École Nationale Supérieure de Techniques Avancées
module : Commande des Systèmes

examen du cours B7–1
“Filtrage bayésien et approximation particulaire”

mercredi 15 octobre 2014, 8:30 à 10:00

Problème

L'objectif de ce problème est d'étudier l'approximation par échantillonnage pondéré d'une distribution de Gibbs–Boltzman associée à une fonction positive et à une distribution de probabilité particulière, décrite par un mélange fini de distributions de probabilité.

On considère donc une distribution de probabilité

$$m(dx) = \sum_{j=1}^{M} w_j \; m_j(dx) \qquad \text{avec} \qquad \sum_{j=1}^{M} w_j = 1 \ ,$$

sous la forme d'un mélange fini de M distributions de probabilités $(m_1, \cdots, m_M)$ affectées des poids positifs normalisés $(w_1, \cdots, w_M)$, et on considère la distribution de Gibbs–Boltzman associée

$$\mu = g \cdot m = \frac{g \; m}{\langle m, g \rangle} \qquad \text{c'est–à–dire que} \qquad \mu(dx) = \frac{\displaystyle\sum_{j=1}^{M} w_j \; g(x) \; m_j(dx)}{\displaystyle\sum_{j=1}^{M} w_j \; \langle m_j, g \rangle} \ ,$$

où g est une fonction positive. On définit aussi la distribution non–normalisée

$$\gamma = g \; m \qquad \text{c'est–à–dire que} \qquad \gamma(dx) = \sum_{j=1}^{M} w_j \; g(x) \; m_j(dx) \ .$$

Au lieu de l'hypothèse faite en cours qu'il est facile de *simuler* une variable aléatoire selon chacune des composantes du mélange, on suppose ici que chacune des composantes du mélange possède une densité par rapport à une mesure positive de référence λ, et que l'expression de cette densité est connue explicitement. En d'autres termes

$$m(dx) = \sum_{j=1}^{M} w_j \; m_j(x) \; \lambda(dx) \qquad \text{avec} \qquad \sum_{j=1}^{M} w_j = 1 \; ,$$

et

$$\gamma(dx) = \sum_{j=1}^{M} w_j \; g(x) \; m_j(x) \; \lambda(dx) \; ,$$

et on suppose qu'il est facile, pour tout $j = 1, \cdots, M$

- d'*évaluer* le produit $g(x) \; m_j(x)$ pour tout $x \in E$.

L'approximation proposée repose sur la factorisation suivante, valide pour chacune des composantes du mélange : pour tout $j = 1, \cdots, M$

$$w_j \; g(x) \; m_j(x) = r_j(x) \; \theta_j \; \pi_j(x) \qquad \text{avec} \qquad r_j(x) = \frac{w_j \; g(x) \; m_j(x)}{\theta_j \; \pi_j(x)} \; ,$$

pour tout $x \in E$, où

- la densité de probabilité π_j ne s'annule qu'aux points $x \in E$ où le produit $g(x) \; m_j(x)$ est déjà nul,
- le poids positif θ_j n'est nul que si le poids w_j est déjà nul,

et on adoptera la convention $0/0 = 0$. En d'autres termes, cette factorisation permet de ré–écrire la distribution non–normalisée sous la forme équivalente

$$\gamma(dx) = \sum_{j=1}^{M} r_j(x) \; \theta_j \; \pi_j(x) \; \lambda(dx) \qquad \text{avec} \qquad \sum_{j=1}^{M} \theta_j = 1 \; ,$$

et on suppose qu'il est facile, pour tout $j = 1, \cdots, M$

- de *simuler* une variable aléatoire de densité π_j,
- et d'*évaluer* la fonction $\pi_j(x)$ pour tout $x \in E$.

Intuitivement, l'approximation proposée consiste à échantillonner selon les densités $(\pi_1, \cdots, \pi_M)$ et selon les poids $(\theta_1, \cdots, \theta_M)$ et à pondérer judicieusement les échantillons obtenus. Pour rendre rigoureuse cette approche jusqu'ici intuitive, l'idée consiste à interpréter la distribution non–normalisée γ comme la distribution marginale (définie sur E) d'une distribution non–normalisée définie sur l'ensemble produit $E' = \{1, \cdots, M\} \times E$.

Soit (I, X) la variable aléatoire à valeurs dans l'ensemble produit $E' = \{1, \cdots, M\} \times E$ et de distribution de probabilité jointe η', vue comme une famille $\eta' = (\eta_1, \cdots, \eta_M)$ de M distributions de probabilité sur l'ensemble E, définie par

$$\mathbb{P}[I = j, X \in dx] = \eta_j(dx) = \theta_j \ \pi_j(x) \ \lambda(dx) \ ,$$

pour tout $j = 1, \cdots, M$.

