

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.**

SOLUTION

De l'équation d'observation exprimée à l'instant $(k - 1)$, on tire

$$V_{k-1} = Y_{k-1} - h_{k-1}(X_{k-1}) ,$$

et en reportant cette expression dans l'équation d'état, on obtient le système

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

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

Le vecteur aléatoire (W_k, V_k) est indépendant de l'état joint précédent (X_{k-1}, Y_{k-1}) , compte tenu que celui-ci s'exprime en fonction des vecteurs aléatoires $X_0, (W_1, \dots, W_{k-1})$ et (V_0, \dots, V_{k-1}) . On en déduit 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

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

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

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.**

SOLUTION

En reportant dans l'équation d'observation l'expression de X_k tirée de l'équation d'état, on obtient le système

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

$$Y_k = h_k(f_k(X_{k-1}, U_{k-1})) + V_k ,$$

ce qui montre que l'état joint présent (X_k, Y_k) dépend de (la composante cachée X_{k-1} de) l'état joint précédent (X_{k-1}, Y_{k-1}) et du vecteur aléatoire (U_{k-1}, V_k) . Par construction, le vecteur aléatoire U_{k-1} et (la composante observée Y_{k-1} de) l'état joint précédent (X_{k-1}, Y_{k-1}) dépendent tous les deux du vecteur aléatoire V_{k-1} , et ne sont donc pas indépendants.

□

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

$$X_k = f_k(X_{k-1}, U_{k-1}) , \quad (1)$$

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

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 .

SOLUTION

Clairement, le vecteur aléatoire

$$\begin{pmatrix} W_k \\ V_{k-1} \end{pmatrix} = \begin{pmatrix} I & -A_k \\ 0 & I \end{pmatrix} \begin{pmatrix} U_{k-1} \\ V_{k-1} \end{pmatrix} ,$$

est gaussien, comme transformation linéaire du vecteur aléatoire (U_{k-1}, V_{k-1}) . En exprimant l'indépendance (et en particulier la décorrélation) des vecteurs aléatoires gaussiens V_{k-1} et W_k , on obtient

$$S_{k-1} = A_k R_{k-1} \quad \text{c'est-à-dire que} \quad A_k = S_{k-1} R_{k-1}^{-1} ,$$

compte tenu que la matrice R_{k-1} est inversible. Par définition, la matrice de covariance Σ_k du vecteur aléatoire gaussien W_k est donnée par

$$\begin{aligned} \Sigma_k &= \mathbb{E}[W_k W_k^*] = \mathbb{E}[(U_{k-1} - A_k V_{k-1}) (U_{k-1} - A_k V_{k-1})^*] \\ &= \mathbb{E}[U_{k-1} U_{k-1}^*] - A_k \mathbb{E}[V_{k-1} U_{k-1}^*] - \mathbb{E}[U_{k-1} V_{k-1}^*] A_k^* + A_k \mathbb{E}[V_{k-1} V_{k-1}^*] A_k^* \\ &= Q_{k-1} - A_k S_{k-1} - S_{k-1} A_k^* + A_k R_{k-1} A_k^* \\ &= Q_{k-1} - S_{k-1} R_{k-1}^{-1} S_{k-1}^* . \end{aligned}$$

□

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

SOLUTION

En reportant la décomposition du vecteur aléatoire U_{k-1} obtenue à la question (iii) dans l'équation d'état (1), on obtient le système suivant

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

qui est bien de la forme du modèle A, avec l'identification

$$f_k(x, v, w) \equiv f_k(x, A_k v + w) ,$$

et on vérifie que les deux suites $\{W_k\}$ et $\{V_k\}$ sont deux bruits blancs indépendants et gaussiens (avec la réserve que le caractère gaussien est sans importance ici).

