# Université de Rennes 1
# Master Recherche SISEA

# Examen du cours UE S3-2
# “Filtre de Kalman et modèles de Markov cachés”
# Jeudi 28 janvier 2010, 10:00 à 12:00
# — Corrigé partiel —

## Problème :

L’objectif de ce problème est d’établir les équations du filtre de Kalman, dans le cas plus général d’un système linéaire observé dans un bruit additif défini par un modèle ARMA (auto–regressive moving average) gaussien. Plus spécifiquement, on suppose que

- les états cachés sont à valeurs dans $\mathbb{R}^m$ et vérifient la relation récurrente

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

  où la suite $\{W_k\}$ est un bruit blanc gaussien, de moyenne nulle et de matrice de covariance $Q^W_k$, et où la condition initiale $X_0$ est un vecteur aléatoire gaussien, de moyenne $\bar{X}_0$ et de matrice de covariance $Q^X_0$,

- les observations sont à valeurs dans $\mathbb{R}^d$ et sont reliées aux états cachés par

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

  où $\{V_k\}$ est une suite de vecteurs aléatoires *dépendants* à valeurs dans $\mathbb{R}^d$, définie par le modèle ARMA$(p,q)$

  $$V_k + A_1\, V_{k-1} + \cdots + A_p\, V_{k-p} = B_0\, e_k + B_1\, e_{k-1} + \cdots + B_q\, e_{k-q} \ ,$$

  paramétré par les matrices $A_1, \cdots, A_p$ et $B_0, B_1, \cdots, B_q$, où la suite $\{e_k\}$ est un bruit blanc gaussien à valeurs dans $\mathbb{R}^d$, de moyenne nulle et de matrice de covariance $Q^E_k$ inversible, et où le vecteur aléatoire $(V_{-p}, \cdots, V_{-1})$ qui permet de démarrer la récurrence est gaussien.

On suppose que les conditions initiales $X_0$ et $(V_{-1}, \cdots, V_{-p})$ et les suites $\{W_k\}$ et $\{e_k\}$ sont mutuellement indépendantes. On suppose aussi que la matrice $B_0$ est inversible (de sorte que finalement la matrice de covariance du vecteur aléatoire gaussien $B_0\, e_k$ est inversible).

(i) **Montrer que la suite $\{(X_k, Y_k)\}$ est gaussienne. Il en résulte que la distribution conditionnelle de l'état caché $X_k$ sachant les observations passées $(Y_0, \cdots, Y_k)$ est une distribution gaussienne, mais quelle hypothèse du cours n'est pas vérifiée ici, qui empêche d'appliquer directement les équations du filtre de Kalman ?**

SOLUTION

Pour tout instant $n$, le vecteur aléatoire $(X_0, Y_0, \cdots, X_n, Y_n)$ s'exprime comme transformation affine du vecteur aléatoire $(X_0, W_1, \cdots, W_n, V_0, V_1, \cdots, V_n)$, et pour tout instant $k$, le vecteur aléatoire $V_k$ s'exprime comme transformation affine du vecteur aléatoire $(e_{-q}, \cdots, e_{-1}, e_0, e_1, \cdots, e_k)$ et de la condition initiale $(V_{-p}, \cdots, V_{-1})$, c'est–à–dire que pour tout instant $n$, le vecteur aléatoire $(X_0, Y_0, \cdots, X_n, Y_n)$ s'exprime comme transformation affine du vecteur aléatoire $(V_{-p}, \cdots, V_{-1}, X_0, W_1, \cdots, W_n, e_{-q}, \cdots, e_{-1}, e_0, e_1, \cdots, e_n)$, qui par hypothèse est un vecteur aléatoire gaussien. Il en résulte que le vecteur aléatoire $(X_0, Y_0, \cdots, X_n, Y_n)$ est gaussien, comme transformation affine d'un vecteur aléatoire gaussien.

En revanche, la suite $\{V_k\}$ n'est pas un bruit blanc, ce qui empêche d'appliquer directement les équations du filtre de Kalman.

$\square$

On admet qu'il existe une représentation d'état pour le bruit d'observation, de la forme

$$U_k = A\, U_{k-1} + B\, e_{k-1} \ ,$$

$$V_k = C\, U_k + B_0\, e_k \ ,$$

où la suite $\{U_k\}$ est à valeurs dans $\mathbb{R}^{rd}$, où la condition initiale $U_0$ est un vecteur aléatoire gaussien indépendant de la suite $\{e_k\}$, et où $A$, $B$ et $C$ sont des matrices de dimension $rd \times rd$, $rd \times d$ et $d \times rd$, respectivement, avec $r = \max(p, q)$. Dans cette représentation d'état, la matrice $A$ dépend des matrices $A_1, \cdots, A_p$ du modèle ARMA et la matrice $B$ dépend des matrices $A_1, \cdots, A_p$ et $B_0, B_1, \cdots, B_q$ du modèle ARMA, mais on se contente ici de supposer l'existence de la représentation d'état proposée.