(i) **Décrire la distribution de probabilité marginale de la variable aléatoire I seulement, et la distribution de probabilité conditionnelle de la variable aléatoire X sachant $I = j$. En déduire un moyen pratique de simuler la variable aléatoire (I, X).**

Solution

En intégrant la distribution de probabilité jointe par rapport à la variable $x \in E$, on obtient

$$\mathbb{P}[I = j] = \int_E \eta_j(dx) = \theta_j \ \int_E \pi_j(x) \ \lambda(dx) = \theta_j \ ,$$

pour tout $j = 1, \cdots, M$, c'est–à–dire que la distribution de probabilité marginale de la variable I seulement est définie par le vecteur de probabilités $(\theta_1, \cdots, \theta_M)$.

D'après la formule de Bayes, on obtient

$$\mathbb{P}[X \in dx \mid I = j] = \frac{\mathbb{P}[I = j, X \in dx]}{\mathbb{P}[I = j]} = \frac{\theta_j \ \pi_j(x) \ \lambda(dx)}{\theta_j} = \pi_j(x) \ \lambda(dx) \ ,$$

pour tout $j = 1, \cdots, M$, c'est–à–dire que la distribution de probabilité conditionnelle de la variable aléatoire X sachant $I = j$ a pour densité π_j par rapport à la mesure positive de référence λ.

Pour simuler la variable aléatoire (I, X), il suffit de simuler d'abord la variable aléatoire I selon le vecteur de probabilités $(\theta_1, \cdots, \theta_M)$, puis de simuler la variable aléatoire X de densité π_I par rapport à la mesure positive de référence λ.

□

Sur l'ensemble produit $E' = \{1, \cdots, M\} \times E$, on introduit la distribution non–normalisée γ', vue comme une famille $\gamma' = (\gamma_1, \cdots, \gamma_M)$ de M distributions non–normalisées sur l'ensemble E, définie par

$$\gamma_j(dx) = r_j(x) \ \theta_j \ \pi_j(x) \ \lambda(dx) \ ,$$

pour tout $j = 1, \cdots, M$.

(ii) **Montrer que la distribution non–normalisée γ peut s'interpréter comme la distribution marginale (définie sur l'ensemble E) de la distribution non–normalisée $\gamma' = (\gamma_1, \cdots, \gamma_M)$ (définie sur l'ensemble produit $E' = \{1, \cdots, M\} \times E$).**

Solution

On obtient la distribution marginale en sommant par rapport à l'indice $j = 1, \cdots, M$, ce qui donne

$$\sum_{j=1}^{M} \gamma_j(dx) = \sum_{j=1}^{M} r_j(x)\ \theta_j\ \pi_j(x)\ \lambda(dx) = \gamma(dx)\ .$$

$\square$

(iii) **Montrer que la distribution non–normalisée γ' définie sur l'ensemble produit $E' = \{1, \cdots, M\} \times E$ peut s'interpréter comme le produit**

- **d'une distribution de probabilité jointe définie sur l'ensemble produit E', vue comme une famille de M distributions de probabilités définies sur l'ensemble E,**
- **et d'une fonction positive définie sur l'ensemble produit E', vue comme une famille de M fonctions positives définies sur l'ensemble E,**

dont on donnera les expressions.

Solution

En raisonnant composante par composante, on voit que

$$\gamma_j(dx) = r_j(x)\ \theta_j\ \pi_j(x)\ \lambda(dx) = r_j(x)\ \eta_j(dx)\ ,$$

pour tout $j = 1, \cdots, M$, c'est–à–dire que la distribution non–normalisée $\gamma' = (\gamma_1, \cdots, \gamma_M)$ définie sur l'ensemble produit $E' = \{1, \cdots, M\} \times E$ s'exprime comme le produit

- de la distribution de probabilité jointe η', vue comme une famille $\eta' = (\eta_1, \cdots, \eta_M)$ de M distributions de probabilité sur l'ensemble E, définie par

$$\eta_j(dx) = \theta_j\ \pi_j(x)\ \lambda(dx)\ ,$$

  pour tout $j = 1, \cdots, M$,

- et de la fonction positive r' définie sur l'ensemble produit E', vue comme une famille $r' = (r_1, \cdots, r_M)$ de M fonctions positives définies sur l'ensemble E, définie par

$$r_j(x) = \frac{w_j\ g(x)\ m_j(x)}{\theta_j\ \pi_j(x)}\ ,$$

  pour tout $j = 1, \cdots, M$.

