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

vendredi 26 octobre 2018, 13:30 à 15:30

Exercice 1

On considère les distributions normalisées μ_n et $\eta_n = \mu_n^-$ définies par

$$\langle \gamma_n, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^{n} g_k(X_k)\,] \qquad \text{et} \qquad \langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle} \;,$$

et par

$$\langle \gamma_n^-, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^{n-1} g_k(X_k)\,] \qquad \text{et} \qquad \langle \mu_n^-, \phi \rangle = \frac{\langle \gamma_n^-, \phi \rangle}{\langle \gamma_n^-, 1 \rangle} \;,$$

respectivement, pour toute fonction mesurable bornée ϕ, où $\{X_k\}$ est une chaîne de Markov caractérisée par

- la distribution de probabilité initiale $\eta_0(dx)$,
- et les noyaux de probabilités de transition $Q_k(x, dx')$, pour tout $k = 1, \cdots, n$,

et où $g_k(x)$ sont des fonctions mesurables bornées (strictement positives) données, appelées fonctions de sélection, pour tout $k = 0, 1, \cdots, n$. On rappelle que le noyau positif $R_k(x, dx')$ est défini par

$$R_k(x, dx') = Q_k(x, dx')\, g_k(x') \;,$$

pour tout $k = 1, \cdots, n$.

On rappelle que l'approximation particulaire *avec redistribution multinomiale* de la distribution normalisée μ_n vérifie le TCL suivant

$$\sqrt{N}\, \langle \mu_n^N - \mu_n, \phi \rangle \Longrightarrow \mathcal{N}(0, v_n(\phi)) \;,$$

en distribution quand $N \uparrow \infty$, avec la variance asymptotique définie par

$$v_n(\phi) = \sum_{k=0}^{n} \frac{\langle \eta_k, |g_k\, R_{k+1:n}\, (\phi - \langle \mu_n, \phi \rangle)\,|^2\, \rangle}{\langle \eta_k, g_k\, R_{k+1:n}\, 1 \rangle^2} \;,$$

où les fonctions

$$R_{k+1:n}\,\phi(x) = R_{k+1}\cdots R_n\,\phi(x)$$

définies pour tout $k = 0, 1\cdots n$, avec la convention $R_{n+1:n}\,\phi(x) = \phi(x)$ pour $k = n$, se calculent de manière rétrograde.

L'objectif de cet exercice est d'obtenir un TCL pour l'approximation particulaire *avec redistribution multinomiale* de la distribution normalisée $\eta_n = \mu_n^-$.

(i) **Montrer que la suite γ_k^- vérifie la relation de récurrence**

$$\gamma_k^- = \gamma_{k-1}^-\,R_k^- \ ,$$

où le noyau positif $R_k^-(x, dx')$ est défini par

$$R_k^-(x, dx') = g_{k-1}(x)\,Q_k(x, dx') \ ,$$

pour tout $k = 1, \cdots, n$.

On considère les noyaux positifs itérés, définis par leur action sur une fonction mesurable bornée ϕ arbitraire, par

$$R_{k+1:n}^-\,\phi(x) = R_{k+1}^-\cdots R_n^-\,\phi(x) \ ,$$

pour tout $k = 0, 1, \cdots, n$, avec la convention $R_{n+1:n}^-\,\phi(x) = \phi(x)$ pour $k = n$.

(ii) **Montrer (par récurrence arrière) l'identité suivante**

$$R_{k+1:n}^- = g_k\,(R_{k+1:n-1}\,Q_n) \ ,$$

pour tout $k = n-1, \cdots, 0$.

[Indication : il s'agit de montrer que $R_{k+1:n}^-(x, dx') = g_k(x)\,(R_{k+1:n-1}\,Q_n)(x, dx')$.]

(iii) **Montrer que**

$$\sqrt{N}\ \langle \eta_n^N - \eta_n, \phi\rangle \Longrightarrow \mathcal{N}(0, v_n^-(\phi)) \ ,$$

en distribution quand $N \uparrow \infty$, où la variance asymptotique vérifie la relation

$$v_n^-(\phi) = v_{n-1}(Q_n\,\phi) + \mathrm{var}(\phi, \eta_n) \ .$$

(iv) **En utilisant l'identité établie à la question (ii), montrer que la variance asymptotique est aussi donnée par l'expression suivante**

$$v_n^-(\phi) = \sum_{k=0}^{n} \frac{\langle \eta_k, |R_{k+1:n}^-\,(\phi - \langle \eta_n, \phi\rangle)\,|^2\,\rangle}{\langle \eta_k, R_{k+1:n}^-\,1\rangle^2} \ .$$

