# École Nationale Supérieure de Techniques Avancées
# examen du cours SOD333
# “Filtrage bayésien et approximation particulaire”

## vendredi 22 octobre 2021, 13:30 à 16:00

## — correction —

On rappelle l'équation récurrente

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

vérifiée par le filtre bayésien, avec

$$\eta_k(dx') = \mu_{k-1}\, Q_k(dx') = \int_E \mu_{k-1}(dx)\, Q_k(x, dx') \ .$$

Si on remplace le filtre $\mu_{k-1}$ par son approximation particulaire

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

alors on obtient le mélange fini

$$\mu_{k-1}^N\, Q_k(dx') = \sum_{i=1}^N w_{k-1}^i \; Q_k(\xi_{k-1}^i, dx') \ ,$$

comme approximation du prédicteur $\eta_k$, et la distribution de probabilité

$$g_k(x') \; [\sum_{i=1}^N w_{k-1}^i \; Q_k(\xi_{k-1}^i, dx')\, ] \ ,$$

définie à une constante de normalisation près, comme approximation du filtre $\mu_k$.

Désormais, on considère la distribution de probabilité cible $\mu(dx')$ sur $E$ définie par la factorisation, ou décomposition d'importance

$$\mu(dx') \propto \underbrace{g_k(x')}_{g(x')} \; \underbrace{\sum_{i=1}^N w_{k-1}^i \; Q_k(\xi_{k-1}^i, dx')}_{\eta(dx')} \ , \qquad (\star)$$

qui permet d'interpréter la distribution de probabilité cible $\mu(dx')$ comme une distribution de Gibbs–Boltzmann, et l'objectif de ce problème est d'étudier différentes approximations particulaires pour la distribution de probabilité cible $\mu(dx')$ et pour la constante de normalisation $\langle \eta, g \rangle$.

(i) **Décrire l'algorithme d'approximation par échantillonnage pondéré de la distribution de probabilité cible $\mu(dx')$, basé sur la décomposition d'importance $(\star)$.**

**Comment les particules sont–elles générées ? Comment les poids sont–ils calculés ? Comment la constante de normalisation $\langle \eta, g \rangle$ est–elle approximée ?**

Solution

On définit successivement l'approximation Monte Carlo du prédicteur

$$\eta^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_k^i} \qquad \text{de sorte que} \qquad \langle \eta^N, g \rangle = \frac{1}{N} \sum_{i=1}^N g(\xi_k^i) \ ,$$

puis l'approximation par échantillonnage pondéré du filtre

$$\mu^N = g \cdot \eta^N = \sum_{i=1}^N w_k^i \, \delta_{\xi_k^i} \ ,$$

où indépendamment pour tout $i = 1, \cdots, N$ la $i$–ème particule $\xi_k^i$ est générée d'après $\eta(dx')$ et reçoit un poids d'après $g(x')$, i.e.

- l'indice $a_k^i$ est généré suivant le vecteur de probabilité $(w_{k-1}^1, \cdots, w_{k-1}^N)$,

puis, en supposant que $a_k^i = a$

- la particule $\xi_k^i$ est générée suivant $Q_k(\xi_{k-1}^a, dx')$,
- et reçoit un poids $w_k^i \propto g(\xi_k^i)$, i.e.

$$w_k^i \propto g_k(\xi_k^i) \ .$$

$\square$

(ii) **Quelle est la variance (non–asymptotique) de l'erreur d'estimation de la constante de normalisation $\langle \eta, g \rangle$ par son approximation particulaire ?**

Solution

On a vu à la question (i) que l'approximation particulaire de la constante de normalisation $\langle \eta, g \rangle$ est définie par

$$\langle \eta^N, g \rangle = \frac{1}{N} \sum_{i=1}^N g(\xi_k^i) \ ,$$

où les variables aléatoires $\xi_k^i$ pour $i = 1, \cdots, N$ sont i.i.d. avec $\eta(dx')$ comme distribution de probabilité commune, de sorte que

$$\mathbb{E}[\langle \eta^N, g \rangle] = \langle \eta, g \rangle \ ,$$

c'est–à–dire que l'estimation est sans biais, et

$$N \, \mathbb{E}|\langle \eta^N, g \rangle - \langle \eta, g \rangle|^2 = \mathrm{var}(g, \eta) = \langle \eta, g^2 \rangle - \langle \eta, g \rangle^2 \ .$$

□

On propose d'étudier de nouvelles approximations particulaires pour la distribution de probabilité cible $\mu(dx')$ et pour la constante de normalisation $\langle \eta, g \rangle$, en appliquant le paradigme suivant :

> Si on peut interpréter la distribution de probabilité cible comme la distribution de probabilité marginale sur $E$ d'une distribution de probabilité jointe définie sur un espace produit, par exemple $\{1, \cdots, N\} \times E$ ou $\{1, \cdots, N\} \times E \times E$, alors on peut obtenir une approximation particulaire de la distribution de probabilité cible de la manière suivante : on construit par échantillonnage pondéré une approximation particulaire de la distribution de probabilité jointe, puis on marginalise sur $E$ cette approximation particulaire.