□

Pour les trois modèles considérés, la décomposition

$$\begin{aligned} &\mathbb{P}[X_k \in dx_k, Y_k \in dy_k \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ &= \mathbb{P}[X_k \in dx_k \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ &\quad \mathbb{P}[Y_k \in dy_k \mid X_{0:k} = x_{0:k}, Y_{0:k-1} = y_{0:k-1}] , \end{aligned} \tag{2}$$

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**

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

Dans les trois modèles

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

qui ne dépend que de x_k seulement, où la seconde égalité résulte de l'indépendance du vecteur aléatoire V_k et des états joints passés $(X_{0:k}, Y_{0:k-1})$, et où la dernière égalité résulte du changement de variable $y_k = h_k(x_k) + v'$, de sorte que

$$\begin{aligned}
& \mathbb{P}[Y_k \in dy_k \mid X_{0:k} = x_{0:k}, Y_{0:k-1} = y_{0:k-1}] \\
&= \mathbb{P}[Y_k \in dy_k \mid X_k = x_k] = q_k^V(y_k - h_k(x_k)) dy_k .
\end{aligned} \tag{3}$$

□

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 $y \in \mathbb{R}^d$.

Par définition

$$\begin{aligned}
& \mathbb{E}[\phi(X_k) \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\
&= \mathbb{E}[\phi(f_k(X_{k-1}, Y_{k-1} - h_{k-1}(X_{k-1}), W_k)) \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\
&= \int \phi(f_k(x_{k-1}, y_{k-1} - h_{k-1}(x_{k-1}), w)) \mathbb{P}[W_k \in dw] \\
&= \int \phi(f_k(x_{k-1}, y_{k-1} - h_{k-1}(x_{k-1}), w)) p_k^W(dw) ,
\end{aligned}$$

qui ne dépend que de (x_{k-1}, y_{k-1}) seulement, où la seconde égalité résulte de l'indépendance du vecteur aléatoire W_k et des états joints passés $(X_{0:k-1}, Y_{0:k-1})$, de sorte que

$$\begin{aligned}
& \mathbb{P}[X_k \in dx_k \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\
&= \mathbb{P}[X_k \in dx_k \mid X_{k-1} = x_{k-1}, Y_{k-1} = y_{k-1}] = P_k(y_{k-1}, x_{k-1}, dx_k) ,
\end{aligned} \tag{4}$$

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

$$\int P_k(y_{k-1}, x_{k-1}, dx_k) \phi(x_k) = \int \phi(f_k(x_{k-1}, y_{k-1} - h_{k-1}(x_{k-1}), w)) p_k^W(dw) ,$$

pour toute fonction test ϕ . Pour simuler un vecteur aléatoire X' selon la distribution de probabilité $P_k(y, x, dx')$, il suffit de simuler un vecteur aléatoire W selon la distribution de probabilité $p_k^W(dw)$, ce qui est supposé facile par hypothèse, et de poser

$$X' = f_k(x, y - h_{k-1}(x), W) ,$$

pour tout $x \in \mathbb{R}^m$ et tout $y \in \mathbb{R}^d$.

En reportant dans la décomposition (2) les identités (4) et (3), on obtient

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

qui ne dépend que de l'état joint précédent (x_{k-1}, y_{k-1}) , d'où on déduit que les états cachés et les observations forment conjointement une chaîne de Markov. □

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.

SOLUTION

Par définition

$$\begin{aligned} & \mathbb{E}[\phi(X_k) \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ &= \mathbb{E}[\phi(f_k(X_{k-1}, U_{k-1})) \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ &= \int \phi(f_k(x_{k-1}, u)) \mathbb{P}[U_{k-1} \in du \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] , \quad (5) \end{aligned}$$

où le vecteur aléatoire U_{k-1} et l'observation Y_{k-1} dépendent tous les deux du vecteur aléatoire V_{k-1} , et ne sont donc 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

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

On remarque que

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

où la seconde égalité résulte de l'indépendance du vecteur aléatoire (U_k, V_k) et des états joints passés et / ou présents $(X_{0:k}, Y_{0:k-1})$, et où la dernière égalité résulte du changement de variable $y_k = h_k(x_k) + v$, de sorte que

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

En intégrant par rapport à la variable u , on obtient

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

et il résulte de la formule de Bayes que

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

de sorte 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 $y \in \mathbb{R}^d$.

