École Nationale Supérieure de Techniques Avancées
Filière : Finance quantitative
Module : Automatique avancée

Examen du cours B7–3 “Filtrage bayésien optimal et approximation particulaire”
Lundi 16 octobre 2006, 13:30 à 16:30

Exercice :

Soit μ une distribution de probabilité sur un espace E, et soit g une fonction positive bornée par 1, telle que $P = \langle \mu, g \rangle > 0$. On suppose que $P \ll 1$, et on veut

- générer un échantillon distribué (approximativement) selon la distribution de Gibbs

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

- et estimer la constante de normalisation (ou fonction de partition) $P = \langle \mu, g \rangle$.

On suppose qu'il existe un noyau markovien M réversible pour la distribution de probabilité μ, c'est–à–dire que

$$\mu(dx)\, M(x, dx') = \mu(dx')\, M(x', dx) \; .$$

(i) **Montrer que le noyau markovien M laisse invariante la distribution de probabilité μ.**

On introduit une suite croissante

$$0 = \beta_0 < \beta_1 < \cdots < \beta_n = 1 \; ,$$

et pour tout $k = 0, 1, \cdots, n$, on considère la distribution de probabilité

$$\mu_k = g^{\beta_k} \cdot \mu = \frac{g^{\beta_k}\, \mu}{\langle \mu, g^{\beta_k} \rangle} \; ,$$

intermédiaire entre $\mu_0 = \mu$ et $\mu_n = g \cdot \mu$.

(ii) **Montrer que**

$$\mu_k = g_k \cdot \mu_{k-1} \ ,$$

avec une fonction de sélection g_k positive, dont on donnera l'expression.

(iii) **Montrer que pour tout $k = 0, 1, \cdots, n$, le noyau markovien**

$$M_k(x, dx') = M(x, dx')\, g^{\beta_k}(x') + (1 - M\, g^{\beta_k}(x))\, \delta_x(dx') \ ,$$

laisse invariante la distribution de probabilité μ_k, et montrer comment simuler, pour $x \in E$ fixé, une variable aléatoire de loi $M_k(x, dx')$.

(iv) **Exprimer μ_k en fonction de μ_{k-1}, avec une étape de mutation faisant intervenir le noyau markovien $M_{k-1}(x, dx')$.**

(v) **Décrire l'algorithme SIR standard pour l'approximation particulaire du flot normalisé μ_n et de la constante de normalisation P.**

Problème :

L'objectif de ce problème est de montrer que pour une certaine classe de systèmes non–linéaires (qu'on peut qualifier de systèmes conditionnellement linéaires gaussiens), le filtre optimal peut s'exprimer, en utilisant la structure particulière du problème, à l'aide d'un filtre optimal sur un espace de dimension réduite, paramétré en chaque point par une distribution de probabilité gaussienne (caractérisée par sa moyenne et sa matrice de covariance) dans l'espace supplémentaire.

On considère le système suivant, où l'état caché $X_k = (X_k^1, X_k^2)$ est partitionné en deux composantes, de dimension m_1 et m_2 respectivement

$$X_k^1 = F_1\, X_{k-1}^1 + f_1 + W_k^1 \ ,$$

$$X_k^2 = F_2\, X_{k-1}^2 + f_2 + W_k^2 \ ,$$

et où l'observation Y_k est relié à l'état caché X_k par la relation

$$Y_k = H\, X_k^1 + h(X_k^2) + V_k \ ,$$

et on note $E_1 = \mathbb{R}^{m_1}$, $E_2 = \mathbb{R}^{m_2}$ et $E = E_1 \times E_2$. On suppose que

- la variable aléatoire X_0^1 est gaussienne, de moyenne m_0^1 et de matrice de covariance P_0^1,
- la suite $\{W_k^1 \, , \, k \geq 1\}$ est un bruit blanc gaussien centré, de matrice de covariance Σ_1 pour tout $k \geq 0$,
- la suite $\{W_k^2 \, , \, k \geq 1\}$ est un bruit blanc gaussien centré, de matrice de covariance Σ_2 pour tout $k \geq 0$,
- la suite $\{V_k \, , \, k \geq 0\}$ est un bruit blanc gaussien centré, de matrice de covariance R *inversible* pour tout $k \geq 0$,
- les bruits $\{W_k^1 \, , \, k \geq 1\}$, $\{W_k^2 \, , \, k \geq 1\}$ et $\{V_k \, , \, k \geq 0\}$ et les conditions initiales X_0^1 et X_0^2 sont mutuellement indépendants,

et on note $\eta_0^2(dx_2)$ la distribution de probabilité de la variable aléatoire X_0^2.

