

École Nationale Supérieure de Techniques Avancées
module : Commande des Systèmes

examen du cours B7–1

“Filtrage bayésien et approximation particulière”

lundi 18 octobre 2010, 8:30 à 10:30

EXERCICE 1 :

L’objectif de cet exercice est d’établir les équations du filtre bayésien, dans le cas plus général d’un système non-linéaire à bruits non-nécessairement gaussiens et non-nécessairement indépendants. En pratique, il s’agira de montrer que chacun des trois modèles présentés ci-dessous rentre dans le cadre des chaînes de Markov partiellement observées, ce qui rend possible en principe l’approximation particulière du filtre bayésien.

Modèle A Le modèle le plus simple considéré ici est

$$X_k = f_k(X_{k-1}, V_{k-1}, W_k) ,$$

$$Y_k = h_k(X_k) + V_k ,$$

où le bruit d’observation apparaît explicitement dans l’équation d’état. Les suites $\{W_k\}$ et $\{V_k\}$ sont deux bruits blancs indépendants, de distribution de probabilité $p_k^W(dw)$ et $q_k^V(v) dv$ respectivement, et on fait l’hypothèse habituelle qu’il est facile de simuler une variable aléatoire selon la distribution de probabilité $p_k^W(dw)$ et qu’il est facile d’évaluer la fonction $q_k^V(v)$ pour tout $v \in \mathbb{R}^d$. Clairement, le bruit $U_{k-1} = (V_{k-1}, W_k)$ dans l’équation d’état et le bruit d’observation V_{k-1} ne sont pas indépendants.

- (i) **En reportant dans l’équation d’état l’expression de V_{k-1} tirée de l’équation d’observation, montrer que l’état joint présent (X_k, Y_k) dépend seulement de l’état joint précédent (X_{k-1}, Y_{k-1}) et du vecteur aléatoire (W_k, V_k) . En déduire que les états cachés et les observations forment conjointement une chaîne de Markov.**

Modèle B À l'autre extrémité, le modèle le plus général considéré ici est

$$\begin{aligned} X_k &= f_k(X_{k-1}, U_{k-1}) , \\ Y_k &= h_k(X_k) + V_k . \end{aligned}$$

La suite $\{(U_k, V_k)\}$ est un bruit blanc, mais le bruit U_k dans l'équation d'état et le bruit d'observation V_k sont supposés dépendants, avec la factorisation

$$\mathbb{P}[U_k \in du, V_k \in dv] = \mathbb{P}[U_k \in du \mid V_k = v] \mathbb{P}[V_k \in dv] = p_k^{U|V}(v, du) q_k^V(v) dv ,$$

de la distribution de probabilité jointe, et on fait l'hypothèse qu'il est facile de simuler un vecteur aléatoire selon la distribution de probabilité conditionnelle $p_k^{U|V}(v, du)$, pour tout $v \in \mathbb{R}^d$, et qu'il est facile d'évaluer la fonction $q_k^V(v)$ pour tout $v \in \mathbb{R}^d$.

- (ii) **Montrer que l'état joint présent (X_k, Y_k) dépend de l'état joint précédent (X_{k-1}, Y_{k-1}) et du vecteur aléatoire (U_{k-1}, V_k) , mais que le vecteur aléatoire U_{k-1} et l'état joint précédent (X_{k-1}, Y_{k-1}) ne sont pas indépendants.**

Modèle C Intermédiaire entre les modèles A et B, le modèle

$$\begin{aligned} X_k &= f_k(X_{k-1}, U_{k-1}) , \\ Y_k &= h_k(X_k) + V_k , \end{aligned} \tag{1}$$

est un cas particulier du modèle B, avec l'hypothèse supplémentaire que la suite $\{(U_k, V_k)\}$ est un bruit blanc *gaussien*. En particulier, le vecteur aléatoire (U_k, V_k) est gaussien et centré, et on note

$$\begin{pmatrix} Q_k & S_k \\ S_k^* & R_k \end{pmatrix} ,$$

sa matrice de covariance, où on suppose que le bloc R_k est inversible.