SOLUTION

En reportant dans l'équation (5) l'expression obtenue en réponse à la question (viii), on obtient

$$\begin{aligned} \mathbb{E}[\phi(X_k) \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ = \int \phi(f_k(x_{k-1}, u)) p_{k-1}^{U|V}(y_{k-1} - h_{k-1}(x_{k-1}), du) , \end{aligned}$$

qui ne dépend que de (x_{k-1}, y_{k-1}) seulement, de sorte que

$$\begin{aligned} \mathbb{P}[X_k \in dx_k \mid X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}] \\ = \mathbb{P}[X_k \in dx_k \mid X_{k-1} = x_{k-1}, Y_{k-1} = y_{k-1}] = P_k(y_{k-1}, x_{k-1}, dx_k) , \end{aligned} \tag{6}$$

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 . Pour simuler un vecteur aléatoire X' selon la distribution de probabilité $P_k(y, x, dx')$, il suffit de simuler un vecteur aléatoire U selon la distribution de probabilité $p_k^{U|V}(y - h_k(x), du)$, ce qui est supposé facile par hypothèse, et de poser

$$X' = f_k(x, U) ,$$

pour tout $x \in \mathbb{R}^m$ et tout $y \in \mathbb{R}^d$.

En reportant dans la décomposition (2) les identités (6) et (3), on obtient

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

qui ne dépend que de l'état joint précédent (x_{k-1}, y_{k-1}) , d'où on déduit que les états cachés et les observations forment conjointement une chaîne de Markov.

□

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).**

SOLUTION

Dans le cas gaussien, la distribution de probabilité conditionnelle $p_k^{U|V}(v, du)$ est gaussienne, de moyenne $S_k R_k^{-1} v$ et de matrice de covariance $Q_k - S_k R_k^{-1} S_k^*$. On en déduit que la distribution de probabilité conditionnelle $p_{k-1}^{U|V}(y_{k-1} - h_{k-1}(x_{k-1}), du)$ est gaussienne, de moyenne $A_k (y_{k-1} - h_{k-1}(x_{k-1}))$ et de matrice de covariance Σ_k . En reportant cette expression dans la définition du noyau de transition introduit à la question (ix), on obtient

$$\begin{aligned} & \int P_k(y_{k-1}, x_{k-1}, dx_k) \phi(x_k) \\ &= \int \phi(f_k(x_{k-1}, u)) \Gamma(du, A_k (y_{k-1} - h_{k-1}(x_{k-1})), \Sigma_k) \\ &= \int \phi(f_k(x_{k-1}, A_k (y_{k-1} - h_{k-1}(x_{k-1})) + w)) \Gamma(dw, 0, \Sigma_k) , \end{aligned}$$

où $\Gamma(dw, m, \Sigma)$ désigne la distribution de probabilité gaussienne de moyenne m et de matrice de covariance Σ .

□

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.**

SOLUTION

Il résulte de la réponse faite à la question (iv) que le modèle C peut également être interprété comme un cas particulier du modèle A, avec l'identification

$$f_k(x, v, w) \equiv f_k(x, A_k v + w) .$$

En faisant la substitution dans la définition du noyau de transition introduit à la question (vi), on obtient

$$\begin{aligned}
 & \int P_k(y_{k-1}, x_{k-1}, dx_k) \phi(x_k) \\
 &= \int \phi(f_k(x_{k-1}, y_{k-1} - h_{k-1}(x_{k-1}), w)) p_k^W(dw) \\
 &\equiv \int \phi(f_k(x_{k-1}, A_k(y_{k-1} - h_{k-1}(x_{k-1})) + w)) p_k^W(dw) ,
 \end{aligned}$$

où $p_k^W(dw)$ est la distribution de probabilité du vecteur aléatoire W_k , c'est-à-dire une distribution de probabilité gaussienne, centrée, de matrice de covariance Σ_k . On retrouve ainsi le résultat obtenu à la question (x).

□

EXERCICE 2 :

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

$$\mu = g \cdot \eta = \frac{g \eta}{\langle \eta, g \rangle}, \quad (7)$$

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**

$$\eta_c = p_c \cdot \eta = \frac{p_c \eta}{\langle \eta, p_c \rangle} \quad \text{avec} \quad p_c(x) = \min\left(1, \frac{g(x)}{c}\right), \quad (8)$$

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

SOLUTION

Il suffit ici de raisonner de manière constructive, en remarquant que $M = 1$ est un majorant naturel de la fonction positive p_c . Pour simuler une variable aléatoire selon la distribution de probabilité η_c , on peut en effet utiliser l'algorithme d'acceptation / rejet associé à la décomposition d'importance (8), ce qui se traduit par les étapes suivantes : on simule indépendamment une variable aléatoire Ξ selon la distribution de probabilité η et une variable U uniforme sur $[0, 1]$: si $p_c(\Xi) \geq U$ (c'est-à-dire avec une probabilité $p_c(\Xi)$) on accepte Ξ , c'est-à-dire qu'on pose $X = \Xi$, et sinon on recommence.