## Notations

De même qu'on utilise indifféremment la notation

$$\mu, \ \eta, \ \text{etc.} \qquad \text{[respectivement } g_k, \ g, \ \text{etc.]}$$

ou la notation

$$\mu(dx'), \ \eta(dx'), \ \text{etc.} \qquad \text{[respectivement } g_k(x'), \ g(x'), \ \text{etc.]}$$

pour désigner des distributions de probabilité [respectivement des fonctions] définies sur $E$, de même on utilisera indifféremment la notation

$$\mu_{\mathrm{aux}}, \ \eta_{\mathrm{aux}}, \ \text{etc.} \qquad \text{[respectivement } g_{\mathrm{aux}}, \ \text{etc.]}$$

ou la notation

$$\mu_{\mathrm{aux}}^a(dx'), \ \eta_{\mathrm{aux}}^a(dx'), \ \text{etc.} \qquad \text{[respectivement } g_{\mathrm{aux}}^a(x'), \ \text{etc.]}$$

pour désigner des distributions de probabilité [respectivement des fonctions] définies sur l'espace produit $\{1, \cdots, N\} \times E$, et de même on utilisera indifféremment la notation

$$\mu_{\rm new},\ \eta_{\rm new},\ \text{etc.} \qquad [\text{respectivement } g_{\rm new},\ \text{etc.}]$$

ou la notation

$$\mu^a_{\rm new}(dz,dx'),\ \eta^a_{\rm new}(dz,dx'),\ \text{etc.} \qquad [\text{respectivement } g^a_{\rm new}(z,x'),\ \text{etc.}]$$

pour désigner des distributions de probabilité [respectivement des fonctions] définies sur l'espace produit $\{1,\cdots,N\} \times E \times E$.

---

### Filtrage particulaire auxiliaire

Le point de départ du filtrage particulaire auxiliaire consiste à interpréter la distribution de probabilité cible $\mu(dx')$ comme la distribution de probabilité marginale sur $E$ associée à la distribution de probabilité jointe

$$\mu^a_{\rm aux}(dx') \propto g_k(x')\ w^a_{k-1}\ Q_k(\xi^a_{k-1},dx')\ ,$$

définie sur l'espace produit $\{1,\cdots,N\} \times E$.

(iii) **Vérifier que la distribution de probabilité marginale définie sur $E$ à partir de la distribution de probabilité jointe $\mu^a_{\rm aux}(dx')$ définie sur l'espace produit $\{1,\cdots,N\} \times E$, coïncide avec la distribution de probabilité cible $\mu(dx')$.**

Solution

En sommant par rapport à l'indice $a \in \{1,\cdots,N\}$, on obtient (à une constante de normalisation près) la distribution de probabilité marginale

$$\sum_{a=1}^N [\, g_k(x')\ w^a_{k-1}\ Q_k(\xi^a_{k-1},dx')\,] = g_k(x') \sum_{a=1}^N w^a_{k-1}\ Q_k(\xi^a_{k-1},dx') = g(x')\ \eta(dx')\ .$$

□

On introduit les poids auxiliaires $(\lambda^1_{\rm aux},\cdots,\lambda^N_{\rm aux})$, qui forment un vecteur de probabilité de dimension $N$, ce qui permet de définir la factorisation, ou décomposition d'importance

$$\mu^a_{\rm aux}(dx') \propto \underbrace{g_k(x')\ \frac{w^a_{k-1}}{\lambda^a_{\rm aux}}}_{g^a_{\rm aux}(x')}\ \underbrace{\lambda^a_{\rm aux}\ Q_k(\xi^a_{k-1},dx')}_{\eta^a_{\rm aux}(dx')}\ , \qquad (\star\star)$$

et d'interpréter la distribution de probabilité $\mu^a_{\rm aux}(dx')$ comme une distribution de Gibbs–Boltzmann.

(iv) **Montrer que**

$$\langle \eta_{\text{aux}}, g_{\text{aux}} \rangle = \langle \eta, g \rangle \ ,$$

**c'est–à–dire que la constante de normalisation $\langle \eta_{\text{aux}}, g_{\text{aux}} \rangle$ associée à la décomposition d'importance ($\star\star$) coïncide avec la constante de normalisation $\langle \eta, g \rangle$ associée à la décomposition d'importance ($\star$).**

Solution

Par définition

$$\begin{aligned} \langle \eta_{\text{aux}}, g_{\text{aux}} \rangle &= \sum_{a=1}^N \int_E g^a_{\text{aux}}(x') \ \eta^a_{\text{aux}}(dx') \\ &= \sum_{a=1}^N \int_E [g_k(x') \ w^a_{k-1} \ Q_k(\xi^a_{k-1}, dx')\,] \\ &= \int_E g_k(x') \ \sum_{a=1}^N w^a_{k-1} \ Q_k(\xi^a_{k-1}, dx') \\ &= \langle \eta, g \rangle \ . \end{aligned}$$

$\square$

(v) **Décrire l'algorithme d'approximation par échantillonnage pondéré de la distribution de probabilité $\mu^a_{\text{aux}}(dx')$, basé sur la décomposition d'importance ($\star\star$).**

