# École Nationale Supérieure de Techniques Avancées
# Filière : Finance quantitative
# Module : Automatique avancée

# Examen du cours B7–3 "Filtrage bayésien optimal et approximation particulaire"
# Jeudi 27 octobre 2005, 13:00 à 16:00

EXERCICE :

On se propose d'estimer la probabilité qu'une chaîne de Markov reste confinée dans un tube, et de simuler cette chaîne de Markov restreinte à ce tube. Il s'agit donc d'estimer la probabilité

$$P_n = \mathbb{P}[X_0 \in A_0, \cdots, X_n \in A_n] ,$$

et la distribution de probabilité conditionnelle

$$\mu_n(dx) = \mathbb{P}[X_n \in dx \mid X_0 \in A_0, \cdots, X_n \in A_n] ,$$

où

- la suite $\{X_k \, , \, k = 0, 1, \cdots, n\}$ est une chaîne de Markov à valeurs dans $E$, de loi initiale
$$\eta_0(dx) = \mathbb{P}[X_0 \in dx] ,$$
et de noyaux de transition
$$Q_k(x, dx') = \mathbb{P}[X_k \in dx' \mid X_{k-1} = x] ,$$
pour tout $k = 1, \cdots, n$,
- les ensembles $A_k \subset E$ sont donnés, pour tout $k = 0, 1, \cdots, n$.

(i) **Montrer que $P_n > 0$ si et seulement si**

$$\eta_0(A_0) > 0 \qquad \textbf{et} \qquad \int_E \mu_{k-1}(dx)\, Q_k(x, A_k) > 0 ,$$

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

Solution

Par définition

$$\mathbb{P}[X_0 \in A_0] = \eta_0(A_0) \qquad \text{et} \qquad \mathbb{P}[X_k \in A_k \mid X_{k-1} = x] = Q_k(x, A_k) \ ,$$

et en utilisant la propriété de Markov, il vient

$$\begin{aligned}
P_k &= \mathbb{P}[X_0 \in A_0, \cdots, X_k \in A_k] \\
&= \mathbb{E}[1_{(X_k \in A_k)} \, 1_{(X_0 \in A_0, \cdots, X_{k-1} \in A_{k-1})}] \\
&= \mathbb{E}[\, \mathbb{P}[X_k \in A_k \mid X_{k-1}] \, 1_{(X_0 \in A_0, \cdots, X_{k-1} \in A_{k-1})}] \\
&= \mathbb{E}[\, Q_k(X_{k-1}, A_k) \mid X_0 \in A_0, \cdots, X_{k-1} \in A_{k-1}\,] \, \mathbb{P}[X_0 \in A_0, \cdots, X_{k-1} \in A_{k-1}] \\
&= P_{k-1} \int_E \mu_{k-1}(dx) \, Q_k(x, A_k) \ ,
\end{aligned}$$

pour tout $k = 1, \cdots, n$, et $P_0 = \eta_0(A_0)$. Par récurrence, on en déduit que $P_n > 0$ si et seulement si

$$\eta_0(A_0) > 0 \qquad \text{et} \qquad \int_E \mu_{k-1}(dx) \, Q_k(x, A_k) > 0 \ ,$$

pour tout $k = 1, \cdots, n$. Compte tenu que le support de $\mu_{k-1}$ est inclus dans $A_{k-1}$, une condition suffisante pour que $P_n > 0$ est que

$$Q_k(x, A_k) > 0 \qquad \text{pour tout } x \in A_{k-1} \ ,$$

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

□

On supposera donc dans toute la suite que ces conditions sont vérifiées.