- (iii) **Montrer que le vecteur aléatoire U_{k-1} peut se décomposer comme**

$$U_{k-1} = A_k V_{k-1} + W_k ,$$

où les vecteurs aléatoires V_{k-1} et W_k sont indépendants et gaussiens, et donner l'expression de la matrice A_k et de la matrice de covariance Σ_k du vecteur aléatoire gaussien W_k .

- (iv) **En déduire que le modèle C peut également être interprété comme un cas particulier du modèle A.**

Pour les trois modèles considérés, la décomposition

$$\begin{aligned} & \mathbb{P}[X_k \in dx_k, Y_k \in dy_k \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ &= \mathbb{P}[X_k \in dx_k \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ & \quad \mathbb{P}[Y_k \in dy_k \mid X_{0:k} = x_{0:k}, Y_{0:k-1} = y_{0:k-1}] , \end{aligned}$$

de la distribution de probabilité conditionnelle de l'état joint présent (X_k, Y_k) sachant les états joints passés $(X_{0:k-1}, Y_{0:k-1})$ est valide en toute généralité, et on fait l'hypothèse habituelle que la distribution de probabilité du bruit d'observation V_k admet une densité $q_k^V(v)$ qu'il est facile d'évaluer en tout point $v \in \mathbb{R}^d$.

(v) **En procédant comme dans le cours, montrer que pour les trois modèles considérés**

$$\begin{aligned} & \mathbb{P}[Y_k \in dy_k \mid X_{0:k} = x_{0:k}, Y_{0:k-1} = y_{0:k-1}] \\ &= \mathbb{P}[Y_k \in dy_k \mid X_k = x_k] = q_k^V(y_k - h_k(x_k)) dy_k . \end{aligned}$$

Il suffit donc de décrire la distribution de probabilité conditionnelle de l'état caché présent X_k sachant les états joints passés $(X_{0:k-1}, Y_{0:k-1})$, c'est-à-dire

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

dans chacun des trois modèles considérés, et on s'intéresse d'abord au modèle A.

(vi) **Pour le modèle A, montrer que**

$$\begin{aligned} & \mathbb{P}[X_k \in dx_k \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ &= \mathbb{P}[X_k \in dx_k \mid X_{k-1} = x_{k-1}, Y_{k-1} = y_{k-1}] = P_k(y_{k-1}, x_{k-1}, dx_k) , \end{aligned}$$

avec un noyau de transition paramétré par y_{k-1} , et défini de manière implicite par

$$\begin{aligned} \mathbb{E}[\phi(X_k) \mid X_{k-1} = x_{k-1}, Y_{k-1} = y_{k-1}] &= \int P_k(y_{k-1}, x_{k-1}, dx_k) \phi(x_k) \\ &= \int \phi(f_k(x_{k-1}, y_{k-1} - h_{k-1}(x_{k-1}), w)) p_k^W(dw) , \end{aligned}$$

pour toute fonction test ϕ . Montrer qu'il est facile de simuler un vecteur aléatoire selon la distribution de probabilité $P_k(y, x, dx')$ ainsi définie, pour tout $x \in \mathbb{R}^m$ et tout $y \in \mathbb{R}^d$.

On s'intéresse ensuite au modèle B.

(vii) **Pour le modèle B, montrer que**

$$\begin{aligned} & \mathbb{E}[\phi(X_k) \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ &= \int \phi(f_k(x_{k-1}, u)) \mathbb{P}[U_{k-1} \in du \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] , \end{aligned}$$

où le vecteur aléatoire U_{k-1} et (la dernière composante observée Y_{k-1} dans) les états joints passés $(X_{0:k-1}, Y_{0:k-1})$ ne sont pas indépendants.