(ii) **On définit la nouvelle variable d'état $Z_k = (X_k, U_k)$. Montrer que les nouveaux états cachés ainsi définis vérifient la relation récurrente**

$$Z_k = F(A)\, Z_{k-1} + M(B)\, e_{k-1} + L\, W_k \ , \qquad (\star)$$

**où on donnera l'expression des matrices $F(A)$, $M(B)$ et $L$ ainsi définies.**

SOLUTION

Clairement

$$Z_k = \begin{pmatrix} X_k \\ \\ U_k \end{pmatrix} = \begin{pmatrix} F\,X_{k-1} + G\,W_k \\ \\ A\,U_{k-1} + B\,e_{k-1} \end{pmatrix}$$

$$= \begin{pmatrix} F & 0 \\ \\ 0 & A \end{pmatrix} \begin{pmatrix} X_{k-1} \\ \\ U_{k-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \\ B \end{pmatrix} e_{k-1} + \begin{pmatrix} G \\ \\ 0 \end{pmatrix} W_k \ ,$$

et les matrices $F(A)$, $M(B)$ et $L$ sont définies par

$$F(A) = \begin{pmatrix} F & 0 \\ \\ 0 & A \end{pmatrix} \ , \qquad M(B) = \begin{pmatrix} 0 \\ \\ B \end{pmatrix} \qquad \text{et} \qquad L = \begin{pmatrix} G \\ \\ 0 \end{pmatrix} \ ,$$

respectivement.

□

(iii) **Montrer que les observations sont reliées aux nouveaux états cachés par**

$$Y_k = H(C)\,Z_k + B_0\,e_k \ , \tag{$\star\star$}$$

**où on donnera l'expression de la matrice $H(C)$ ainsi définie.**

Solution

Clairement

$$Y_k = H\,X_k + V_k = H\,X_k + C\,U_k + B_0\,e_k = \begin{pmatrix} H & C \end{pmatrix} \begin{pmatrix} X_k \\ \\ U_k \end{pmatrix} + B_0\,e_k \ ,$$

et la matrice $H(C)$ est définie par

$$H(C) = \begin{pmatrix} H & C \end{pmatrix} \ .$$

□

(iv) **Montrer que la suite $\{(Z_k, Y_k)\}$ est gaussienne. Il en résulte que la distribution conditionnelle du nouvel état caché $Z_k$ sachant les observations passées $(Y_0, \cdots, Y_k)$ est une distribution gaussienne, mais quelle hypothèse du cours n'est pas vérifiée ici, qui empêche d'appliquer directement les équations du filtre de Kalman ?**

Solution

Pour tout instant $n$, le vecteur aléatoire $(Z_0, Y_0, \cdots, Z_n, Y_n)$ s'exprime comme transformation affine du vecteur aléatoire $(Z_0, W_1, \cdots, W_n, e_0, e_1, \cdots, e_n)$, qui par hypothèse est un vecteur aléatoire gaussien, donc le vecteur aléatoire $(Z_0, Y_0, \cdots, Z_n, Y_n)$ est gaussien, comme transformation affine d'un vecteur aléatoire gaussien.

En revanche, le bruit d'état $(e_{k-1}, W_k)$ et le bruit d'observation $e_k$ ne forment pas deux suites indépendantes, ce qui empêche d'appliquer directement les équations du filtre de Kalman.

□

(v) **Montrer qu'on peut écrire le modèle $(\star)$, $(\star\star)$ sous la forme**

$$Z_k = F(A, B, C)\, Z_{k-1} + M(B_0, B)\, Y_{k-1} + L\, W_k \ ,$$

$$Y_k = H(C)\, Z_k + B_0\, e_k \ ,$$

**où on donnera l'expression des matrices $F(A, B, C)$ et $M(B_0, B)$ ainsi définies. On pourra par exemple utiliser l'équation $(\star\star)$ exprimée à l'instant $(k-1)$ pour obtenir $e_{k-1}$ en fonction de $Y_{k-1}$ et $Z_{k-1}$, et reporter dans $(\star)$ l'expression obtenue.**