Exercice 2

Généralités : Soit η une distribution de probabilité sur E. Sans perte de généralité — comme le montre l'exemple des noyaux de Metropolis–Hastings introduits ci–dessous à la question (i) — on suppose l'existence d'un noyau markovien M réversible pour la distribution de probabilité η, c'est–à–dire que

$$\eta(dx)\, M(x, dx') = \eta(dx')\, M(x', dx) \ .$$

En intégrant par rapport à la variable x, on constate que $\eta\, M = \eta$, c'est–à–dire que le noyau markovien M laisse η invariante. On remarque aussi que

$$\begin{aligned} \langle \eta, u\, M\, v \rangle \ &= \int_E \eta(dx)\, u(x) \int_E M(x, dx')\, v(x') \\ &= \int_E \eta(dx')\, v(x') \int_E M(x', dx)\, u(x) = \langle \eta, v\, M\, u \rangle \ , \end{aligned}$$

pour toute fonction u et v définies sur E.

Soit λ une mesure positive dominant η et soit Q un noyau markovien dominé par la même mesure positive λ, c'est–à–dire que

$$\eta(dx) = p(x)\, \lambda(dx) \qquad \text{et} \qquad Q(x, dx') = q(x, x')\ \lambda(dx') \ .$$

(i) **Montrer que le noyau markovien M défini par**

$$M(x, dx') = r(x, x')\ Q(x, dx') + (1 - r(x))\ \delta_x(dx') \ ,$$

avec

$$r(x, x') = \min(1, \frac{p(x')\ q(x', x)}{p(x)\ q(x, x')}) \qquad \textbf{et} \qquad r(x) = \int_E r(x, x')\ Q(x, dx') \ ,$$

pour tout $x, x' \in E$, est réversible pour la distribution de probabilité η.

(ii) **Montrer que pour générer une variable aléatoire Ξ distribuée selon $M(x, dx')$, où l'état initial $x \in E$ est fixé, il suffit de générer une variable aléatoire X' selon $Q(x, dx') = q(x, x')\, dx'$, et de poser ensuite**

$$\Xi = \begin{cases} X', & \textbf{avec probabilité } r(x, X'), \\ x, & \textbf{avec probabilité } (1 - r(x, X')), \end{cases}$$

c'est–à–dire que partant de l'état $x \in E$, on accepte la proposition X' avec probabilité $r(x, X')$ et on conserve l'état initial x sinon.

On définit la distribution de Gibbs–Boltzmann

$$\mu^\bullet = g \cdot \eta = \frac{g\,\eta}{\langle \eta, g \rangle} \ ,$$

où g une fonction positive bornée et où l'intégrale $\langle \eta, g \rangle$ est supposée strictement positive.

(iii) **Montrer que le noyau markovien**

$$M^\bullet(x, dx') = M(x, dx')\, \frac{g(x')}{\lambda} + (1 - \frac{M\,g(x)}{\lambda})\, \delta_x(dx') \qquad \textbf{où} \qquad \lambda = \sup_{x \in E} g(x) \ ,$$

laisse invariante la distribution de Gibbs–Boltzmann $\mu^\bullet$.

(iv) **Montrer que pour générer une variable aléatoire $X^\bullet$ distribuée selon $M^\bullet(x, dx')$, où $x \in E$ est fixé, il suffit de générer une variable aléatoire Ξ distribuée selon $M(x, dx')$, et de poser ensuite**

$$X^\bullet = \begin{cases} \Xi, & \textbf{avec probabilité } \dfrac{g(\Xi)}{\lambda}, \\ x, & \textbf{avec probabilité } (1 - \dfrac{g(\Xi)}{\lambda}), \end{cases}$$

c'est–à–dire que partant de l'état $x \in E$, on accepte la proposition Ξ avec probabilité $\dfrac{g(\Xi)}{\lambda}$ et on conserve l'état initial x sinon.

