k} = y_{0:k}]$ de l'état caché sachant les observations, est une distribution gaussienne, dont on notera $\widehat{X}_k(y_{0:k})$ le vecteur moyenne et $P_k(y_{0:k})$ la matrice de covariance.

La fin du problème consiste donc

- à vérifier l'hypothèse de récurrence au rang $k = 0$,
- à montrer que si l'hypothèse de récurrence est vérifiée au rang $(k-1)$, alors elle est vérifiée au rang k,
- et incidemment, à obtenir une relation de récurrence permettant de calculer $\widehat{X}_k(y_{0:k})$ et $P_k(y_{0:k})$, en fonction de $\widehat{X}_{k-1}(y_{0:k-1})$, de $P_{k-1}(y_{0:k-1})$ et de y_k.

(vi) **Vérifier l'hypothèse de récurrence au rang $k = 0$.**

(vii) **À partir de l'hypothèse de récurrence au rang $(k-1)$ et de la réponse aux questions (iii) et (iv), montrer que la distribution de probabilité conditionnelle**

$$\mathbb{P}[X_k \in dx' \mid Y_{0:k-1} = y_{0:k-1}] \ ,$$

de l'état présent sachant les observations passées, coïncide avec la distribution de probabilité du vecteur aléatoire

$$X' = F_k(y_{0:k-1})\, X + f_k(y_{0:k-1}) + G_k(y_{0:k-1})\, W_k \ ,$$

où le vecteur aléatoire X est gaussien, indépendant de W_k, de moyenne $\widehat{X}_{k-1}(y_{0:k-1})$ et de matrice de covariance $P_{k-1}(y_{0:k-1})$. En déduire qu'il s'agit d'une distribution de probabilité gaussienne, dont on donnera l'expression de la moyenne $\widehat{X}_k^-(y_{0:k-1})$ et de la matrice de covariance $P_k^-(y_{0:k-1})$.

(viii) **À partir de la factorisation (3) et de la réponse aux questions (v) et (vii), montrer que la distribution de probabilité conditionnelle jointe**

$$\mathbb{P}[X_k \in dx', Y_k \in dy' \mid Y_{0:k-1} = y_{0:k-1}] \ ,$$

de l'état présent et de l'observation présente sachant les observations passées, coïncide avec la distribution de probabilité jointe du vecteur aléatoire

$$\begin{pmatrix} X' \\ \\ Y' \end{pmatrix} = \begin{pmatrix} X' \\ \\ H_k(y_{0:k-1})\, X' + h_k(y_{0:k-1}) + V_k \end{pmatrix} \ ,$$

où le vecteur aléatoire X' est gaussien, indépendant de V_k, de moyenne $\widehat{X}_k^-(y_{0:k-1})$ et de matrice de covariance $P_k^-(y_{0:k-1})$. En déduire que la distribution de probabilité conditionnelle

$$\mathbb{P}[X_k \in dx' \mid Y_{0:k} = y_{0:k}]$$

de l'état présent sachant les observations, est une distribution de probabilité gaussienne, dont on donnera l'expression de la moyenne $\widehat{X}_k(y_{0:k})$ et de la matrice de covariance $P_k(y_{0:k})$.

(ix) **En quoi le résultat obtenu était–il prévisible dès le début, au moins d'un point de vue intuitif ?**

Exercice

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

- *loi initiale* $\nu = (\nu_i)$
$$\nu_i = \mathbb{P}[X_0 = i] \ , \qquad \text{pour tout } i \in E,$$
- *matrice de transition* $\pi = (\pi_{i,j})$
$$\pi_{i,j} = \mathbb{P}[X_{k+1} = j \mid X_k = i] \ , \qquad \text{pour tout } i, j \in E.$$

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

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

où la suite $\{V_k\}$ est un bruit blanc gaussien centré de dimension d, de matrice de covariance identité, indépendant de la chaîne de Markov $\{X_k\}$.

La fonction h définie sur E à valeurs dans $\mathbb{R}^d$ est caractérisée par la donnée d'une famille $h = (h_i)$ de N vecteurs de $\mathbb{R}^d$, la fonction $\Sigma = s\, s^*$ définie sur E à valeurs dans les matrices $d \times d$ symétriques définies positives (inversibles) est caractérisée par la donnée d'une famille $\Sigma = (\Sigma_i)$ de N matrices $d \times d$ symétriques définies positives, et on a

$$\mathbb{P}[Y_k \in dy \mid X_k = i] = \frac{1}{\sqrt{\det(2\pi\, \Sigma_i)}} \ \exp\{-\tfrac{1}{2}\, (y - h_i)^*\, \Sigma_i^{-1}\, (y - h_i)\}\, dy \ ,$$

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

Le but de cet exercice est de proposer un algorithme d'estimation pour la famille $h = (h_i)$ des vecteurs moyennes et pour la famille $\Sigma = (\Sigma_i)$ des matrices de covariance, au vu des observations $Y_{0:n} = (Y_0, \cdots, Y_n)$.

