# École Nationale Supérieure de Techniques Avancées module : Commande des Systèmes

# examen du cours OROC–SC–FP "Filtrage bayésien et approximation particulaire"

# vendredi 27 octobre 2017, 13:30 à 15:30

## Problème

L'objectif de ce problème est d'étudier une stratégie alternative d'approximation par échantillonnage pondéré d'une distribution de Gibbs–Boltzmann

$$\mu = g \cdot \eta = \frac{g\,\eta}{\langle \eta, g \rangle} = \frac{\gamma}{\langle \gamma, 1 \rangle} \qquad \text{soit} \qquad \langle \mu, \phi \rangle = \frac{\langle \eta, g\,\phi \rangle}{\langle \eta, g \rangle} = \frac{\langle \gamma, \phi \rangle}{\langle \gamma, 1 \rangle}$$

lorsque la fonction de pondération / sélection $g$ est bornée supérieurement, positive (mais pas strictement positive), et d'intégrale non–nulle par rapport à la distribution de probabilité $\eta$, c'est–à–dire lorsque

$$g(x) \geq 0 \text{ pour tout } x \in E \qquad \text{et} \qquad \sup_{x \in E} g(x) = M < \infty \qquad \text{et} \qquad \langle \eta, g \rangle > 0 \ ,$$

et on souhaite en particulier se prémunir contre le risque d'une extinction du système de particules.

Pour tout entier $N$, on définit les approximations

$$\eta_N = S^N(\eta) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_i} \qquad \text{et} \qquad \mu_N = g \cdot \eta_N = \frac{\displaystyle\sum_{i=1}^N g(\xi_i)\ \delta_{\xi_i}}{\displaystyle\sum_{i=1}^N g(\xi_i)}$$

où $(\xi_1, \cdots, \xi_i, \cdots)$ est une suite de variables aléatoires i.i.d. générées selon la distribution de probabilité $\eta$. L'approximation $\mu_N$ n'est bien définie que si la constante de normalisation

$$\langle \eta_N, g \rangle = \frac{1}{N} \sum_{i=1}^N g(\xi_i) > 0$$

est non nulle (strictement positive).

Pour toute fonction $\phi$ mesurable bornée (et donc intégrable par rapport à la distribution de probabilité $\mu$), on définit la fonction recentrée $\bar{\phi} = \phi - \langle \mu, \phi \rangle$, et pour tout entier $N$, on pose

$$A_N = \sum_{i=1}^{N} [g(\xi_i)\, \bar{\phi}(\xi_i) - \langle \eta, g\, \bar{\phi} \rangle] = \sum_{i=1}^{N} g(\xi_i)\, \bar{\phi}(\xi_i) \qquad \text{et} \qquad D_N = \sum_{i=1}^{N} g(\xi_i) \ .$$

(i) **Sous réserve que $\langle \eta_N, g \rangle \neq 0$, établir l'expression suivante**

$$\langle \mu_N - \mu, \phi \rangle = \frac{\langle \eta_N - \eta, g\, \bar{\phi} \rangle}{\langle \eta_N, g \rangle} = \frac{A_N}{D_N} \ ,$$

**pour l'erreur d'approximation.**

Solution

Par définition

$$\langle \mu_N - \mu, \phi \rangle = \frac{\langle \eta_N, g\, \phi \rangle}{\langle \eta_N, g \rangle} - \langle \mu, \phi \rangle = \frac{\langle \eta_N, g\, (\phi - \langle \mu, \phi \rangle) \rangle}{\langle \eta_N, g \rangle} = \frac{\langle \eta_N - \eta, g\, (\phi - \langle \mu, \phi \rangle) \rangle}{\langle \eta_N, g \rangle} \ ,$$

compte tenu que

$$\langle \eta, g\, (\phi - \langle \mu, \phi \rangle) \rangle = \langle \eta, g \rangle\ \langle \mu, \phi - \langle \mu, \phi \rangle \rangle = 0 \ ,$$

et en terme de sommes empiriques

$$\langle \mu_N - \mu, \phi \rangle = \frac{\langle \eta_N, g\, \bar{\phi} \rangle}{\langle \eta_N, g \rangle} = \frac{\displaystyle\sum_{i=1}^{N} g(\xi_i)\, \bar{\phi}(\xi_i)}{\displaystyle\sum_{i=1}^{N} g(\xi_i)} = \frac{A_N}{D_N} \ .$$

□

La difficulté bien sûr consiste à controler la constante de normalisation $\langle \eta_N, g \rangle$ et à garantir qu'elle soit bien strictement positive. Pour tout $H > 0$ (destiné à tendre vers l'infini), on introduit à cet effet la variable aléatoire

$$T = T_H = \inf\{N \in \mathbb{N}^* \ : \ \sum_{i=1}^{N} g(\xi_i) \geq H\, M\} \ ,$$

à valeur entière (on rappelle que $M$ dénote le supremum (ou juste une borne) de la fonction $g$ supposée bornée supérieurement).