**Comment les particules sont–elles générées ? Comment les poids sont–ils calculés ? Comment la constante de normalisation $\langle \eta_{\text{aux}}, g_{\text{aux}} \rangle$ est–elle approximée ? Comment la distribution de probabilité cible $\mu(dx')$ est–elle approximée, après marginalisation ?**

Solution

On définit successivement l'approximation Monte Carlo

$$\eta^N_{\text{aux}} = \frac{1}{N} \ \sum_{i=1}^N \delta_{(a^i_k,\, \xi^i_k)} \qquad \text{de sorte que} \qquad \langle \eta^N_{\text{aux}}, g_{\text{aux}} \rangle = \frac{1}{N} \ \sum_{i=1}^N g^{a^i_k}_{\text{aux}}(\xi^i_k) \ ,$$

puis l'approximation par échantillonnage pondéré

$$\mu^N_{\text{aux}} = \sum_{i=1}^N w^i_k \ \delta_{(a^i_k,\, \xi^i_k)} \qquad \text{et sa marginale} \qquad \mu^N = \sum_{i=1}^N w^i_k \ \delta_{\xi^i_k} \ ,$$

où indépendamment pour tout $i = 1, \cdots, N$ la $i$–ème particule $(a^i_k, \xi^i_k)$ est générée d'après $\eta^a_{\text{aux}}(dx')$ et reçoit un poids donné par $g^a_{\text{aux}}(x')$, i.e.

- l'indice $a_k^i$ est généré suivant le vecteur de probabilité $(\lambda_{\rm aux}^1, \cdots, \lambda_{\rm aux}^N)$,

puis, en supposant que $a_k^i = a$

- la particule $\xi_k^i$ est générée suivant la distribution de probabilité $Q_k(\xi_{k-1}^a, dx')$,
- et reçoit le poids $w_k^i \propto g_{\rm aux}^a(\xi_k^i)$, i.e.

$$w_k^i \propto g_k(\xi_k^i) \; \frac{w_{k-1}^a}{\lambda_{\rm aux}^a} \; .$$

$\square$

Pour chaque vecteur de probabilité $(\lambda_{\rm aux}^1, \cdots, \lambda_{\rm aux}^N)$, on obtient une décomposition d'importance de la distribution de probabilité $\mu_{\rm aux}^a(dx')$ et une approximation particulaire par échantillonnage pondéré.

(vi) **Montrer que la classe d'approximations particulaires décrite à la question (v) contient l'approximation particulaire décrite à la question (i) comme cas particulier.**

[Indice : Pour un choix particulier du vecteur de probabilité $(\lambda_{\rm aux}^1, \cdots, \lambda_{\rm aux}^N)$, l'approximation particulaire décrite à la question (v) coïncide avec l'approximation particulaire décrite à la question (i).]

Solution

Si $\lambda_{\rm aux}^a = w_{k-1}^a$ pour tout $a \in \{1, \cdots, N\}$ alors l'algorithme d'approximation particulaire décrit à la question (v) se réduit à l'algorithme d'approximation particulaire décrit à la question (i).

On en déduit le résultat de comparaison suivant. Si on minimise par rapport au vecteur de probabilité $(\lambda_{\rm aux}^1, \cdots, \lambda_{\rm aux}^N)$ l'expression de la variance de l'erreur d'estimation de la constante de normalisation $\langle \eta_{\rm aux}, g_{\rm aux} \rangle$ par son approximation particulaire décrite à la question (v), alors la variance minimale est plus petite que la variance obtenue pour un choix quelconque du vecteur de probabilité $(\lambda_{\rm aux}^1, \cdots, \lambda_{\rm aux}^N)$, et en particulier elle est plus petite que la variance correspondant au choix $(w_{k-1}^1, \cdots, w_{k-1}^N)$ c'est–à–dire plus petite que la variance de l'erreur d'estimation de la constante de normalisation $\langle \eta, g \rangle$ par son approximation particulaire décrite à la question (i).

$\square$

(vii) **Montrer que la variance (non–asymptotique) de l'erreur d'estimation de la constante de normalisation $\langle \eta_{\rm aux}, g_{\rm aux} \rangle$ par son approximation particulaire est égale à**

$$\frac{1}{N} \; [ \sum_{a=1}^N \frac{|w_{k-1}^a|^2}{\lambda_{\rm aux}^a} \; |u_a|^2 - \langle \eta, g \rangle^2 \; ] \; ,$$

**avec**

$$u_a = \{ \int_E |g_k(x')|^2 \; Q_k(\xi^a_{k-1}, dx') \}^{1/2} \; ,$$

**pour tout** $a = 1, \cdots, N$**.**

Solution

On a vu à la question (v) que l'approximation particulaire de la constante de normalisation $\langle \eta_{\text{aux}}, g_{\text{aux}} \rangle$ est définie par

$$\langle \eta^N_{\text{aux}}, g_{\text{aux}} \rangle = \frac{1}{N} \; \sum_{i=1}^N g^{a^i_k}_{\text{aux}}(\xi^i_k) \; ,$$

où les variables aléatoires $(a^i_k, \xi^i_k)$ pour $i = 1, \cdots, N$ sont i.i.d. avec $\eta^a_{\text{aux}}(dx')$ comme distribution de probabilité commune, de sorte que

$$\mathbb{E}[\langle \eta^N_{\text{aux}}, g_{\text{aux}} \rangle] = \langle \eta_{\text{aux}}, g_{\text{aux}} \rangle = \langle \eta, g \rangle \; ,$$

c'est–à–dire que l'estimation est sans biais, et

$$N \; \mathbb{E}|\langle \eta^N_{\text{aux}}, g_{\text{aux}} \rangle - \langle \eta_{\text{aux}}, g_{\text{aux}} \rangle|^2 = \text{var}(g_{\text{aux}}, \eta_{\text{aux}}) = \langle \eta_{\text{aux}}, g^2_{\text{aux}} \rangle - \langle \eta_{\text{aux}}, g_{\text{aux}} \rangle^2 \; .$$