Application à l'estimation d'une probabilité de dépassement : Soit V une fonction définie sur E et à valeurs positives réelles, et soit $c_* > 0$ un seuil strictement positif. Soit X une variable aléatoire à valeurs dans E et distribuée selon η. On souhaite

- estimer la probabilité de dépassement du seuil c_*, soit

$$p_* = \mathbb{P}[V(X) \geq c_*] \ ,$$

- générer un échantillon distribué (approximativement) selon la loi conditionnelle au dépassement du seuil c_*, soit

$$\mu_*(dx) = \mathbb{P}[X \in dx \mid V(X) \geq c_*] \ .$$

Si la probabilité $p_* \ll 1$ est très petite, il est pertinent d'introduire une suite croissante

$$0 = c_0 < c_1 < \cdots < c_k < \cdots < c_n = c_* \ ,$$

de seuils intermédiaires entre le seuil trivial $c_0 = 0$ et le seuil fixé c_*, avec l'idée de tester $(V(X) \geq c_k)$ successivement pour $k = 0, 1, \cdots, n$, au lieu de tester $(V(X) \geq c_*)$ directement.

Pour tout $k = 1, \cdots, n$ on considère la distribution de probabilité

$$\mu_k = g_k \cdot \eta = \frac{g_k\, \eta}{\langle \eta, g_k \rangle} \ ,$$

où par définition $g_k(x) = 1_{(V(x) \geq c_k)}$.

(v) **Montrer que**

$$\langle \eta, g_k \rangle = \mathbb{P}[V(X) \geq c_k] \qquad \textbf{et} \qquad \mu_k(dx) = \mathbb{P}[X \in dx \mid V(X) \geq c_k] \ ,$$

pour tout $k = 0, 1, \cdots, n$**.**

(vi) **En déduire que**

$$\langle \mu_{k-1}, g_k \rangle = \mathbb{P}[V(X) \geq c_k \mid V(X) \geq c_{k-1}] \ ,$$

pour tout $k = 1, \cdots, n$**.**

On va donc ici

- estimer la probabilité p_* vue comme probabilité de dépassement du seuil final, soit

$$p_* = \mathbb{P}[V(X) \geq c_n] = \langle \eta, g_n \rangle \ ,$$

- générer un échantillon distribué (approximativement) selon la loi conditionnelle μ_*, vue comme loi conditionnelle au dépassement du seuil final, soit

$$\mu_*(dx) = \mathbb{P}[X \in dx \mid V(X) \geq c_n] = \mu_n(dx) \ .$$

(vii) **Montrer que**

$$g_k = g_k\, g_{k-1} \ ,$$

et en déduire que la distribution de probabilité μ_k **vérifie aussi la relation**

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

en terme de la distribution de probabilité μ_{k-1}**.**

(viii) **En utilisant les résultats préliminaires obtenus aux questions (iii) et (iv), montrer que le noyau markovien**

$$M_k(x, dx') = M(x, dx')\, g_{k-1}(x') + (\, 1 - M\, g_{k-1}(x)\,)\, \delta_x(dx') \ ,$$

laisse invariante la distribution de probabilité μ_{k-1}**, et que pour simuler une variable aléatoire** X_k **distribuée selon** $M_k(x, dx')$**, où** $x \in E$ **est fixé, il suffit de générer une variable aléatoire** Ξ **distribuée selon** $M(x, dx')$ **et de poser ensuite :** $X_k = \Xi$ **si** $V(\Xi) \geq c_{k-1}$**, et** $X_k = x$ **sinon.**

(ix) **En déduire que la suite $\{\mu_k\}$ vérifie l'equation récurrente suivante**

$$\mu_{k-1} \xrightarrow{\text{mutation}} \eta_k = \mu_{k-1}\, M_k \xrightarrow{\text{pondération}} \mu_k = g_k \cdot \eta_k \ ,$$

et que

$$\langle \eta_k, g_k \rangle = \langle \mu_{k-1}, g_k \rangle = \mathbb{P}[V(X) \geq c_k \mid V(X) \geq c_{k-1}] \ ,$$

pour tout $k = 1, \cdots, n$.

(x) **Montrer que**

$$p_* = \langle \eta, g_n \rangle = \langle \eta, g_0 \rangle \prod_{k=1}^{n} \langle \eta_k, g_k \rangle \ .$$

(xi) **Montrer comment mettre en œuvre l'algorithme SIR, par exemple avec *redistribution multinomiale*, pour estimer la probabilité p_* de dépassement du seuil c_* et pour générer un échantillon distribué (approximativement) selon la loi μ_* conditionnelle au dépassement du seuil c_*.**

Discuter le rôle du noyau markovien M (utilisé dans la définition du noyau markovien M_k à la question (viii)).