$\square$

(iv) **En utilisant cette factorisation, proposer une approximation par échantillonnage pondéré de la distribution non–normalisée $\gamma' = (\gamma_1, \cdots, \gamma_M)$ définie sur l'ensemble produit $E' = \{1, \cdots, M\} \times E$.**

Solution

Pour obtenir une approximation par échantillonnage pondéré de la distribution non–normalisée $\gamma' = (\gamma_1, \cdots, \gamma_M)$, il suffit

- de simuler un N–échantillon $((I_1, \xi_1), \cdots, (I_N, \xi_N))$ distribué conjointement selon η', vu comme une famille $\eta' = (\eta_1, \cdots, \eta_M)$ de M distributions de probabilité sur l'ensemble E, définie par

$$\eta_j(dx) = \theta_j \; \pi_j(x) \; \lambda(dx) \; ,$$

  pour tout $j = 1, \cdots, M$, c'est–à–dire que indépendamment pour tout $i = 1, \cdots, N$

  - la variable aléatoire I_i est distribuée selon le vecteur de probabilités $(\theta_1, \cdots, \theta_M)$,
  - et la variable aléatoire ξ_i a pour densité π_{I_i} par rapport à la mesure positive de référence λ,

- et d'évaluer les poids

$$r_{I_i}(\xi_i) = \frac{w_{I_i} \; g(\xi_i) \; m_{I_i}(\xi_i)}{\theta_{I_i} \; \pi_{I_i}(\xi_i)} \; ,$$

  pour tout $i = 1, \cdots, N$.

On obtient ainsi l'approximation

$$\gamma' \approx \widehat{\gamma}'_N = \frac{1}{N} \; \sum_{i=1}^N r_{I_i}(\xi_i) \; \delta_{(I_i, \, \xi_i)} \; ,$$

sous la forme d'une distribution empirique pondérée sur l'espace produit $E' = \{1, \cdots, M\} \times E$. Concrètement, pour toute fonction mesurable bornée ϕ' définie sur l'ensemble produit E', vue comme une famille $\phi' = (\phi_1, \cdots, \phi_M)$ de M fonctions mesurables bornées définies sur l'ensemble E, on a

$$\langle \gamma', \phi' \rangle = \sum_{j=1}^M \langle \gamma_j, \phi_j \rangle \approx \langle \widehat{\gamma}'_N, \phi' \rangle = \frac{1}{N} \; \sum_{i=1}^N r_{I_i}(\xi_i) \; \phi_{I_i}(\xi_i) \; ,$$

□

(v) **En utilisant le résultat énoncé à la question (ii), en déduire une approximation par échantillonnage pondéré de la distribution non–normalisée γ définie sur l'ensemble E.**

Solution

En particulier, si la fonction ϕ' dépend seulement de la variable $x \in E$, c'est–à–dire si $\phi_j = \phi$ pour tout $j = 1, \cdots, M$, alors

$$\langle \gamma, \phi \rangle = \langle \sum_{j=1}^{M} \gamma_j, \phi \rangle = \sum_{j=1}^{M} \langle \gamma_j, \phi \rangle \approx \frac{1}{N} \sum_{i=1}^{N} r_{I_i}(\xi_i)\, \phi(\xi_i) \ ,$$

d'où l'approximation

$$\gamma \approx \widehat{\gamma}_N = \frac{1}{N} \sum_{i=1}^{N} r_{I_i}(\xi_i)\, \delta_{\xi_i} \ .$$

□

(vi) **En déduire une approximation par échantillonnage pondéré de la distribution de Gibbs–Boltzmann μ définie sur l'ensemble E.**

Solution

En particulier pour $\phi \equiv 1$, on obtient l'approximation suivante

$$\langle \gamma, 1 \rangle \approx \langle \widehat{\gamma}_N, 1 \rangle = \frac{1}{N} \sum_{i=1}^{N} r_{I_i}(\xi_i) \ ,$$

pour la constante de normalisation. En suivant le principe vu en cours pour l'approximation par échantillonnage pondéré d'une distribution de Gibbs–Boltzmann, on utilise le même échantillon pour approcher numérateur et dénominateur, et on obtient l'approximation suivante