En fait, compte tenu de l'identité $\langle \eta_{\text{aux}}, g_{\text{aux}} \rangle = \langle \eta, g \rangle$ montrée à la question (iv), la variable aléatoire $\langle \eta^N_{\text{aux}}, g_{\text{aux}} \rangle$ fournit aussi une estimation sans biais de la constante de normalisation $\langle \eta, g \rangle$, et la variance de l'erreur d'estimation vérifie

$$N \; \mathbb{E}|\langle \eta^N_{\text{aux}}, g_{\text{aux}} \rangle - \langle \eta, g \rangle|^2 = \langle \eta_{\text{aux}}, g^2_{\text{aux}} \rangle - \langle \eta, g \rangle^2 \; .$$

La variance de l'erreur d'estimation est controlée par

$$\begin{aligned}
\langle \eta_{\text{aux}}, g^2_{\text{aux}} \rangle &= \sum_{a=1}^N \int_E |g^a_{\text{aux}}(x')|^2 \; \eta^a_{\text{aux}}(dx') \\
&= \sum_{a=1}^N \int_E |g_k(x') \; \frac{w^a_{k-1}}{\lambda^a_{\text{aux}}}|^2 \; \lambda^a_{\text{aux}} \; Q_k(\xi^a_{k-1}, dx') \\
&= \sum_{a=1}^N \frac{|w^a_{k-1}|^2}{\lambda^a_{\text{aux}}} \; \int_E |g_k(x') \; Q_k(\xi^a_{k-1}, dx') \\
&= \sum_{a=1}^N \frac{|w^a_{k-1}|^2}{\lambda^a_{\text{aux}}} \; |u_a|^2 \; ,
\end{aligned}$$

avec

$$u_a = \{ \int_E |g_k(x')|^2 \; Q_k(\xi^a_{k-1}, dx') \}^{1/2} \; ,$$

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

$\square$

(viii) **Pour quel choix particulier du vecteur de probabilité $(\lambda^1_{\rm aux}, \cdots, \lambda^N_{\rm aux})$, la variance de l'erreur d'estimation étudiée à la question (vii) est–elle minimale ? Monter que cette variance minimale est égale à**

$$\frac{1}{N}\,[\,[\sum_{a=1}^N w^a_{k-1}\,u_a\,]^2 - \langle \eta, g \rangle^2\,]\ .$$

**Comment cette variance se compare–t–elle à la variance obtenue à la question (ii) ?**

Solution

Pour minimiser l'expression de la variance par rapport au vecteur $(\lambda^1_{\rm aux}, \cdots, \lambda^N_{\rm aux})$, sous la contrainte qu'il s'agisse d'un vecteur de probabilité, on introduit le multiplicateur de Lagrange $\mu$ et le lagrangien

$$\sum_{a=1}^N \frac{|w^a_{k-1}|^2}{\lambda^a_{\rm aux}}\,|u_a|^2 + \mu\,(\sum_{a=1}^N \lambda^a_{\rm aux} - 1)\ .$$

La condition d'optimalité du premier ordre donne

$$|\frac{w^a_{k-1}}{\lambda^a_{\rm aux}}|^2\,|u_a|^2 = \mu\ ,$$

de sorte que

$$\lambda^a_{\rm opt} \propto w^a_{k-1}\,u_a\ .$$

Si on utilise cette expression, on obtient

$$\begin{aligned}
\min_{\lambda_{\rm aux}} \langle \eta_{\rm aux}, g^2_{\rm aux} \rangle &= \min_{\lambda_{\rm aux}} \sum_{a=1}^N \frac{|w^a_{k-1}\,u_a|^2}{\lambda^a_{\rm aux}} \\
&= \sum_{a=1}^N \frac{|w^a_{k-1}\,u_a|^2}{\lambda^a_{\rm opt}} \\
&= \sum_{a=1}^N \frac{|w^a_{k-1}\,u_a|^2}{w^a_{k-1}\,u_a} \times \sum_{a=1}^N w^a_{k-1}\,u_a \\
&= [\sum_{a=1}^N w^a_{k-1}\,u_a\,]^2 \\
&\leq \sum_{a=1}^N w^a_{k-1}\,|u_a|^2\ ,
\end{aligned}$$

où la majoration découle de l'inégalité de Cauchy–Schwartz. On remarque que

$$\sum_{a=1}^{N} w_{k-1}^a \, |u_a|^2 = \sum_{a=1}^{N} w_{k-1}^a \int_E |g_k(x')|^2 \, Q_k(\xi_{k-1}^a, dx') = \langle \eta, g^2 \rangle \; .$$

On en déduit (ou on retrouve ce qui était attendu) que

$$\min_{\lambda_{\mathrm{aux}}} \langle \eta_{\mathrm{aux}}, g_{\mathrm{aux}}^2 \rangle = [\sum_{a=1}^{N} w_{k-1}^a \, u_a \,]^2 \leq \sum_{a=1}^{N} w_{k-1}^a \, |u_a|^2 = \langle \eta, g^2 \rangle \; .$$

c'est–à–dire que le filtre particulaire auxiliaire avec le choix optimal du vecteur de probabilité auxiliaire est plus performant (possède une variance plus petite) que le filtre particulaire classique.