On se propose donc d'étudier la distribution de probabilité conditionnelle jointe du vecteur aléatoire (U_k, Y_k) sachant les états joints passés $(X_{0:k}, Y_{0:k-1})$

(viii) **Montrer que**

$$\begin{aligned} & \mathbb{P}[U_k \in du, Y_k \in dy_k \mid X_{0:k} = x_{0:k}, Y_{0:k-1} = y_{0:k-1}] \\ &= p_k^{U|V}(y_k - h_k(x_k), du) q_k^V(y_k - h_k(x_k)) dy_k , \end{aligned}$$

et en déduire que

$$\mathbb{P}[U_k \in du \mid X_{0:k} = x_{0:k}, Y_{0:k} = y_{0:k}] = p_k^{U|V}(y_k - h_k(x_k), du) .$$

(ix) **En reportant cette expression exprimée à l'instant $(k-1)$ dans l'expression obtenue à la question (vii), montrer que**

$$\begin{aligned} & \mathbb{P}[X_k \in dx_k \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ &= \mathbb{P}[X_k \in dx_k \mid X_{k-1} = x_{k-1}, Y_{k-1} = y_{k-1}] = P_k(y_{k-1}, x_{k-1}, dx_k) , \end{aligned}$$

avec un noyau de transition paramétré par y_{k-1} , et défini de manière implicite par

$$\int P_k(y_{k-1}, x_{k-1}, dx_k) \phi(x_k) = \int \phi(f_k(x_{k-1}, u)) p_{k-1}^{U|V}(y_{k-1} - h_{k-1}(x_{k-1}), du) ,$$

pour toute fonction test ϕ . Montrer qu'il est facile de simuler un vecteur aléatoire selon la distribution de probabilité $P_k(y, x, dx')$ ainsi définie, pour tout $x \in \mathbb{R}^m$ et tout $y \in \mathbb{R}^d$.

On s'intéresse enfin au modèle C, vu d'abord comme un cas particulier du modèle B.

- (x) **En utilisant l'expression de la distribution de probabilité conditionnelle $p_k^{U|V}(v, du)$ dans le cas gaussien correspondant au modèle C, donner la forme particulière que prend dans ce cas le noyau de transition introduit à la question (ix).**

On a vu en cours que dans le cas gaussien, il est facile de simuler un vecteur aléatoire selon la distribution de probabilité conditionnelle $p_k^{U|V}(v, du)$, pour tout $v \in \mathbb{R}^d$, et on considérera donc ce point comme acquis.

- (xi) **Compte tenu que le modèle C peut aussi être interprété comme un cas particulier du modèle A, donner la forme particulière que prend dans ce cas le noyau de transition introduit à la question (vi), et comparer avec la réponse obtenue à la question précédente.**

EXERCICE 2 :

L'objectif de cet exercice est d'étudier un algorithme de simulation Monte Carlo pour une distribution de Gibbs–Boltzmann

$$\mu = g \cdot \eta = \frac{g \eta}{\langle \eta, g \rangle}, \quad (2)$$

appelé algorithme de *rejet contrôlé*, intermédiaire entre la méthode d'acceptation / rejet et la méthode d'échantillonnage pondéré. La constante de normalisation $\langle \eta, g \rangle$ n'est pas nécessairement connue, et on connaît seulement un majorant M de la fonction positive bornée g .

L'algorithme de rejet contrôlé est défini de la manière suivante, pour une constante positive $c > 0$ donnée : on simule une variable aléatoire Ξ selon la distribution de probabilité η , et avec une probabilité

$$p_c(\Xi) = \min\left(1, \frac{g(\Xi)}{c}\right),$$

on accepte Ξ , c'est-à-dire qu'on pose $X = \Xi$, et sinon on recommence.

Le rôle de la constante positive $c > 0$ est d'accepter toute proposition Ξ dont le poids $g(\Xi) \geq c$ est supérieur au seuil, et d'accepter avec probabilité $0 \leq g(\Xi)/c < 1$ seulement une proposition Ξ dont le poids $0 \leq g(\Xi) < c$ est inférieur au seuil.