Intuitivement, il s'agit d'exploiter la remarque suivante : si la suite $\{X_k^2 \, , \, k \geq 0\}$ est observée, alors le système ci–dessus est linéaire et gaussien.

Notation : Dans tout le problème, on utilise la notation $\Gamma(dx, m, P)$ pour désigner la distribution de probabilité gaussienne de moyenne m et de matrice de covariance P, et si la matrice P est *inversible* alors on utilise la notation $q(x-m, P)$ pour désigner la densité de probabilité correspondante, c'est–à–dire

$$q(x-m, P) \propto \frac{1}{\sqrt{\det P}} \ \exp\{-\tfrac{1}{2}\,(x-m)^*\,P^{-1}\,(x-m)\,\}\ .$$

(i) **Montrer que les deux suites $\{X_k^1\,,\,k \geq 0\}$ et $\{X_k^2\,,\,k \geq 0\}$ forment deux chaînes de Markov indépendantes. Donner l'expression de leur distributions de probabilité initiales et de leurs noyaux de transition respectifs $Q^1(x_1, dx'_1)$ et $Q^2(x_2, dx'_2)$.**

(ii) **Montrer que la probabilité d'émission possède une densité**

$$\mathbb{P}[Y_k \in dy \mid X_k^1 = x_1, X_k^2 = x_2] = g(y, x_1, x_2)\ dy\ ,$$

dont on donnera l'expression. En déduire l'expression de la fonction de vraisemblance définie par

$$g_k(x_1, x_2) = g(Y_k, x_1, x_2)\ .$$

Flot paramétré

On introduit le flot linéaire (non–normalisé) défini par

$$\begin{aligned}\langle \gamma_n, \phi \rangle \ &= \int_E \cdots \int_E \phi(x_n^1, x_n^2)\ \prod_{k=0}^n g_k(x_k^1, x_k^2)\ \mathbb{P}[X_{0:n}^1 \in dx_{0:n}^1, X_{0:n}^2 \in dx_{0:n}^2] \\ &= \mathbb{E}[\phi(X_n)\ \prod_{k=0}^n g_k(X_k)\,]\ ,\end{aligned}$$

pour toute fonction mesurable bornée ϕ définie sur E, où l'espérance porte seulement sur les variables $X_{0:n}$, les variables $Y_{0:n}$ étant considérées comme fixées.

(iii) **Montrer que le flot linéaire peut aussi s'écrire**

$$\langle \gamma_n, \phi \rangle = \mathbb{E}[\phi_2(X_n^2)\ \langle \gamma_n^{1|2}, \phi_1 \rangle\,]\ ,$$

pour toute fonction mesurable bornée $\phi = \phi_1 \otimes \phi_2$ définie sur $E = E_1 \times E_2$, et en particulier

$$\langle \gamma_n, 1 \rangle = \mathbb{E}[\,\langle \gamma_n^{1|2}, 1 \rangle\,]\ ,$$

pour $\phi \equiv 1$, où le flot linéaire paramétré est défini par

$$\begin{aligned}\langle \gamma_n^{1|2}, \phi_1 \rangle &= \int_{E_1} \cdots \int_{E_1} \phi_1(x_n^1) \prod_{k=0}^{n} g_k(x_k^1, X_k^2)\, \mathbb{P}[X_{0:n}^1 \in dx_{0:n}^1] \\ &= \mathbb{E}[\phi_1(X_n^1) \prod_{k=0}^{n} g_k^{1|2}(X_k^1)]\ ,\end{aligned}$$

pour toute fonction mesurable bornée ϕ_1 définie sur E_1, où cette dernière espérance porte seulement sur les variables $X_{0:n}^1$, les variables $Y_{0:n}$ et $X_{0:n}^2$ étant considérées comme fixées. On donnera en particulier l'expression de la fonction de vraisemblance $g_k^{1|2}(x_1)$.