$\square$

### Filtrage particulaire avec transition markovienne auxiliaire

Le point de départ de ce nouveau concept de filtrage particulaire auxiliaire consiste à introduire un noyau de probabilités de transition auxiliaire de l'ensemble fini $\{1, \cdots, N\}$ vers $E$, c'est–à–dire un vecteur $(\pi_{\mathrm{new}}^1(dz), \cdots, \pi_{\mathrm{new}}^N(dz))$ de distributions de probabilité sur $E$, et à interpréter la distribution de probabilité cible $\mu(dx')$ comme la distribution de probabilité marginale sur $E$ associée à la distribution de probabilité jointe

$$\mu_{\mathrm{new}}^a(dz, dx') \propto g_k(x') \; w_{k-1}^a \; \pi_{\mathrm{new}}^a(dz) \; Q_k(\xi_{k-1}^a, dx') \; ,$$

définie sur l'espace produit $\{1, \cdots, N\} \times E \times E$.

(ix) **Vérifier que la distribution de probabilité marginale définie sur $E$ à partir de la distribution de probabilité jointe $\mu_{\mathrm{new}}^a(dz, dx')$ définie sur l'espace produit $\{1, \cdots, N\} \times E \times E$, coïncide avec la distribution de probabilité cible $\mu(dx')$.**

#### Solution

En sommant par rapport à l'indice $a \in \{1, \cdots, N\}$ et en intégrant par rapport à la variable $z \in E$, on obtient (à une constante de normalisation près) la distribution de probabilité marginale

$$
\begin{aligned}
\sum_{a=1}^N \int_E [\, g_k(x') \; w_{k-1}^a \; \pi_{\rm new}^a(dz) \; Q_k(\xi_{k-1}^a, dx') \,] \\
&= g_k(x') \; \sum_{a=1}^N w_{k-1}^a \; [\int_E \pi_{\rm new}^a(dz)] \; Q_k(\xi_{k-1}^a, dx') \\
&= g_k(x') \; \sum_{a=1}^N w_{k-1}^a \; Q_k(\xi_{k-1}^a, dx') \\
&= g(x') \; \eta(dx') \; .
\end{aligned}
$$

$\square$

On suppose désormais que les noyaux de probabilités de transition $Q_k(x, dx')$ possèdent une densité, i.e. on suppose que

$$
Q_k(x, dx') = q_k(x, x') \; dx \; ,
$$

ce qui permet de définir la factorisation, ou décomposition d'importance

$$
\mu_{\rm new}^a(dz, dx') \propto \underbrace{g_k(x') \; \frac{q_k(\xi_{k-1}^a, x')}{q_k(z, x')} \; \frac{w_{k-1}^a}{\lambda_{\rm new}^a}}_{g_{\rm new}^a(z, x')} \; \underbrace{\lambda_{\rm new}^a \; \pi_{\rm new}^a(dz) \; Q_k(z, dx')}_{\eta_{\rm new}^a(dz, dx')} \; , \qquad (\star\star\star)
$$

et d'interpréter la distribution de probabilité $\mu_{\rm new}^a(dz, dx')$ comme une distribution de Gibbs–Boltzmann.

(x) **Montrer que**

$$
\langle \eta_{\rm new}, g_{\rm new} \rangle = \langle \eta, g \rangle \; ,
$$

**c'est–à–dire que la constante de normalisation $\langle \eta_{\rm new}, g_{\rm new} \rangle$ associée à la décomposition d'importance $(\star\star\star)$ coïncide avec la constante de normalisation $\langle \eta, g \rangle$ associée à la décomposition d'importance $(\star)$.**

Solution

Par définition

$$\begin{aligned}
\langle \eta_{\rm new}, g_{\rm new} \rangle &= \sum_{a=1}^{N} \int_E \int_E g^a_{\rm new}(z, x') \; \eta^a_{\rm new}(dz, dx') \\
&= \sum_{a=1}^{N} \int_E \int_E [g_k(x') \; w^a_{k-1} \; \pi^a_{\rm new}(dz) \; Q_k(\xi^a_{k-1}, dx') \,] \\
&= \int_E g_k(x') \; \sum_{a=1}^{N} w^a_{k-1} \; [\int_E \pi^a_{\rm new}(dz)] \; Q_k(\xi^a_{k-1}, dx') \\
&= \int_E g_k(x') \; \sum_{a=1}^{N} w^a_{k-1} \; Q_k(\xi^a_{k-1}, dx') \\
&= \langle \eta, g \rangle \; .
\end{aligned}$$