On considère d'abord le cas où $X_{0:n} = (X_0, \cdots, X_n)$ et $Y_{0:n} = (Y_0, \cdots, Y_n)$ sont observés. La fonction de log–vraisemblance pour l'estimation de $h = (h_i)$ et de $\Sigma = (\Sigma_i)$, à partir de l'observation conjointe de $X_{0:n} = (X_0, \cdots, X_n)$ et de $Y_{0:n} = (Y_0, \cdots, Y_n)$, s'écrit

$$\ell(h, \Sigma) = -\tfrac{1}{2} \sum_{k=0}^{n} \log(2\pi\, \Sigma(X_k)) - \tfrac{1}{2} \sum_{k=0}^{n} (Y_k - h(X_k))^*\, \Sigma^{-1}(X_k)\, (Y_k - h(X_k)) \ .$$

ou de manière équivalente

$$\ell(h, M) = \tfrac{1}{2} \sum_{k=0}^{n} \log\det(2\pi\, M(X_k)) - \tfrac{1}{2} \sum_{k=0}^{n} (Y_k - h(X_k))^*\, M(X_k)\, (Y_k - h(X_k)) \ ,$$

en introduisant la famille $M = (M_i)$ où $M_i = \Sigma_i^{-1}$ pour tout $i \in E$.

(i) **En utilisant les décompositions**

$$h(X_k) = \sum_{i \in E} 1_{(X_k = i)} \, h_i \qquad \textbf{et} \qquad M(X_k) = \sum_{i \in E} 1_{(X_k = i)} \, M_i \ ,$$

montrer que

$$\ell(h, M) \ = \ \tfrac{1}{2} \sum_{i \in E} \sum_{k=0}^{n} 1_{(X_k = i)} \, [\, \log \det(2\pi \, M_i) - (Y_k - h_i)^* \, M_i \, (Y_k - h_i) \,] \ .$$

Rappel Si $A = (a^{lm})$ est une matrice $d \times d$ inversible, alors on rappelle

- le développement du déterminant $\det A$ le long de la ligne l (ou bien de la colonne m)

$$\det A = \sum_{m=1}^{d} a^{lm} \, (-1)^{l+m} \ \det A^{lm} = \sum_{l=1}^{d} a^{lm} \, (-1)^{l+m} \ \det A^{lm} \ ,$$

  où la matrice A^{lm} désigne la matrice construite à partir de la matrice A en supprimant la ligne l et la colonne m,

- l'expression du coefficient b^{lm} à l'intersection de la ligne l et de la colonne m de la matrice inverse $A^{-1} = B = (b^{lm})$

$$b^{lm} = (-1)^{l+m} \, \frac{\det A^{lm}}{\det A} \ ,$$

et on en déduit que

$$\frac{\partial}{\partial a^{lm}} \det A = (-1)^{l+m} \ \det A^{lm} = (-1)^{l+m} \, \frac{\det A^{lm}}{\det A} \ \det A = b^{lm} \ \det A \ ,$$

soit

$$\frac{\partial}{\partial a^{lm}} \log \det A = b^{lm} \ .$$

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

$$\widehat{h}_i = \frac{\displaystyle\sum_{k=0}^{n} 1_{(X_k = i)} \, Y_k}{\displaystyle\sum_{k=0}^{n} 1_{(X_k = i)}} \qquad \textbf{et} \qquad \widehat{\Sigma}_i = \frac{\displaystyle\sum_{k=0}^{n} 1_{(X_k = i)} \, (Y_k - \widehat{h}_i) \, (Y_k - \widehat{h}_i)^*}{\displaystyle\sum_{k=0}^{n} 1_{(X_k = i)}} \ ,$$

respectivement.

On considère ensuite le cas où seulement $Y_{0:n} = (Y_0, \cdots, Y_n)$ est observé. On remarque que

$$\widehat{\Sigma}_i = \frac{\displaystyle\sum_{k=0}^{n} 1(X_k = i)\, (Y_k - \widehat{h}_i)\, (Y_k - \widehat{h}_i)^*}{\displaystyle\sum_{k=0}^{n} 1(X_k = i)} = \frac{\displaystyle\sum_{k=0}^{n} 1(X_k = i)\, Y_k\, Y_k^*}{\displaystyle\sum_{k=0}^{n} 1(X_k = i)} - \widehat{h}_i\, \widehat{h}_i^* \ ,$$

et on pose

$$N_i = \sum_{k=0}^{n} 1(X_k = i) \ , \qquad O_i = \sum_{k=0}^{n} 1(X_k = i)\, Y_k \qquad \text{et} \qquad Q_i = \sum_{k=0}^{n} 1(X_k = i)\, Y_k\, Y_k^* \ .$$

En supposant un modèle $\mathbf{M}' = (\nu', \pi', h', \Sigma')$ donné, on se propose de redéfinir les estimateurs $\widehat{h}_i$ et $\widehat{\Sigma}_i$ en remplaçant N_i, O_i et Q_i par leurs espérances conditionnelles respectives

$$\widehat{N}_i = \mathbb{E}'[N_i \mid Y_{0:n}] \ , \qquad \widehat{O}_i = \mathbb{E}'[O_i \mid Y_{0:n}] \qquad \text{et} \qquad \widehat{Q}_i = \mathbb{E}'[Q_i \mid Y_{0:n} \ ,$$

par rapport à $Y_{0:n}$ et calculées sous le modèle $\mathbf{M}'$.

(iii) **Donner l'expression des estimateurs $\widehat{N}_i = \mathbb{E}'[N_i \mid Y_{0:n}]$, $\widehat{O}_i = \mathbb{E}'[O_i \mid Y_{0:n}]$ et $\widehat{Q}_i = \mathbb{E}'[Q_i \mid Y_{0:n}]$ à l'aide des variables forward et backward correspondant au modèle $\mathbf{M}'$.**

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