On définit aussi

$$\begin{aligned}\langle \gamma_{n|n-1}^{1|2}, \phi_1 \rangle &= \int_{E_1} \cdots \int_{E_1} \phi_1(x_n^1) \prod_{k=0}^{n-1} g_k(x_k^1, X_k^2)\, \mathbb{P}[X_{0:n}^1 \in dx_{0:n}^1] \\ &= \mathbb{E}[\phi_1(X_n^1) \prod_{k=0}^{n-1} g_k^{1|2}(X_k^1)]\ ,\end{aligned}$$

pour toute fonction mesurable bornée ϕ_1 définie sur E_1, et les flots normalisés $\mu_n^{1|2}$ et $\eta_n^{1|2} = \mu_{n|n-1}^{1|2}$ associés.

(iv) **Etablir l'équation récurrente**

$$\mu_{k-1}^{1|2} \xrightarrow{\text{prédiction}} \eta_k^{1|2} = \mu_{k|k-1}^{1|2} = \mu_{k-1}^{1|2}\, Q^1 \xrightarrow{\text{correction}} \mu_k^{1|2} = g_k^{1|2} \cdot \eta_k^{1|2}\ ,$$

pour tout $k \geq 1$, avec la condition initiale

$$\mu_0^{1|2} = g_0^{1|2} \cdot \eta_0^{1|2} \qquad \textbf{et} \qquad \eta_0^{1|2} = \eta_0^1\ ,$$

pour $k = 0$, et montrer que la constante de normalisation vérifie

$$\langle \gamma_n^{1|2}, 1 \rangle = \prod_{k=0}^{n} \langle \eta_k^{1|2}, g_k^{1|2} \rangle\ .$$

On pourra exprimer $\gamma_{k|k-1}^{1|2}$ en fonction de $\gamma_{k-1}^{1|2}$ et $\gamma_k^{1|2}$ en fonction de $\gamma_{k|k-1}^{1|2}$, puis normaliser.

On admet que les flots normalisés $\mu_n^{1|2}$ et $\mu_{n|n-1}^{1|2}$ associés sont des distributions de probabilités gaussiennes

$$\mu_n^{1|2}(dx_1) = \Gamma(dx_1, \widehat{X}_n^{1|2}, P_n^{1|2}) \qquad \text{et} \qquad \mu_{n|n-1}^{1|2}(dx_1) = \Gamma(dx_1, \widehat{X}_{n|n-1}^{1|2}, P_{n|n-1}^{1|2})\ ,$$

de moyenne $\widehat{X}_n^{1|2}$ et $\widehat{X}_{n|n-1}^{1|2}$ respectivement, et de matrice de covariance $P_n^{1|2}$ et $P_{n|n-1}^{1|2}$ respectivement, avec des récurrences qui sont essentiellement celles du filtre de Kalman. On introduit à cet effet les applications suivantes : prédiction

$$m_1 \longmapsto M^{\rm pr}(m_1) = F_1\, m_1 + f_1 \ ,$$

et correction (mise–à–jour)

$$(x_2, m_1) \longmapsto M_k^{\rm up}(x_2, m_1) = m_1 + P_{k|k-1}^{1|2}\, H^*\, \Xi_k^{-1}\, (Y_k - H\, m_1 - h(x_2)) \ ,$$

et les matrices de covariance associées

$$P_{k|k-1}^{1|2} = F_1\, P_{k-1}^{1|2}\, F_1^* + \Sigma_1 \ ,$$

et

$$P_k^{1|2} = P_{k|k-1}^{1|2} - P_{k|k-1}^{1|2}\, H^*\, \Xi_k^{-1}\, H\, P_{k|k-1}^{1|2} \ ,$$

ou par définition

$$\Xi_k = H\, P_{k|k-1}^{1|2}\, H^* + R$$

On a alors

$$\widehat{X}_{k|k-1}^{1|2} = M^{\rm pr}(\widehat{X}_{k-1}^{1|2}) \qquad \text{et} \qquad \widehat{X}_k^{1|2} = M_k^{\rm up}(X_k^2, \widehat{X}_{k|k-1}^{1|2}) \ ,$$

avec la condition initiale

$$\widehat{X}_{0|-1}^{1|2} = m_0^1 \qquad \text{et} \qquad P_{0|-1}^{1|2} = P_0^1 \ ,$$

pour $k = 0$.