Les étapes ainsi décrites coïncident exactement avec les étapes de l'algorithme de rejet contrôlé. En conséquence, une variable aléatoire simulée selon l'algorithme de rejet contrôlé ne diffère pas d'une variable aléatoire qui serait simulée selon l'algorithme

d'acceptation / rejet associé à la décomposition d'importance (8), et cette variable a pour distribution de probabilité η_c , par construction.

□

- (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 (8)) avec l'expression de la probabilité d'acceptation dans l'algorithme d'acceptation / rejet pour la décomposition d'importance d'origine (7).

SOLUTION

Par définition, la probabilité d'acceptation dans l'algorithme de rejet contrôlé (vu comme algorithme d'acceptation / rejet pour la décomposition d'importance (8)) est

$$\mathbb{P}[p_c(\Xi) \geq U] = \mathbb{E}[p_c(\Xi)] = \int_E \min(1, \frac{g(x)}{c}) \eta(dx) ,$$

tandis que la probabilité d'acceptation dans l'algorithme d'acceptation / rejet pour la décomposition d'importance d'origine (7) est

$$\mathbb{P}[g(\Xi) \geq MU] = \mathbb{E}[\frac{g(\Xi)}{M}] = \int_E \frac{g(x)}{M} \eta(dx) .$$

Si $0 < c \leq M$, alors

$$1 \geq \frac{g(x)}{M} \text{ et } \frac{g(x)}{c} \geq \frac{g(x)}{M} \quad \text{de sorte que} \quad \min(1, \frac{g(x)}{c}) \geq \frac{g(x)}{M} ,$$

pour tout $x \in E$, et on en déduit que

$$\int_E \min(1, \frac{g(x)}{c}) \eta(dx) \geq \int_E \frac{g(x)}{M} \eta(dx) ,$$

c'est-à-dire que la probabilité d'acceptation dans l'algorithme de rejet contrôlé (vu comme algorithme d'acceptation / rejet pour la décomposition d'importance (8)) est plus grande que probabilité d'acceptation dans l'algorithme d'acceptation / rejet pour la décomposition d'importance d'origine (7). Si en revanche $c \geq M$, alors

$$\min(1, \frac{g(x)}{c}) = \frac{g(x)}{c} \leq \frac{g(x)}{M} ,$$

pour tout $x \in E$, et on en déduit que

$$\int_E \min(1, \frac{g(x)}{c}) \eta(dx) \leq \int_E \frac{g(x)}{M} \eta(dx) ,$$

et la conclusion est inversée, c'est-à-dire que la probabilité d'acceptation dans l'algorithme de rejet contrôlé (vu comme algorithme d'acceptation / rejet pour la décomposition d'importance (8)) est plus petite que probabilité d'acceptation dans l'algorithme d'acceptation / rejet pour la décomposition d'importance d'origine (7). Il n'y a donc aucun intérêt à choisir un seuil $c \geq M$. Les deux probabilités d'acceptation sont bien sûr égales dans le cas particulier où $c = M$.

□

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.**

SOLUTION

Si $c \geq M$, alors nécessairement $p_c(x) = \frac{g(x)}{c}$ pour tout $x \in E$, de sorte que $\eta_c = \mu$, et le rapport $\frac{g(x)}{p_c(x)} \equiv c$ est constant pour tout $x \in E$. Dans ce cas, les variables aléatoires (ξ^1, \dots, ξ^N) sont simulées selon la distribution de probabilité μ en utilisant l'algorithme d'acceptation / rejet, et ne sont pas pondérées. L'approximation proposée se réduit donc à la distribution empirique (non-pondérée) associée à un échantillon généré selon la distribution de probabilité μ en utilisant 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.**

SOLUTION

Dans le cas limite où $c \rightarrow 0$, alors nécessairement $p_c(x) \equiv 1$ pour tout $x \in E$, de sorte que $\eta_c = \eta$ et $\frac{g(x)}{p_c(x)} = g(x)$ pour tout $x \in E$. Dans ce cas, la proposition est toujours acceptée, les variables aléatoires (ξ^1, \dots, ξ^N) sont simulées selon la distribution de probabilité η , et leurs poids (non-normalisés) sont obtenues en évaluant la fonction g . L'approximation proposée se réduit donc à la distribution empirique pondérée obtenue par échantillonnage pondéré, avec un échantillon généré selon la distribution de probabilité η et pondéré en évaluant la fonction positive g .