$$\mu = \frac{\gamma}{\langle \gamma, 1 \rangle} \approx \frac{\widehat{\gamma}_N}{\langle \widehat{\gamma}_N, 1 \rangle} = \frac{\displaystyle\sum_{i=1}^{N} r_{I_i}(\xi_i)\, \delta_{\xi_i}}{\displaystyle\sum_{j=1}^{N} r_{I_j}(\xi_j)} = \sum_{i=1}^{N} w_i\, \delta_{\xi_i} \ ,$$

avec les nouveaux poids définis pour tout $i = 1, \cdots, N$ par

$$w_i = \frac{r_{I_i}(\xi_i)}{\displaystyle\sum_{j=1}^{N} r_{I_j}(\xi_j)} \ .$$

□

On considère l'approximation

$$\gamma \approx \widehat{\gamma}_N = \frac{1}{N} \sum_{i=1}^{N} r_{I_i}(\xi_i)\, \delta_{\xi_i} \ ,$$

obtenue en réponse à la question (v), où indépendemment pour tout $i = 1, \cdots, N$ la variable aléatoire (I_i, ξ_i) à pour distribution de probabilité jointe $\eta' = (\eta_1, \cdots, \eta_M)$ définie sur l'ensemble produit $E' = \{1, \cdots, M\} \times E$. Concrètement

$$\langle \gamma, \phi \rangle \approx \langle \widehat{\gamma}_N, \phi \rangle = \frac{1}{N} \sum_{i=1}^{N} r_{I_i}(\xi_i)\ \phi(\xi_i)\ ,$$

pour toute fonction mesurable bornée ϕ définie sur l'ensemble E.

(vii) **Montrer que $\langle \widehat{\gamma}_N, \phi \rangle$ est un estimateur non–biaisé de $\langle \gamma, \phi \rangle$.**

Solution

Par définition

$$\begin{aligned}
\mathbb{E}\,\langle \widehat{\gamma}_N, \phi \rangle &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[r_{I_i}(X_i)\ \phi(X_i)] \\
&= \mathbb{E}[r_I(X)\ \phi(X)] \\
&= \sum_{j=1}^{M} \int_E r_j(x)\ \phi(x)\ \theta_j\ \pi_j(x)\ \lambda(dx) \\
&= \int_E \phi(x)\ [\sum_{j=1}^{M} r_j(x)\ \theta_j\ \pi_j(x)\ \lambda(dx)\,] \\
&= \int_E \phi(x)\ \gamma(dx) = \langle \gamma, \phi \rangle\ .
\end{aligned}$$

□

(viii) **Calculer la variance $\mathbb{E}|\,\langle \widehat{\gamma}_N, \phi \rangle - \langle \gamma, \phi \rangle\,|^2$ de l'erreur d'approximation.**

Solution

Par différence

$$\begin{aligned}
\langle \widehat{\gamma}_N, \phi \rangle - \langle \gamma, \phi \rangle &= \frac{1}{N} \sum_{i=1}^{N} r_{I_i}(X_i)\ \phi(X_i) - \langle \gamma, \phi \rangle \\
&= \frac{1}{N} \sum_{i=1}^{N} [r_{I_i}(X_i)\ \phi(X_i) - \langle \gamma, \phi \rangle\,]\ ,
\end{aligned}$$

et par indépendance

$$\begin{aligned}
\mathbb{E}|\,\langle \widehat{\gamma}_N, \phi \rangle - \langle \gamma, \phi \rangle\,|^2 \;&=\; \frac{1}{N^2}\;\sum_{i=1}^N \mathbb{E}|\, r_{I_i}(X_i)\;\phi(X_i) - \langle \gamma, \phi \rangle\,|^2 \\
&= \frac{1}{N}\;\mathbb{E}|\, r_I(X)\;\phi(X) - \langle \gamma, \phi \rangle\,|^2 \\
&= \frac{1}{N}\;[\,\mathbb{E}|\, r_I(X)\;\phi(X)\,|^2 - |\langle \gamma, \phi \rangle|^2\,]\;.
\end{aligned}$$

Par définition

$$\begin{aligned}
\mathbb{E}|\, r_I(X)\;\phi(X)\,|^2 \;&=\; \sum_{j=1}^M \int_E |r_j(x)\;\phi(x)|^2\;\theta_j\;\pi_j(x)\;\lambda(dx) \\
&= \sum_{j=1}^M \int_E |\phi(x)|^2\;(\frac{w_j\;g(x)\;m_j(x)}{\theta_j\;\pi_j(x)})^2\;\theta_j\;\pi_j(x)\;\lambda(dx) \\
&= \sum_{j=1}^M \int_E |\phi(x)|^2\;g^2(x)\;\frac{w_j\;m_j(x)}{\theta_j\;\pi_j(x)}\;w_j\;m_j(x)\;\lambda(dx) \\
&= \sum_{j=1}^M \int_E |\phi(x)|^2\;g^2(x)\;(\frac{w_j\;m_j(x)}{\theta_j\;\pi_j(x)} - 1)\;w_j\;m_j(x)\;\lambda(dx) \\
&\qquad + \sum_{j=1}^M \int_E |\phi(x)|^2\;g^2(x)\;w_j\;m_j(x)\;\lambda(dx)\;,
\end{aligned}$$