- (i) **Montrer que cet algorithme correspond à l'algorithme d'acceptation / rejet pour la distribution de Gibbs–Boltzmann**

$$\eta_c = p_c \cdot \eta = \frac{p_c \eta}{\langle \eta, p_c \rangle} \quad \text{avec} \quad p_c(x) = \min\left(1, \frac{g(x)}{c}\right), \quad (3)$$

c'est-à-dire en particulier que la variable aléatoire X simulée selon cet algorithme a pour distribution de probabilité η_c .

- (ii) **Comparer l'expression de la probabilité d'acceptation dans l'algorithme de rejet contrôlé (vu comme algorithme d'acceptation / rejet pour la décomposition d'importance (3)) avec l'expression de la probabilité d'acceptation dans l'algorithme d'acceptation / rejet pour la décomposition d'importance d'origine (2).**

On introduit les approximations suivantes

$$\mu \approx \frac{\sum_{i=1}^N \frac{g(\xi^i)}{p_c(\xi^i)} \delta_{\xi^i}}{\sum_{i=1}^N \frac{g(\xi^i)}{p_c(\xi^i)}} = \sum_{i=1}^N w^i \delta_{\xi^i} \quad \text{et} \quad \langle \eta, g \rangle \approx \frac{1}{N} \sum_{i=1}^N \frac{g(\xi^i)}{p_c(\xi^i)},$$

pour la distribution de Gibbs–Boltzmann et pour la constante de normalisation, respectivement, où les variables aléatoires (ξ^1, \dots, ξ^N) sont i.i.d. de distribution de probabilité commune η_c , et où les poids positifs (w^1, \dots, w^N) sont définis par

$$w^i = \frac{\frac{g(\xi^i)}{p_c(\xi^i)}}{\sum_{j=1}^N \frac{g(\xi^j)}{p_c(\xi^j)}} \quad \text{pour tout } i = 1, \dots, N.$$

- (iii) Dans le cas particulier où $c \geq M$, montrer que η_c coïncide avec μ , et en déduire que l’approximation proposée se réduit dans ce cas à la distribution empirique associée à un échantillon généré selon l’algorithme d’acceptation / rejet.
- (iv) Dans le cas limite où $c \rightarrow 0$, montrer que η_c coïncide avec η , et en déduire que l’approximation proposée se réduit dans ce cas à la distribution empirique pondérée obtenue par échantillonnage pondéré.
- (v) Montrer que la distribution μ peut aussi se représenter comme la distribution de Gibbs–Boltzmann

$$\mu = \frac{g}{p_c} \cdot \eta_c = \frac{\frac{g}{p_c} \eta_c}{\langle \eta_c, \frac{g}{p_c} \rangle},$$

et en déduire que l’approximation proposée peut s’interpréter comme un algorithme d’échantillonnage pondéré associé à cette nouvelle décomposition d’importance.

On rappelle que la distance du χ^2 définie par

$$\chi^2(\mu, \mu') = \int_E \left(\frac{d\mu}{d\mu'}(x) \right)^2 \mu'(dx) - 1,$$

permet d’évaluer la qualité de la distribution d’importance μ' pour l’approximation de la distribution de probabilité cible μ . On admettra que les variables aléatoires

$$\min(g(\Xi), c) \quad \text{et} \quad \max(g(\Xi), c) g(\Xi),$$

sont positivement corrélées, c’est-à-dire que

$$\mathbb{E}[\min(g(\Xi), c)] \mathbb{E}[\max(g(\Xi), c) g(\Xi)] \leq \mathbb{E}[\min(g(\Xi), c) \max(g(\Xi), c) g(\Xi)].$$

- (vi) Comparer les distances $\chi^2(\mu, \eta)$ et $\chi^2(\mu, \eta_c)$ entre la distribution de Gibbs–Boltzmann μ et chacune des deux distributions d'importance proposées, η et η_c .
- (vii) Au vu des réponses aux questions (ii) et (vi), que penser de l'algorithme de rejet contrôlé, par rapport à l'algorithme d'acceptation / rejet et à l'algorithme d'échantillonnage pondéré ?