(ii) **Montrer que la variable aléatoire $T = T_H$ à valeur entière est un temps d'arrêt pour la tribu engendrée par les variables** $(\xi_1, \cdots, \xi_i, \cdots)$

- **p.s. fini,**
- **et intégrable (de moyenne finie).**

Solution

Clairement, l'évènement

$$\{T \leq N\} = \{\sum_{i=1}^{N} g(\xi_i) \geq H\, M\} \subset \sigma(\xi_1, \cdots, \xi_N)$$

dépend seulement des variables aléatoires $(\xi_1, \cdots, \xi_N)$. D'après la loi forte des grands nombres

$$\frac{1}{N} \sum_{i=1}^{N} g(\xi_i) \longrightarrow \langle \eta, g \rangle > 0 \qquad \text{de sorte que} \qquad \sum_{i=1}^{N} g(\xi_i) \longrightarrow \infty$$

p.s. quand $N \uparrow \infty$, c'est–à–dire que le seuil $H\, M$ est franchi après un nombre fini de termes.

La méthode de Chernoff donne

$$\begin{aligned}
\mathbb{P}[T > N] &= \mathbb{P}[\sum_{i=1}^{N} g(\xi_i) < H\, M] \\
&= \mathbb{P}[\exp\{-\lambda \sum_{i=1}^{N} g(\xi_i)\} > \exp\{-\lambda\, H\, M\}] \\
&\leq \exp\{\lambda\, H\, M\}\; \mathbb{E}[\exp\{-\lambda \sum_{i=1}^{N} g(\xi_i)\}] \\
&\leq \exp\{\lambda\, H\, M\}\; (\mathbb{E}[\exp\{-\lambda\, g(X)\}])^N \\
&\leq \exp\{\lambda\, H\, M\}\; r^N(\lambda)\ ,
\end{aligned}$$

et on remarque que

$$r(\lambda) = \mathbb{E}[\exp\{-\lambda\, g(X)\}] = \int \exp\{-\lambda\, g(x)\}\, \eta(dx) < 1\ ,$$

si et seulement si $\langle \eta, g \rangle > 0$. En effet

$$\begin{aligned}
\langle \eta, g \rangle = 0 \quad &\Leftrightarrow \quad g(x) = 0 \text{ pour } \eta\text{–presque tout } x \\
&\Leftrightarrow \quad \exp\{-\lambda\, g(x)\} = 1 \text{ pour } \eta\text{–presque tout } x \\
&\Leftrightarrow \quad r(\lambda) = 1\ .
\end{aligned}$$

On introduit la suite décroissante

$$F_N = \mathbb{P}[T \geq N] \qquad \text{avec} \qquad F_1 = 1 \ ,$$

et on remarque que

$$F_N = \mathbb{P}[T = N] + F_{N+1} \ .$$

La différence

$$\sum_{N=1}^{K} N \ (F_N - F_{N+1}) - \sum_{N=1}^{K} F_N = \sum_{N=1}^{K} ((N-1)\, F_N - N \ F_{N+1}) = -K \ F_{K+1} \ ,$$

tend vers zéro quand $K \uparrow \infty$, de sorte que

$$\sum_{N=1}^{K} N \ (F_N - F_{N+1}) \qquad \text{converge ssi} \qquad \sum_{N=1}^{K} F_N \qquad \text{converge}$$

avec des limites égales, et il suffit de vérifier que

$$\begin{aligned} \sum_{N=1}^{K} F_N = \sum_{N=0}^{K-1} F_{N+1} &= \sum_{N=0}^{K-1} \mathbb{P}[T > N] \\ &\leq \exp\{\lambda\, H\, M\} \ \sum_{N=0}^{K-1} r^N(\lambda) \\ &\leq \exp\{\lambda\, H\, M\} \ \frac{1}{1 - r(\lambda)} < \infty \ . \end{aligned}$$

Il en résulte que

$$\begin{aligned} \mathbb{E}[T] = \sum_{N=1}^{\infty} N \ \mathbb{P}[T = N] &= \sum_{N=1}^{\infty} N \ (F_N - F_{N+1}) \\ &= \sum_{N=1}^{\infty} F_N \\ &\leq \exp\{\lambda\, H\, M\} \ \frac{1}{1 - r(\lambda)} < \infty \ . \end{aligned}$$

□

On introduit les approximations $\eta_T$ et $\mu_T$ et les variables $A_T$ et $D_T$, où les sommes portent sur un nombre aléatoire de termes (il suffit de remplacer $N$ par $T$ dans les formules).