Solution

Compte tenu que la matrice $B_0$ est inversible, par hypothèse, l'équation $(\star\star)$ exprimée à l'instant $(k-1)$ donne

$$e_{k-1} = B_0^{-1}\, (Y_{k-1} - H\, X_{k-1} - C\, U_{k-1}) \ ,$$

et en reportant dans $(\star)$, on obtient

$$U_k = A\, U_{k-1} + B\, e_{k-1} = A\, U_{k-1} + B\, B_0^{-1}\, (Y_{k-1} - H\, X_{k-1} - C\, U_{k-1}) \ ,$$

de sorte que

$$Z_k = \begin{pmatrix} X_k \\ U_k \end{pmatrix} = \begin{pmatrix} F\, X_{k-1} + G\, W_k \\ A\, U_{k-1} + B\, B_0^{-1}\, (Y_{k-1} - H\, X_{k-1} - C\, U_{k-1}) \end{pmatrix}$$

$$= \begin{pmatrix} F & 0 \\ -B\, B_0^{-1}\, H & A - B\, B_0^{-1}\, C \end{pmatrix} \begin{pmatrix} X_{k-1} \\ U_{k-1} \end{pmatrix}$$

$$+ \begin{pmatrix} 0 \\ B\, B_0^{-1} \end{pmatrix} Y_{k-1} + \begin{pmatrix} G \\ 0 \end{pmatrix} W_k \ ,$$

et les matrices $F(A,B,C)$ et $M(B,B_0)$ sont définies par

$$F(A,B,C) = \begin{pmatrix} F & 0 \\ & \\ -B\,B_0^{-1}\,H & A - B\,B_0^{-1}\,C \end{pmatrix} \qquad \text{et} \qquad M(B,B_0) = \begin{pmatrix} 0 \\ \\ B\,B_0^{-1} \end{pmatrix} .$$

□

Bien que l'observation apparaisse dans l'équation d'état, on est maintenant capable de calculer récursivement le vecteur moyenne $\widehat{Z}_k$ et la matrice de covariance $\Sigma_k$ de la distribution conditionnelle du nouvel état caché $Z_k$ sachant les observations passées $Y_{0:k} = (Y_0, \cdots, Y_k)$. On introduira aussi le vecteur moyenne $\widehat{Z}_k^-$ et la matrice de covariance $\Sigma_k^-$ de la distribution conditionnelle du nouvel état caché $Z_k$ sachant les observations passées $Y_{0:k-1} = (Y_0, \cdots, Y_{k-1})$.

Dans les questions suivantes, il s'agira juste d'adapter les démonstrations vues en cours, en prenant soin d'indiquer en quoi la présence de l'observation dans l'équation d'état modifie les équations.

(vi) **Exprimer le vecteur moyenne $\widehat{Z}_k^-$ et la matrice de covariance $\Sigma_k^-$ en fonction de $\widehat{Z}_{k-1}$ et de $\Sigma_{k-1}$.**

Solution

En utilisant l'équation d'état obtenue à la question (v), on obtient

$$\begin{aligned} \widehat{Z}_k^- &= \mathbb{E}[Z_k \mid Y_{0:k-1}] \\ &= F(A,B,C)\,\mathbb{E}[Z_{k-1} \mid Y_{0:k-1}] + M(B,B_0)\,\mathbb{E}[Y_{k-1} \mid Y_{0:k-1}] + L\,\mathbb{E}[W_k \mid Y_{0:k-1}] \\ &= F(A,B,C)\,\widehat{Z}_{k-1} + M(B,B_0)\,Y_{k-1} , \end{aligned}$$

où on a utilisé dans la dernière égalité l'indépendance des vecteurs aléatoires $W_k$ et $Y_{0:k-1}$. Par différence

$$Z_k - \widehat{Z}_k^- = F(A,B,C)\,(Z_{k-1} - \widehat{Z}_{k-1}) + L\,W_k ,$$

de sorte que

$$\begin{aligned}
\Sigma_k^- &= \mathbb{E}[\,(Z_k - \widehat{Z}_k^-)\,(Z_k - \widehat{Z}_k^-)^*] \\
&= \mathbb{E}[\,(F(A,B,C)\,(Z_{k-1} - \widehat{Z}_{k-1}) + L\,W_k)\,(F(A,B,C)\,(Z_{k-1} - \widehat{Z}_{k-1}) + L\,W_k)^*] \\
&= F(A,B,C)\,\mathbb{E}[\,(Z_{k-1} - \widehat{Z}_{k-1})\,(Z_{k-1} - \widehat{Z}_{k-1})^*]\,F^*(A,B,C) + L\,\mathbb{E}[W_k\,W_k^*]\,L^* \\
&\qquad + F(A,B,C)\,\mathbb{E}[\,(Z_{k-1} - \widehat{Z}_{k-1})\,W_k^*]\,L^* \\
&\qquad + L\,\mathbb{E}[W_k\,(Z_{k-1} - \widehat{Z}_{k-1})^*]\,F^*(A,B,C) \\
&= F(A,B,C)\,\Sigma_{k-1}\,F^*(A,B,C) + L\,Q_k^W\,L^*\ ,
\end{aligned}$$