□

(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.

SOLUTION

Par définition

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

avec l'identité corespondante

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

pour les constantes de normalisation, de sorte que

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

Pour approcher la distribution de Gibbs–Boltzmann μ , on peut en effet utiliser l'algorithme d'échantillonnage pondéré associé à cette nouvelle décomposition d'importance, ce qui se traduit par l'approximation suivante

$$\mu \approx \frac{\sum_{i=1}^N \frac{g(\xi^i)}{p_c(\xi^i)} \delta_{\xi^i}}{\sum_{i=1}^N \frac{g(\xi^i)}{p_c(\xi^i)}} = \sum_{i=1}^N w^i \delta_{\xi^i},$$

où les variables aléatoires (ξ^1, \dots, ξ^N) sont i.i.d. de distribution de probabilité commune η_c , par exemple obtenu par l'algorithme d'acceptation / rejet ou de manière équivalente (au vu de la réponse à la question (i)) par l'algorithme de rejet contrôlé, et où les poids positifs (w^1, \dots, w^N) sont définis par

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

Les étapes ainsi décrites coïncident exactement avec les étapes de l'approximation proposée. □

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 .**

SOLUTION

On remarque que

$$\left\langle \eta_c, \frac{g}{p_c} \right\rangle = \frac{\langle \eta, g \rangle}{\langle \eta, p_c \rangle},$$

de sorte que

$$\frac{d\mu}{d\eta_c} = \frac{\frac{g}{p_c}}{\left\langle \eta_c, \frac{g}{p_c} \right\rangle} = \frac{g}{\langle \eta, g \rangle} / \frac{p_c}{\langle \eta, p_c \rangle} = \frac{\langle \eta, p_c \rangle}{\langle \eta, g \rangle} \frac{g}{p_c},$$

et par définition

$$\begin{aligned} 1 + \chi^2(\mu, \eta_c) &= \int_E \left(\frac{d\mu}{d\eta_c}(x) \right)^2 \eta_c(dx) = \frac{\langle \eta, p_c \rangle^2}{\langle \eta, g \rangle^2} \int_E \frac{g^2(x)}{p_c^2(x)} \frac{p_c(x)}{\langle \eta, p_c \rangle} \eta(dx) \\ &= \frac{\langle \eta, p_c \rangle}{\langle \eta, g \rangle^2} \int_E \frac{g^2(x)}{p_c(x)} \eta(dx) , \end{aligned}$$

à comparer avec

$$1 + \chi^2(\mu, \eta) = \int_E \left(\frac{d\mu}{d\eta}(x) \right)^2 \eta(dx) = \frac{1}{\langle \eta, g \rangle^2} \int_E g^2(x) \eta(dx) .$$

On remarque que

$$\frac{g(x)}{\min(1, \frac{g(x)}{c})} = \max(g(x), c) \quad \text{et} \quad \min(1, \frac{g(x)}{c}) = \frac{1}{c} \min(g(x), c) ,$$

pour tout $x \in E$, de sorte que

$$\begin{aligned} \int_E \frac{g^2(x)}{p_c(x)} \eta(dx) &= \int_E \frac{g^2(x)}{\min(1, \frac{g(x)}{c})} \eta(dx) \\ &= \int_E \max(g(x), c) g(x) \eta(dx) = \mathbb{E}[\max(g(\Xi), c) g(\Xi)] , \end{aligned}$$

et

$$\begin{aligned} \langle \eta, p_c \rangle &= \int_E p_c(x) \eta(dx) = \int_E \min(1, \frac{g(x)}{c}) \eta(dx) \\ &= \frac{1}{c} \int_E \min(g(x), c) \eta(dx) = \frac{1}{c} \mathbb{E}[\min(g(\Xi), c)] , \end{aligned}$$