(v) **Montrer que**

$$\langle \eta_k^{1|2}, g_k^{1|2} \rangle = \int_{E_1} \Gamma(dx_1, \widehat{X}_{k|k-1}^{1|2}, P_{k|k-1}^{1|2})\; g_k^{1|2}(x_1) = q(Y_k - H\, m_k^{1|2} - h(X_k^2), \Xi_k) \ ,$$

avec la notation $m_k^{1|2} = \widehat{X}_{k|k-1}^{1|2}$ **pour tout** $k \geq 1$**, et**

$$\langle \eta_0^{1|2}, g_0^{1|2} \rangle = \int_{E_1} \Gamma(dx_1, m_0^1, P_0^1)\; g_0^{1|2}(x_1) = q(Y_0 - H\, m_0^1 - h(X_0^2), \Xi_0) \ ,$$

pour $k = 0$**.**

Flot hybride

On pose $X_k^{2,1} = (X_k^2, m_k^{1|2})$ pour tout $k \geq 0$, et en particulier $X_0^{2,1} = (X_0^2, m_0^1)$ pour $k = 0$.

(vi) **Montrer que la suite $\{X_k^{2,1}\ ,\ k \geq 0\}$ est une chaîne de Markov, à valeurs dans $E^{2,1} = E_2 \times E_1$, de distribution de probabilité initiale**

$$\eta_0^{2,1}(dx_2, dm_1) = \mathbb{P}[X_0^2 \in dx_2, m_0^{1|2} \in dm_1] = \eta_0^2(dx_2)\ \delta_{m_0^1}(dm_1)\ ,$$

et de noyau de transition

$$\begin{aligned} Q_k^{2,1}(x_2, m_1, dx'_2, dm'_1) &= \mathbb{P}[X_k^2 \in dx'_2, m_k^{1|2} \in dm'_1 \mid X_{k-1}^2 = x_2, m_{k-1}^{1|2} = m_1] \\ &= \Gamma(dx'_2, F_2\, x_2 + f_2, \Sigma_2)\ \delta_{M^{\mathrm{pr}} \circ M_{k-1}^{\mathrm{up}}(x_2, m_1)}(dm'_1)\ . \end{aligned}$$

On pose

$$g_k^{2,1}(x_2, m_1) = q(Y_k - H\, m_1 - h(x_2), \Xi_k)\ ,$$

pour tout $k \geq 0$, et on introduit le flot linéaire hybride défini par

$$\langle \gamma_n^{2,1}, F \rangle = \mathbb{E}[F(X_n^{2,1}) \prod_{k=0}^n g_k^{2,1}(X_k^{2,1})]\ ,$$

pour toute fonction F mesurable bornée définie sur $E^{2,1} = E_2 \times E_1$, et en particulier

$$\langle \gamma_n^{2,1}, 1 \rangle = \mathbb{E}[\prod_{k=0}^n g_k^{2,1}(X_k^{2,1})]\ ,$$

pour $F \equiv 1$.

(vii) **Montrer que**

$$\prod_{k=0}^n g_k^{2,1}(X_k^{2,1}) = \prod_{k=0}^n \langle \eta_k^{1|2}, g_k^{1|2} \rangle = \langle \gamma_n^{1|2}, 1 \rangle\ ,$$

et en déduire que

$$\langle \gamma_n^{2,1}, 1 \rangle = \langle \gamma_n, 1 \rangle\ ,$$

c'est–à–dire que la constante de normalisation du flot linéaire hybride coïncide avec la constante de normalisation du flot linéaire introduit au début du problème.

On définit aussi

$$\langle \gamma_{n|n-1}^{2,1}, F \rangle = \mathbb{E}[F(X_n^{2,1}) \prod_{k=0}^{n-1} g_k^{2,1}(X_k^{2,1})]\ ,$$

pour toute fonction F mesurable bornée définie sur $E^{2,1} = E_2 \times E_1$, et les flots normalisés $\mu_n^{2,1}$ et $\eta_n^{2,1} = \mu_{n|n-1}^{2,1}$ associés.