(iii) **Montrer que p.s. la constante de normalisation $\langle \eta_T, g \rangle$ est non nulle (strictement positive), et en déduire que l'approximation $\mu_T$ est bien définie.**

Solution

Nécessairement $D_T \geq H\,M$, et par définition

$$\langle \eta_T, g \rangle = \frac{D_T}{T} \geq \frac{H\,M}{T} \ ,$$

et la borne inférieure est strictement positive puisque le dénominateur $T$ est fini p.s.

$\square$

(iv) **Montrer l'encadrement suivant**

$$H\,M \leq D_T \leq (H+1)\,M \ .$$

Solution

Par définition

$$H\,M \leq D_T = D_{T-1} + g(\xi_T) < H\,M + M \ .$$

$\square$

On rappelle les deux identités de Wald

*Soit $(X_1, \cdots, X_i, \cdots)$ des variables aléatoires i.i.d. et soit*

$$S_N = \sum_{i=1}^{N} X_i$$

*la somme partielle définie pour tout entier $N$. Soit $\tau$ un temps d'arrêt (pour la filtration engendrée par les variables aléatoires introduites ci–dessus) et intégrable, c'est–à–dire tel que $\mathbb{E}[\tau] < \infty$.*

*(i) Si $m = \mathbb{E}[X_1] < \infty$, alors*

$$\mathbb{E}[S_\tau] = m\,\mathbb{E}[\tau] \ .$$

*(ii) Si en outre $\sigma^2 = \mathbb{E}|X_1 - m|^2 < \infty$, alors*

$$\mathbb{E}|S_\tau - m\,\tau|^2 = \sigma^2\,\mathbb{E}[\tau] \ .$$

(v) **En utilisant la première identité de Wald, montrer que**

$$\mathbb{E}[T] \leq (H+1)\,\frac{M}{\langle \eta, g \rangle} \ .$$

Solution

On pose $X_i = g(\xi_i)$ pour tout indice $i$ entier non nul, et on vérifie que

$$m = \mathbb{E}[X_1] = \langle \eta, g \rangle \ .$$

En appliquant la première identité de Wald à la somme $D_N$ et au temps d'arrêt $T$, et en utilisant la majoration obtenue en réponse à la question (iv), on a

$$\mathbb{E}[D_T] = \langle \eta, g \rangle \ \mathbb{E}[T] < (H+1)\, M \ ,$$

de sorte que

$$\mathbb{E}[T] < (H+1)\ \frac{M}{\langle \eta, g \rangle} \ .$$

□

(vi) **En utilisant la deuxième identité de Wald, montrer que**

$$\mathbb{E}|A_T|^2 \leq (H+1)\, \frac{M}{\langle \eta, g \rangle}\ \langle \eta, g^2\, \bar{\phi}^2 \rangle \ .$$

Solution

On pose $X_i = g(\xi_i)\, \bar{\phi}(\xi_i)$ pour tout indice $i$ entier non nul, et on vérifie que

$$m = \mathbb{E}[X_1] = \langle \eta, g\, \bar{\phi} \rangle = \langle \eta, g\, (\phi - \langle \mu, \phi \rangle) \rangle = \langle \eta, g \rangle\ \langle \mu, \phi - \langle \mu, \phi \rangle \rangle = 0 \ ,$$

et

$$\sigma^2 = \mathbb{E}|X_1 - m|^2 = \mathbb{E}|X_1|^2 = \langle \eta, g^2\, \bar{\phi}^2 \rangle \ .$$

En appliquant la deuxième identité de Wald à la somme $A_N$ et au temps d'arrêt $T$, et en utilisant la majoration obtenue en réponse à la question (v), on a

$$\mathbb{E}|A_T|^2 = \langle \eta, g^2\, \bar{\phi}^2 \rangle\ \mathbb{E}[T] < (H+1)\ \frac{M}{\langle \eta, g \rangle}\ \langle \eta, g^2\, \bar{\phi}^2 \rangle \ .$$