□

(xi) **Décrire l'algorithme d'approximation par échantillonnage pondéré de la distribution de probabilité $\mu^a_{\rm new}(dz, dx')$, basé sur la décomposition d'importance ($\star\star\star$).**

**Comment les particules sont–elles générées ? Comment les poids sont–ils calculés ? Comment la constante de normalisation $\langle \eta_{\rm new}, g_{\rm new} \rangle$ est–elle approximée ? Comment la distribution de probabilité cible $\mu(dx')$ est–elle approximée, après marginalisation ?**

SOLUTION

On définit successivement l'approximation Monte Carlo

$$\eta^N_{\rm new} = \frac{1}{N} \sum_{i=1}^{N} \delta_{(a^i_k, \, \zeta^i_k, \, \xi^i_k)} \qquad \text{de sorte que} \qquad \langle \eta^N_{\rm new}, g_{\rm new} \rangle = \frac{1}{N} \sum_{i=1}^{N} g^{a^i_k}_{\rm new}(\zeta^i_k, \xi^i_k) \; ,$$

puis l'approximation par échantillonnage pondéré

$$\mu^N_{\rm new} = g_{\rm new} \cdot \eta^N_{\rm new} = \sum_{i=1}^{N} w^i_k \; \delta_{(a^i_k, \, \zeta^i_k, \, \xi^i_k)} \qquad \text{et sa marginale} \qquad \mu^N = \sum_{i=1}^{N} w^i_k \; \delta_{\xi^i_k} \; ,$$

où indépendemment pour tout $i = 1, \cdots, N$ la $i$–ème particule $(a^i_k, \zeta^i_k, \xi^i_k)$ est générée d'après $\eta^a_{\rm new}(dz, dx')$ et reçoit un poids donné par $g^a_{\rm new}(z, x')$, i.e.

- l'indice $a^i_k$ est généré suivant le vecteur de probabilité $(\lambda^1_{\rm new}, \cdots, \lambda^N_{\rm new})$, puis, en supposant que $a_k^i = a$

- la particule $\zeta_k^i$ est générée suivant la distribution de probabilité $\pi_{\text{new}}^a(dz)$,

puis, en supposant que $\zeta_k^i = z$

- la particule $\xi_k^i$ est générée suivant la distribution de probabilité $Q_k(z, dx')$,
- et reçoit le poids $w_k^i \propto g_{\text{new}}^a(z, \xi_k^i)$, i.e.

$$w_k^i \propto g_k(\xi_k^i) \; \frac{q_k(\xi_{k-1}^a, \xi_k^i)}{q_k(z, \xi_k^i)} \; \frac{w_{k-1}^a}{\lambda_{\text{new}}^a} \; .$$

□

Pour chaque vecteur $(\pi_{\text{new}}^1(dz), \cdots, \pi_{\text{new}}^N(dz))$ de distributions de probabilité sur $E$ et pour chaque vecteur de probabilité $(\lambda_{\text{new}}^1, \cdots, \lambda_{\text{new}}^N)$, on obtient une décomposition d'importance de la distribution de probabilité $\mu_{\text{new}}^a(dz, dx')$ et une approximation particulaire par échantillonnage pondéré.

(xii) **Montrer que la classe d'approximations particulaires décrite à la question (xi) contient la classe d'approximations particulaires décrite à la question (v), et a fortiori contient l'approximation particulaire décrite à la question (i) comme cas particulier.**

[Indice : Pour un choix particulier du vecteur $(\pi_{\text{new}}^1(dz), \cdots, \pi_{\text{new}}^N(dz))$ de distributions de probabilité sur $E$, l'approximation particulaire décrite à la question (xi) coïncide avec l'approximation particulaire décrite à la question (v).]

Solution

Si $\pi_{\text{new}}^a(dz) = \delta_{\xi_{k-1}^a}(dz)$ pour tout $a \in \{1, \cdots, N\}$ alors dans l'algorithme d'approximation particulaire décrit à la question (xi) l'instruction

- la particule $\zeta_k^i$ est générée suivant la distribution de probabilité $\pi_{\text{new}}^a(dz)$,

signifie juste que $\zeta_k^i = \xi_{k-1}^a$, et l'algorithme devient

- l'indice $a_k^i$ est généré suivant le vecteur de probabilité $(\lambda_{\text{new}}^1, \cdots, \lambda_{\text{new}}^N)$,

puis, en supposant que $a_k^i = a$

- la particule $\xi_k^i$ est générée suivant la distribution de probabilité $Q_k(\xi_{k-1}^a, dx')$,

- et reçoit le poids $w_k^i \propto g_{\text{new}}^a(\xi_{k-1}^a, \xi_k^i)$, i.e.

$$w_k^i \propto g_k(\xi_k^i) \; \frac{w_{k-1}^a}{\lambda_{\text{new}}^a} \; ,$$

et on reconnait l'algorithme d'approximation particulaire décrit à la question (v).