(viii) **Etablir l'équation récurrente**

$$\mu_{k-1}^{2,1} \xrightarrow{\text{prédiction}} \eta_k^{2,1} = \mu_{k|k-1}^{2,1} = \mu_{k-1}^{2,1}\, Q_k^{2,1} \xrightarrow{\text{correction}} \mu_k^{2,1} = g_k^{2,1} \cdot \eta_k^{2,1} \ ,$$

pour tout $k \geq 1$, avec la condition initiale

$$\mu_0^{2,1} = g_0^{2,1} \cdot \eta_0^{2,1} \ ,$$

pour $k = 0$, et montrer que la constante de normalisation vérifie

$$\langle \gamma_n^{2,1}, 1 \rangle = \prod_{k=0}^{n} \langle \eta_k^{2,1}, g_k^{2,1} \rangle \ .$$

On pourra exprimer $\gamma_{k|k-1}^{2,1}$ en fonction de $\gamma_{k-1}^{2,1}$ et $\gamma_k^{2,1}$ en fonction de $\gamma_{k|k-1}^{2,1}$, puis normaliser.

(ix) **Si dans la définition du flot linéaire hybride on fait le choix**

$$F(x_2, m_1) = \phi_2(x_2) \int_{E_1} \Gamma(dx_1, M_n^{\rm up}(x_2, m_1), P_n^{1|2})\, \phi_1(x_1) \ ,$$

montrer que

$$\langle \gamma_n^{2,1}, F \rangle = \langle \gamma_n, \phi \rangle \ ,$$

pour toute fonction mesurable bornée $\phi = \phi_1 \otimes \phi_2$ définie sur $E = E_1 \times E_2$. Plus généralement si

$$T_n\, \phi(x_2, m_1) = \int_{E_1} \Gamma(dx_1, M_n^{\rm up}(x_2, m_1), P_n^{1|2})\, \phi(x_1, x_2) \ ,$$

montrer que

$$\langle \gamma_n^{2,1}, T_n\, \phi \rangle = \langle \gamma_n, \phi \rangle \ ,$$

pour toute fonction mesurable bornée ϕ définie sur E, c'est–à–dire que

$$\gamma_n = \gamma_n^{2,1}\, T_n \qquad \textbf{et} \qquad \mu_n = \mu_n^{2,1}\, T_n \ .$$

Approximation particulaire

Pour approcher numériquement la constante de normalisation $\langle \gamma_n, 1 \rangle$ et le flot normalisé μ_n, on peut

- utiliser une approche directe,

- ou bien approcher numériquement la constante de normalisation $\langle \gamma_n^{2,1}, 1 \rangle$ et le flot normalisé $\mu_n^{2,1}$, au vu de la réponse aux questions (vii) et (ix).

On recherche donc une approximation du flot non–linéaire hybride $\mu_k^{2,1}$ sous la forme de la distribution de probabilité empirique pondérée

$$\mu_k^{2,1} \approx \mu_k^{2,1,N} = \sum_{i=1}^N w_k^i \, \delta_{(\xi_k^i,\, m_k^i)} \qquad \text{avec} \qquad \sum_{i=1}^N w_k^i = 1 \ ,$$

associée à une population de N particules caractérisée par leurs positions dans $E^{2,1} = E_2 \times E_1$ et par leurs poids positifs.

(x) **En déduire l'approximation suivante**

$$\mu_k^N = \mu_k^{2,1,N} \, T_k = \sum_{i=1}^N w_k^i \, \delta_{\xi_k^i} \, \Gamma(\,\cdot\,, M_k^{\rm up}(\xi_k^i, m_k^i), P_k^{1|2}) \ ,$$

pour le flot normalisé $\mu_k = \mu_k^{2,1} \, T_k$. Décrire les approximations particulaires des distributions de probabilité marginales sur E_1 et sur E_2.

(xi) **Décrire l'algorithme SIR standard pour l'approximation particulaire du flot normalisé $\mu_k^{2,1}$ et de la constante de normalisation $\langle \gamma_k^{2,1}, 1 \rangle$.**

Dans cet algorithme, la composante appartenant au sous–espace E_2 évolue donc de façon aléatoire, et fait l'objet d'une approximation particulaire, dans laquelle les particules explorent le sous–espace E_2, et se voient attachées une autre composante appartenant au sous–espace E_1 définie par des relations déterministes, essentiellement celles de filtres de Kalman. En ce sens, cet algorithme hybride exploite au mieux la structure particulière du système conditionnellement linéaire gaussien considéré dans ce problème, puisque moins de particules sont en principe nécessaires pour explorer convenablement le sous–espace E_2, plutôt que l'espace complet $E = E_1 \times E_2$.