et on vérifie que

$$\begin{aligned}
\sum_{j=1}^M \int_E |\phi(x)|^2\;g^2(x)\;w_j\;m_j(x)\;\lambda(dx) \;&=\; \int_E |\phi(x)|^2\;g^2(x)\;[\,\sum_{j=1}^M w_j\;m_j(x)\;\lambda(dx)\,] \\
&= \int_E |\phi(x)|^2\;g^2(x)\;m(dx) = \langle m, g^2\,|\phi|^2\,\rangle\;,
\end{aligned}$$

et que

$$\langle m, g^2\,|\phi|^2\,\rangle - |\langle \gamma, \phi \rangle|^2 = \langle m, g^2\,|\phi|^2\,\rangle - |\langle m, g\,\phi \rangle|^2 = \mathrm{var}(g\,\phi, m)\;.$$

On en déduit que

$$\begin{aligned}
N\;\mathbb{E}|\,\langle \widehat{\gamma}_N, \phi \rangle - \langle \gamma, \phi \rangle\,|^2 &= \mathrm{var}(g\,\phi, m) \\
&+ \sum_{j=1}^M \int_E |\phi(x)|^2\;g^2(x)\;(\frac{w_j\;m_j(x)}{\theta_j\;\pi_j(x)} - 1)\;w_j\;m_j(x)\;\lambda(dx)\;.
\end{aligned}$$

$\square$

Commentaire

Cette stratégie s'applique notamment dans le cadre du filtrage bayésien, et produit une classe d'approximations particulaires connues sous le sigle APF, pour *auxiliary particle filter*. Concrètement, si on dispose d'une approximation particulaire du filtre bayésien sous la forme

$$\mu_{k-1}^N = \sum_{i=1}^N w_{k-1}^i \, \delta_{\xi_{k-1}^i} \qquad \text{avec} \qquad \sum_{i=1}^N w_{k-1}^j = 1 \ ,$$

alors on obtient à l'instant suivant une approximation particulaire du filtre bayésien sous la forme

$$g_k(x') \; (\mu_{k-1}^N \, Q_k)(dx') = \sum_{i=1}^N w_{k-1}^i \; g_k(x') \; Q_k(\xi_{k-1}^i, dx') \qquad \text{avec} \qquad \sum_{i=1}^N w_{k-1}^j = 1 \ ,$$

à une constante multiplicative près. On remarque que cette distribution non–normalisée est bien de la forme étudiée ici, c'est–à–dire

$$\gamma(dx') = \sum_{i=1}^N w_i \; g(x') \; m_i(dx') \ .$$

avec les identifications

- $g(x') = g_k(x')$ pour tout $x' \in E$,
- $w_i = w_{k-1}^i$ et $m_i(dx') = Q_k(\xi_{k-1}^i, dx')$ pour tout $i = 1, \cdots, N$.

L'approximation particulaire proposée par le filtre APF repose sur la factorisation suivante : pour tout $i = 1, \cdots, N$

$$w_{k-1}^i \; g_k(x') \; m_k^i(x') = r_k^i(x') \; \theta_k^i \; \pi_k^i(x') \qquad \text{avec} \qquad r_k^i(x') = \frac{w_{k-1}^i \; g_k(x') \; m_k^i(x')}{\theta_k^i \; \pi_k^i(x')} \ ,$$

pour tout $x' \in E$, où on s'autorise à définir le nouveau poids $\theta_k^i = \theta(\xi_{k-1}^i)$ comme fonction de la position ξ_{k-1}^i de la particule à l'instant précédent. Cette flexibilité offerte permet par exemple de définir

$$\theta_k^i = \int_E g_k(x') \; m_k^i(dx') \ ,$$

si cette expression est disponible explicitement, ou sinon

$$\theta_k^i = g_k(\int_E x' \; m_k^i(dx')) \ ,$$

qui est plus souvent disponible, ce qui permet de sélectionner les particules dans la population courante dont on sait prédire que les descendants seront consistants avec l'observation recueillie à l'instant suivant.