□

(vii) **En déduire que l'erreur d'approximation vérifie**

$$\mathbb{E}|\langle \mu_T - \mu, \phi \rangle|^2 = \mathbb{E}|\frac{A_T}{D_T}|^2 \leq \frac{1}{H}\ \frac{H+1}{H}\ \frac{\langle \eta, g \rangle}{M}\ \frac{\langle \eta, g^2\, \bar{\phi}^2 \rangle}{\langle \eta, g \rangle^2} \ .$$

Solution

En utilisant la minoration obtenue en réponse à la question (iv), et la majoration obtenue en réponse à la question (vi), on a

$$\mathbb{E}|\langle \mu_T - \mu, \phi \rangle|^2 = \mathbb{E}|\frac{A_T}{D_T}|^2 \leq (\frac{1}{H\,M})^2\, \mathbb{E}|A_T|^2 \leq (\frac{1}{H\,M})^2\, (H+1)\, \frac{M}{\langle \eta, g \rangle}\, \langle \eta, g^2\, \bar{\phi}^2 \rangle \ .$$

□

## Sélection binaire

On s'intéresse à présent au cas particulier où la fonction de pondération / sélection $g$ est binaire, c'est–à–dire à valeurs 0 ou 1, et on va montrer qu'il est possible d'obtenir des résultats plus précis voire explicites, quitte à modifier légèrement l'algorithme.

(viii) **En appliquant directement les résultats précédents, montrer que l'erreur d'approximation vérifie**

$$\mathbb{E}|\langle \mu_T - \mu, \phi \rangle|^2 = \mathbb{E}|\frac{A_T}{D_T}|^2 \leq \frac{1}{H}\, \frac{H+1}{H}\, \mathrm{var}(\phi, \mu) \ .$$

Solution

Il suffit d'appliquer le résultat obtenu en réponse à la question (vii) avec $M = 1$ et avec

$$\frac{\langle \eta, g^2\, \bar{\phi}^2 \rangle}{\langle \eta, g \rangle} = \frac{\langle \eta, g\, \bar{\phi}^2 \rangle}{\langle \eta, g \rangle} = \langle \mu, \bar{\phi}^2 \rangle = \langle \mu, |\phi - \langle \mu, \phi \rangle|^2 \rangle = \mathrm{var}(\phi, \mu) \ ,$$

compte tenu que $g^2 = g$.

□

(ix) **Si $H$ est un entier, montrer que le temps d'arrêt peut alors s'écrire**

$$T = T_H = \inf\{N \in \mathbb{N}^* \ : \ \sum_{i=1}^{N} g(\xi_i) = H\} \ ,$$

**c'est–à–dire comme le nombre d'expériences nécessaires pour obtenir exactement $H$ succès dans un tirage de Bernoulli où la probabilité de succès est $0 < p = \langle \eta, g \rangle \leq 1$.**

Solution

Compte tenu que la fonction de pondération / sélection est à valeurs 0 ou 1, quand $N$ augmente d'une unité la variable $D_N$ reste constante ou bien augmente d'une unité, de telle sorte que le seuil entier $H$ est atteint exactement, avant d'être dépassé.

□

(x) **Montrer que $D_T = H$ et $D_{T-1} = H - 1$.**

SOLUTION

Le seuil entier $H$ est atteint exactement pour $N = T$, de sorte que $D_T = H$. Par ailleurs, la variable $D_N$ a nécessairement augmenté d'une unité pour $N = T$, sinon le seuil entier $H$ aurait déjà été atteint, c'est–à–dire que $D_T - D_{T-1} = 1$ et $D_{T-1} = H - 1$.

□

(xi) **En reprenant les étapes suivies aux questions (v) et (vi), montrer que l'erreur d'approximation vérifie**

$$\mathbb{E}|\langle \mu_T - \mu, \phi \rangle|^2 = \mathbb{E}|\frac{A_T}{D_T}|^2 = \frac{1}{H} \operatorname{var}(\phi, \mu) \ ,$$