où on a utilisé dans la dernière égalité l'indépendance de $W_k$ et $(Z_{k-1}, Y_{0:k-1})$.

---
$\square$

(vii) **En introduisant la suite des innovations, exprimer le vecteur moyenne $\widehat{Z}_k$ et la matrice de covariance $\Sigma_k$ en fonction de $\widehat{Z}_k^-$ et de $\Sigma_k^-$.**

(viii) **Comment exprimer très simplement le vecteur moyenne $\widehat{X}_k$ et la matrice de covariance $P_k$ de la distribution conditionnelle de l'état caché $X_k$ sachant les observations passées $Y_{0:k} = (Y_0, \cdots, Y_k)$, en fonction de $\widehat{Z}_k$ et de $\Sigma_k$ ?**

## Exercice :

L'objectif de cet exercice est de montrer que les équations de Baum obtenues pour les modèles de Markov cachés avec des observations indépendantes conditionnellement aux états cachés, peuvent se généraliser facilement au cas où les observations forment une chaîne de Markov conditionnellement aux états cachés.

On considère une chaîne de Markov $\{X_k\}$ à espace d'état fini $E = \{1, \cdots, N\}$, de loi initiale

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

et de matrice de transition

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

La suite $\{X_k\}$ n'est pas observée directement, mais on observe la suite $\{Y_k\}$ à valeurs dans $\mathbb{R}^d$, définie par

$$Y_k = h(X_k, Y_{k-1}) + V_k \ ,$$

où la suite $\{V_k\}$ est un bruit blanc, avec une densité

$$\mathbb{P}[V_k \in dv] = q_k^V(v)\, dv \ ,$$

et où la condition initiale $Y_0$ possède une densité

$$\mathbb{P}[Y_0 \in dy] = g_0(y)\, dy \ .$$

On suppose que la condition initiale $Y_0$ et les suites $\{X_k\}$ et $\{V_k\}$ sont mutuellement indépendantes.

(i) **Donner l'expression des densités d'émission**

$$\mathbb{P}[Y_k \in dy' \mid X_k = j, Y_{k-1} = y] = g^j(y, y')\, dy' \ ,$$

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

Solution

Compte tenu que le vecteur aléatoire $V_k$ et le vecteur aléatoire $(X_k, Y_{k-1})$ sont indépendants, on a immédiatement

$$
\begin{aligned}
\mathbb{E}[\phi(Y_k) \mid X_k = j, Y_{k-1} = y] &= \mathbb{E}[\phi(h(X_k, Y_{k-1}) + V_k) \mid X_k = j, Y_{k-1} = y] \\
&= \mathbb{E}[\phi(h_j(y) + V_k) \mid X_k = j, Y_{k-1} = y] \\
&= \int_{\mathbb{R}^d} \phi(h_j(y) + v) \, \mathbb{P}[V_k \in dv \mid X_k = j, Y_{k-1} = y] \\
&= \int_{\mathbb{R}^d} \phi(h_j(y) + v) \, \mathbb{P}[V_k \in dv] \\
&= \int_{\mathbb{R}^d} \phi(h_j(y) + v) \, q_k^V(v) \, dv \\
&= \int_{\mathbb{R}^d} \phi(y') \, q_k^V(y' - h_j(y)) \, dy' \ ,
\end{aligned}
$$

pour toute fonction $\phi$ définie sur $\mathbb{R}^d$, d'où le résultat avec

$$
g^j(y, y') = q_k^V(y' - h_j(y)) \ ,
$$

où par définition $h(x', y) = h_j(y)$ si $x' = j$, pour tout $y, y' \in \mathbb{R}^d$ et pour tout $j \in E$.