Compte tenu que les variables aléatoires

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

sont positivement corrélées, on en déduit que

$$\begin{aligned} \langle \eta, g \rangle^2 (1 + \chi^2(\mu, \eta_c)) &= \langle \eta, p_c \rangle \int_E \frac{g^2(x)}{p_c(x)} \eta(dx) \\ &= \frac{1}{c} \mathbb{E}[\min(g(\Xi), c)] \mathbb{E}[\max(g(\Xi), c) g(\Xi)] \\ &\leq \frac{1}{c} \mathbb{E}[\min(g(\Xi), c) \max(g(\Xi), c) g(\Xi)] = \mathbb{E}[g^2(\Xi)] \\ &= \int_E g^2(x) \eta(dx) = \langle \eta, g \rangle^2 (1 + \chi^2(\mu, \eta)) , \end{aligned}$$

de sorte que

$$\chi^2(\mu, \eta_c) \leq \chi^2(\mu, \eta) ,$$

c'est-à-dire que l'algorithme d'échantillonnage pondéré avec la distribution d'importance η_c offre une meilleure performance en terme de variance que l'algorithme d'échantillonnage pondéré avec la distribution d'importance η .

□

(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 ?**

SOLUTION

Soit à calculer l'intégrale suivante

$$I = \langle \mu, \phi \rangle = \frac{\langle \eta, g \phi \rangle}{\langle \eta, g \rangle} = \frac{\langle \eta_c, \frac{g}{p_c} \phi \rangle}{\langle \eta_c, \frac{g}{p_c} \rangle} ,$$

selon l'une ou l'autre des trois méthodes suivantes

- échantillonnage exact par acceptation / rejet pour la distribution de probabilité μ avec la décomposition d'importance originale (7), d'où la variance, normalisée par le nombre total d'itérations dans l'algorithme d'acceptation / rejet

$$\frac{M}{\langle \eta, g \rangle} \text{var}(\phi, \mu) ,$$

- échantillonnage pondéré pour la distribution de probabilité μ avec la décomposition d'importance originale (7), d'où la variance asymptotique, normalisée par la taille de l'échantillon

$$\frac{\langle \eta, g^2 |\phi - \langle \mu, \phi \rangle|^2 \rangle}{\langle \eta, g \rangle^2} \leq \frac{M}{\langle \eta, g \rangle} \text{var}(\phi, \mu) ,$$

compte tenu que $0 \leq g(x) \leq M$ pour tout $x \in E$,

- échantillonnage pondéré pour la distribution de probabilité μ avec la nouvelle décomposition d'importance introduite à la question (v), et échantillonnage exact par acceptation / rejet pour la distribution de probabilité η_c avec la décomposition d'importance (8), d'où la variance asymptotique, normalisée par le nombre total d'itérations dans l'algorithme d'acceptation / rejet

$$\frac{1}{\langle \eta, p_c \rangle} \frac{\langle \eta_c, (\frac{g}{p_c})^2 |\phi - \langle \mu, \phi \rangle|^2 \rangle}{\langle \eta_c, \frac{g}{p_c} \rangle^2} = \frac{\langle \eta, \frac{g^2}{p_c} |\phi - \langle \mu, \phi \rangle|^2 \rangle}{\langle \eta, g \rangle^2} \geq \frac{\langle \eta, g^2 |\phi - \langle \mu, \phi \rangle|^2 \rangle}{\langle \eta, g \rangle^2} ,$$

compte tenu que $0 \leq p_c(x) \leq 1$ pour tout $x \in E$.

En résumé

- au vu de la réponse à la question (ii), l'algorithme de rejet contrôlé utilise moins d'itérations dans son étape d'acceptation / rejet que l'algorithme d'acceptation / rejet, ou de manière équivalente, sa probabilité d'acceptation est plus grande,
- au vu de la réponse à la question (vi), la distribution d'importance η_c utilisée dans l'algorithme de rejet contrôlé est plus proche de la distribution cible π , au sens de la distance du χ^2 , que ne l'est la distribution d'importance η utilisée dans l'algorithme d'échantillonnage pondéré,
- en revanche, au vu des calculs effectués ci-dessus, c'est l'algorithme d'échantillonnage pondéré qui possède la plus petite variance asymptotique, normalisée par le nombre de variables aléatoires simulées, fussent-elles rejetées, c'est-à-dire convenablement normalisé par le temps de calcul.

□