**exactement.**

SOLUTION

En appliquant la première identité de Wald à la somme $D_N$ et au temps d'arrêt $T$, et en utilisant l'identité $D_T = H$, on a

$$\mathbb{E}[D_T] = \langle \eta, g \rangle \ \mathbb{E}[T] = H \ ,$$

de sorte que

$$\mathbb{E}[T] = \frac{H}{\langle \eta, g \rangle} \ .$$

En appliquant la deuxième identité de Wald à la somme $A_N$ et au temps d'arrêt $T$, et en utilisant l'expression obtenue ci–dessus, on a

$$\mathbb{E}|A_T|^2 = \langle \eta, g^2 \, \bar{\phi}^2 \rangle \ \mathbb{E}[T] = H \ \frac{\langle \eta, g^2 \, \bar{\phi}^2 \rangle}{\langle \eta, g \rangle} = H \operatorname{var}(\phi, \mu) \ ,$$

grâce à la remarque déjà faite dans la réponse à la question (viii). Finalement, en utilisant à nouveau l'identité $D_T = H$, on a

$$\mathbb{E}|\langle \mu_T - \mu, \phi \rangle|^2 = \mathbb{E}|\frac{A_T}{D_T}|^2 = \frac{1}{H^2} \ \mathbb{E}|A_T|^2 = \frac{1}{H} \operatorname{var}(\phi, \mu) \ .$$

□

On rappelle que le nombre $T$ d'expériences nécessaires pour obtenir exactement $H$ succès dans un tirage de Bernoulli où la probabilité de succès est $p$, suit une loi binomiale inverse (ou loi de Pascal), caractérisée par

$$\mathbb{P}[T = N] = \binom{N-1}{H-1} \ p^H \ (1-p)^{N-H} \qquad \text{pour tout entier } N \geq H.$$

(xii) **Commenter la formule ci–dessus. Montrer que**

$$\mathbb{E}[\frac{1}{T-1}] = \frac{p}{H-1} \ .$$

Solution

L'évènement $\{T = N\}$, en d'autres termes

le $H$–ème succès se produit à la $N$–ème tentative,

est la conjonction de deux évènements indépendants

$(H-1)$ succès sont constatés après $(N-1)$ tentatives,

et

un succès se produit à la $N$–ème tentative.

La probabilité du second évènement est $p$. Le premier évènement est la disjonction exclusive de plusieurs évènements, chacun avec la même probabilité $p^{H-1}\ (1-p)^{N-H}$, et en nombre égal à celui des différentes manières d'obtenir $(H-1)$ succès après $(N-1)$ tentatives, c'est–à–dire égal au nombre de sous–ensembles de $(H-1)$ éléments dans un ensemble de $(N-1)$ éléments, soit $\binom{N-1}{H-1}$.

Par définition

$$\begin{aligned} \mathbb{E}[\frac{1}{T-1}] &= \sum_{N=H}^{\infty} \frac{1}{N-1}\ \mathbb{P}[T=N] \\ &= \sum_{N=H}^{\infty} \frac{1}{N-1}\ \frac{(N-1)!}{(N-H)!\ (H-1)!}\ p^H\ (1-p)^{N-H} \\ &= \frac{p}{H-1}\ \sum_{N=H}^{\infty} \frac{(N-2)!}{(N-H)!\ (H-2)!}\ p^{H-1}\ (1-p)^{N-H}\ , \end{aligned}$$

et en introduisant la notation $J = H-1$ et le changement de variable $M = N-1$, on vérifie que

$$\begin{aligned} \sum_{N=H}^{\infty} \frac{(N-2)!}{(N-H)!\ (H-2)!}\ p^{H-1}\ (1-p)^{N-H} &= \sum_{M=J}^{\infty} \frac{(M-1)!}{(M-J)!\ (J-1)!}\ p^J\ (1-p)^{M-J} \\ &= 1\ . \end{aligned}$$

□

On introduit l'approximation $\eta_T^\bullet = \eta_{T-1}$, où il suffit de remplacer $N$ par $T-1$ (au lieu de $T$) dans les formules.

(xiii) **Montrer que la constante de normalisation vérifie**

$$\langle \eta_T^\bullet, g \rangle = \frac{H-1}{T-1} \ ,$$

**et en déduire qu'il s'agit d'une approximation non–biaisée de la constante de normalisation $\langle \eta, g \rangle$.**

Solution

Par définition, et en utilisant l'identité obtenue à la question (x), on a

$$\langle \eta_T^\bullet, g \rangle = \frac{D_{T-1}}{T-1} = \frac{H-1}{T-1} \ ,$$

et en utilisant le résultat obtenu en réponse à la question (xii), on obtient

$$\mathbb{E}\langle \eta_T^\bullet, g \rangle = (H-1)\ \mathbb{E}[\frac{1}{T-1}] = (H-1)\ \frac{p}{H-1} = p = \langle \eta, g \rangle \ .$$

□