On en déduit le résultat de comparaison suivant. Si on minimise par rapport au vecteur $(\pi_{\text{new}}^1(dz), \cdots, \pi_{\text{new}}^N(dz))$ de distributions de probabilité sur $E$ l'expression de la variance de l'erreur d'estimation de la constante de normalisation $\langle \eta_{\text{new}}, g_{\text{new}} \rangle$ par son approximation particulaire décrite à la question (xi), alors la variance minimale est plus petite que la variance obtenue pour un choix quelconque du vecteur $(\pi_{\text{new}}^1(dz), \cdots, \pi_{\text{new}}^N(dz))$ de distributions de probabilité sur $E$, et en particulier elle est plus petite que la variance correspondant au choix $(\delta_{\xi_{k-1}^1}(dz), \cdots, \delta_{\xi_{k-1}^N}(dz))$ c'est–à–dire plus petite que la variance de l'erreur d'estimation de la constante de normalisation $\langle \eta_{\text{aux}}, g_{\text{aux}} \rangle$ par son approximation particulaire décrite à la question (v).

$\square$

(xiii) **Montrer que la variance (non–asymptotique) de l'erreur d'estimation de la constante de normalisation $\langle \eta_{\text{new}}, g_{\text{new}} \rangle$ par son approximation particulaire est égale à**

$$\frac{1}{N} \; [\sum_{a=1}^N \frac{|w_{k-1}^a|^2}{\lambda_{\text{new}}^a} \; \int_E |u_a(z)|^2 \, \pi_{\text{new}}^a(dz) - \langle \eta, g \rangle^2 \, ] \; ,$$

**avec**

$$u_a(z) = \{ \int_E |g_k(x')|^2 \; |\frac{q_k(\xi_{k-1}^a, x')}{q_k(z, x')}|^2 \; Q_k(z, dx') \}^{1/2} \; ,$$

**pour tout** $a = 1, \cdots, N$ **et pour tout** $z \in E$.

**Dans le cas particulier où** $z = \xi_{k-1}^a$**, montrer que** $u_a(\xi_{k-1}^a) = u_a$ **expression déjà vue à la question (vii), pour tout** $a = 1, \cdots, N$.

Solution

On a vu à la question (xi) que l'approximation particulaire de la constante de normalisation $\langle \eta_{\text{new}}, g_{\text{new}} \rangle$ est définie par

$$\langle \eta_{\text{new}}^N, g_{\text{new}} \rangle = \frac{1}{N} \; \sum_{i=1}^N g_{\text{new}}^{a_k^i}(\zeta_k^i, \xi_k^i) \; ,$$

où les variables aléatoires $(a_k^i, \zeta_k^i, \xi_k^i)$ pour $i = 1, \cdots, N$ sont i.i.d. avec $\eta_{\text{new}}^a(dz, dx')$ comme distribution de probabilité commune, de sorte que

$$\mathbb{E}[\langle \eta_{\text{new}}^N, g_{\text{new}} \rangle] = \langle \eta_{\text{new}}, g_{\text{new}} \rangle \; ,$$

c'est–à–dire que l'estimation est sans biais, et

$$N\ \mathbb{E}|\langle \eta^N_{\text{new}}, g_{\text{new}}\rangle - \langle \eta_{\text{new}}, g_{\text{new}}\rangle|^2 = \text{var}(g_{\text{new}}, \eta_{\text{new}}) = \langle \eta_{\text{new}}, g^2_{\text{new}}\rangle - \langle \eta_{\text{new}}, g_{\text{new}}\rangle^2\ .$$

En fait, compte tenu de l'identité $\langle \eta_{\text{new}}, g_{\text{new}}\rangle = \langle \eta, g\rangle$ montrée à la question (x), la variable aléatoire $\langle \eta^N_{\text{new}}, g_{\text{new}}\rangle$ fournit aussi une estimation sans biais de la constante de normalisation $\langle \eta, g\rangle$, et la variance de l'erreur d'estimation vérifie

$$N\ \mathbb{E}|\langle \eta^N_{\text{new}}, g_{\text{new}}\rangle - \langle \eta, g\rangle|^2 = \langle \eta_{\text{new}}, g^2_{\text{new}}\rangle - \langle \eta, g\rangle^2\ .$$

La variance est controlée par

$$\begin{aligned}
\langle \eta_{\text{new}}, g^2_{\text{new}}\rangle &= \sum_{a=1}^N \int_E \int_E |g^a_{\text{new}}(z,x')|^2\ \eta^a_{\text{new}}(dz,dx') \\
&= \sum_{a=1}^N \int_E \int_E |g_k(x')\ \frac{q_k(\xi^a_{k-1},x')}{q_k(z,x')}\ \frac{w^a_{k-1}}{\lambda^a_{\text{new}}}|^2\ \lambda^a_{\text{new}}\ \pi^a_{\text{new}}(dz)\ Q_k(z,dx') \\
&= \sum_{a=1}^N \frac{|w^a_{k-1}|^2}{\lambda^a_{\text{new}}}\ \int_E [\int_E |g_k(x')|^2\ |\frac{q_k(\xi^a_{k-1},x')}{q_k(z,x')}|^2\ Q_k(z,dx')\,]\ \pi^a_{\text{new}}(dz) \\
&= \sum_{a=1}^N \frac{|w^a_{k-1}|^2}{\lambda^a_{\text{new}}}\ \int_E |u_a(z)|^2\, \pi^a_{\text{new}}(dz)\ ,
\end{aligned}$$

avec

$$u_a(z) = \{\int_E |g_k(x')|^2\ |\frac{q_k(\xi^a_{k-1},x')}{q_k(z,x')}|^2\ Q_k(z,dx')\}^{1/2}\ ,$$

pour tout $a = 1, \cdots, N$ et pour tout $z \in E$. Dans le cas particulier où $z = \xi^a_{k-1}$, cette expression devient

$$u_a(\xi^a_{k-1}) = \{\int_E |g_k(x')|^2\ Q_k(\xi^a_{k-1},dx')\}^{1/2} = u_a\ ,$$

déjà vue à la question (vii) pour tout $a = 1, \cdots, N$.