$\square$

(ii) **Montrer que conditionnellement aux états cachés $(X_0, \cdots, X_n)$ les observations $(Y_0, \cdots, Y_n)$ forment une chaîne de Markov, dont la distribution de probabilité de transition ne dépend que de l'état caché à la fin de la transition, et peut donc s'exprimer à l'aide de la densité d'émission introduite à la question (i), c'est–à–dire que la loi conditionnelle jointe peut se factoriser de la façon suivante**

$$
\begin{aligned}
&\mathbb{P}[Y_0 \in dy_0, \cdots, Y_n \in dy_n \mid X_0 = i_0, \cdots, X_n = i_n] \\
&\qquad = \mathbb{P}[Y_0 \in dy_0] \prod_{k=1}^{n} \mathbb{P}[Y_k \in dy_k \mid X_k = i_k, Y_{k-1} = y_{k-1}] \\
&\qquad = g(y_0) \, dy_0 \prod_{k=1}^{n} g^{i_k}(y_{k-1}, y_k) \, dy_k \ ,
\end{aligned}
$$

**pour tout $y_0, \cdots, y_n \in \mathbb{R}^d$ et pour tout $i_0, \cdots, i_n \in E$.**

Solution

En toute généralité (par exemple par récurrence)

$$\mathbb{P}[Y_0 \in dy_0, \cdots, Y_n \in dy_n \mid X_0 = i_0, \cdots, X_n = i_n]$$

$$= \mathbb{P}[Y_0 \in dy_0 \mid X_0 = i_0, \cdots, X_n = i_n]$$

$$\prod_{k=1}^{n} \mathbb{P}[Y_k \in dy_k \mid X_0 = i_0, \cdots, X_n = i_n, Y_0 = y_0, \cdots, Y_{k-1} = y_{k-1}] \ ,$$

et compte tenu que le vecteur aléatoire $V_k$ et le vecteur aléatoire $(X_k, Y_0, \cdots, Y_{k-1})$ sont indépendants, on a immédiatement

$$\mathbb{E}[\phi(Y_k) \mid X_0 = i_0, \cdots, X_n = i_n, Y_0 = y_0, \cdots, Y_{k-1} = y_{k-1}]$$

$$= \mathbb{E}[\phi(h(X_k, Y_{k-1}) + V_k) \mid X_0 = i_0, \cdots, X_n = i_n, Y_0 = y_0, \cdots, Y_{k-1} = y_{k-1}]$$

$$= \mathbb{E}[\phi(h_{i_k}(y_{k-1}) + V_k) \mid X_0 = i_0, \cdots, X_n = i_n, Y_0 = y_0, \cdots, Y_{k-1} = y_{k-1}]$$

$$= \int_{\mathbb{R}^d} \phi(h_{i_k}(y_{k-1}) + v) \, \mathbb{P}[V_k \in dv \mid X_0 = i_0, \cdots, X_n = i_n, Y_0 = y_0, \cdots, Y_{k-1} = y_{k-1}]$$

$$= \int_{\mathbb{R}^d} \phi(h_{i_k}(y_{k-1}) + v) \, \mathbb{P}[V_k \in dv]$$

$$= \int_{\mathbb{R}^d} \phi(h_{i_k}(y_{k-1}) + v) \, q_k^V(v) \, dv$$

$$= \int_{\mathbb{R}^d} \phi(y_k) \, q_k^V(y_k - h_{i_k}(y_{k-1})) \, dy_k$$

$$= \int_{\mathbb{R}^d} \phi(y_k) \, g^{i_k}(y_{k-1}, y_k) \, dy_k \ ,$$

pour toute fonction $\phi$ définie sur $\mathbb{R}^d$, de sorte que

$$\mathbb{P}[Y_k \in dy_k \mid X_0 = i_0, \cdots, X_n = i_n, Y_0 = y_0, \cdots, Y_{k-1} = y_{k-1}] = g^{i_k}(y_{k-1}, y_k) \, dy_k \ ,$$

ne dépend que de $i_k$ et $y_{k-1}$. Par ailleurs

$$\mathbb{P}[Y_0 \in dy_0 \mid X_0 = i_0, \cdots, X_n = i_n] = \mathbb{P}[Y_0 \in dy_0] = g(y_0) \, dy_0 \ .$$

$\square$

On pose $p_k = (p_k^i)$ avec

$$p_k^i = \mathbb{P}[X_k = i \mid Y_{0:k}] \qquad \text{pour tout } i \in E.$$

Dans la question suivante, il s'agira juste d'adapter la démonstration vue en cours, en prenant soin d'indiquer en quoi la présence de l'observation précédente dans la densité d'émission modifie les équations.

(iii) **Donner l'équation de Baum vérifiée par la suite $\{p_k\}$.**