$\square$

(xiv) **Pour quel choix particulier du vecteur de probabilité $(\lambda^1_{\text{aux}}, \cdots, \lambda^N_{\text{aux}})$, et du vecteur $(\pi^1_{\text{new}}(dz), \cdots, \pi^N_{\text{new}}(dz))$ de distributions de probabilité sur $E$, la variance de l'erreur d'estimation étudiée à la question (xiii) est–elle minimale ? Montrer que cette variance minimale est égale à**

$$\frac{1}{N}\ [\,[\sum_{a=1}^N w^a_{k-1}\, u^\bullet_a\,] - \langle \eta, g\rangle^2\,]\ ,$$

**où $u^\bullet_a = \min_{z \in E} u_a(z)$ pour tout $a = 1, \cdots, N$.**

**Comment cette variance se compare–t–elle à la variance obtenue à la question (vii) ?**

[Indice : On supposera pour simplifier qu'il existe un unique minimiseur $z_a^\bullet$ pour lequel la valeur minimum $u_a^\bullet$ est atteinte.]

Solution

Pour tout $a \in \{1, \cdots, N\}$, on note $u_a^\bullet$ le minimum de $u_a(z)$ quand $z \in E$, de sorte que $u_a^\bullet \leq u_a(\xi_{k-1}^a) = u_a$ en particulier, et en supposant qu'il existe un unique minimiseur $z^\bullet(a)$ où le minimum est atteint, alors

$$\int_E |u_a(z)|^2 \, \pi_{\text{new}}^a(dz) \geq |u_a^\bullet|^2 \ ,$$

et la borne inférieure est atteinte pour le choix $\pi_{\text{opt}}^a(dz) = \delta_{z^\bullet(a)}(dz)$. Si on utilise cette expression, on obtient

$$\begin{aligned}
\min_{\pi_{\text{new}}} \langle \eta_{\text{new}}, g_{\text{new}}^2 \rangle &= \min_{\pi_{\text{new}}} \sum_{a=1}^N \frac{|w_{k-1}^a|^2}{\lambda_{\text{new}}^a} \int_E |u_a(z)|^2 \, \pi_{\text{new}}^a(dz) \\
&= \sum_{a=1}^N \frac{|w_{k-1}^a|^2}{\lambda_{\text{new}}^a} \int_E |u_a(z)|^2 \, \pi_{\text{opt}}^a(dz) \\
&= \sum_{a=1}^N \frac{|w_{k-1}^a|^2}{\lambda_{\text{new}}^a} \, |u_a^\bullet|^2 \ ,
\end{aligned}$$

et le minimum de cette expression par rapport aux poids auxiliaires est atteint pour

$$\lambda_{\text{opt}}^a \propto w_{k-1}^a \, u_a^\bullet \ ,$$

de sorte que, en raisonnant comme dans la réponse à la question (viii)

$$\begin{aligned}
\min_{(\lambda_{\text{new}}, \pi_{\text{new}})} \langle \eta_{\text{new}}, g_{\text{new}}^2 \rangle &= \min_{\lambda_{\text{new}}} \min_{\pi_{\text{new}}} \langle \eta_{\text{new}}, g_{\text{new}}^2 \rangle \\
&= \min_{\lambda_{\text{new}}} \sum_{a=1}^N \frac{|w_{k-1}^a \, u_a^\bullet|^2}{\lambda_{\text{new}}^a} \\
&= \sum_{a=1}^N \frac{|w_{k-1}^a \, u_a^\bullet|^2}{\lambda_{\text{opt}}^a} \\
&= \sum_{a=1}^N \frac{|w_{k-1}^a \, u_a^\bullet|^2}{w_{k-1}^a \, u_a^\bullet} \times \sum_{a=1}^N w_{k-1}^a \, u_a^\bullet \\
&= [\sum_{a=1}^N w_{k-1}^a \, u_a^\bullet]^2 \\
&\leq [\sum_{a=1}^N w_{k-1}^a \, u_a]^2 \ ,
\end{aligned}$$

où la majoration résulte de la relation $u_a^\bullet \leq u_a$ valide pour tout $a = 1, \cdots, N$. On en déduit (ou on retrouve ce qui était attendu) que le filtre particulaire avec transition markovienne auxiliaire avec le choix optimal du vecteur de probabilités auxiliaire et du noyau de transition auxiliaire est plus performant (possède une variance plus petite) que le filtre particulaire auxiliaire avec le choix optimal du vecteur de probabilités auxiliaire.

$\square$