(ii) **Montrer que la probabilité $P_n$ et la distribution de probabilité conditionnelle $\mu_n$ peuvent s'exprimer à l'aide du flot non–normalisé $\gamma_n$ défini par**

$$\langle \gamma_n, \phi \rangle = \mathbb{E}[\, \phi(X_n) \, 1_{(X_0 \in A_0, \cdots, X_n \in A_n)}] \ ,$$

**pour toute fonction $\phi$ mesurable bornée définie sur $E$.**

Solution

En particulier pour $\phi \equiv 1$

$$\langle \gamma_n, 1 \rangle = \mathbb{P}[X_0 \in A_0, \cdots, X_n \in A_n] \ ,$$

de sorte que

$$\frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle} = \mathbb{E}[\phi(X_n) \mid X_0 \in A_0, \cdots, X_n \in A_n] \ ,$$

pour fonction $\phi$ définie sur $E$, c'est–à–dire que

$$P_n = \langle \gamma_n, 1 \rangle \qquad \text{et} \qquad \langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle} \ .$$

□

(iii) **Donner les équations vérifiées par le flot normalisé et par la constante de normalisation, c'est–à–dire exprimer $\mu_k$ en fonction de $\mu_{k-1}$, et exprimer $P_k$ en fonction de $P_{k-1}$.**

Solution

Il suffit de donner l'équation vérifiée par le flot non–normalisé, c'est–à–dire d'exprimer $\gamma_k$ en fonction de $\gamma_{k-1}$. En posant $g_k(x) = 1_{(x \in A_k)}$, et en utilisant la propriété de Markov, il vient

$$\begin{aligned}
\langle \gamma_k, \phi \rangle &= \mathbb{E}[\phi(X_k)\, g_k(X_k)\, 1_{(X_0 \in A_0, \cdots, X_{k-1} \in A_{k-1})}] \\
&= \mathbb{E}[\, \mathbb{E}[\phi(X_k)\, g_k(X_k) \mid X_{k-1}]\, 1_{(X_0 \in A_0, \cdots, X_{k-1} \in A_{k-1})}] \\
&= \mathbb{E}[Q_k(\phi\, g_k)(X_{k-1})\, 1_{(X_0 \in A_0, \cdots, X_{k-1} \in A_{k-1})}] \\
&= \langle \gamma_{k-1}, Q_k(\phi\, g_k) \rangle \ ,
\end{aligned}$$

pour toute fonction $\phi$ définie sur $E$, et en particulier pour $\phi \equiv 1$

$$\langle \gamma_k, 1 \rangle = \langle \gamma_{k-1}, Q_k\, g_k \rangle = \langle \mu_{k-1}, Q_k\, g_k \rangle\, \langle \gamma_{k-1}, 1 \rangle \ ,$$

et d'autre part

$$\langle \mu_k, \phi \rangle = \frac{\langle \gamma_k, \phi \rangle}{\langle \gamma_k, 1 \rangle} = \frac{\langle \gamma_{k-1}, Q_k(\phi\, g_k) \rangle}{\langle \gamma_{k-1}, Q_k\, g_k \rangle} = \frac{\langle \mu_{k-1}\, Q_k, \phi\, g_k \rangle}{\langle \mu_{k-1}\, Q_k, g_k \rangle} \ .$$

On en déduit que

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

et que

$$P_k = \langle \eta_k, g_k \rangle\, P_{k-1} \ .$$

□

(iv) **En déduire un algorithme particulaire pour l'approximation du flot normalisé $\mu_n$ et de la constante de normalisation $P_n = \langle \gamma_n, 1 \rangle$.**

Solution

L'algorithme SIR standard, aussi appelé algorithme *bootstrap*, pour ce problème propose une approximation sous la forme d'une distribution empirique pondérée

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \, \delta_{\xi_k^i} \qquad \text{avec} \qquad \sum_{i=1}^N w_k^i = 1 \ ,$$

telle que

$$\mu_{k-1}^N \longrightarrow \eta_k^N = S^N(\mu_{k-1}^N \, Q_k) \longrightarrow \mu_k^N = g_k \cdot \eta_k^N \ ,$$

et que

$$P_k \approx P_k^N = \langle \eta_k^N, g_k \rangle \; P_{k-1}^N \ ,$$

ce qui donne pour $k = 0$

- étape de mutation : indépendamment pour tout $i = 1, \cdots, N$, on génère une variable aléatoire $\xi_0^i$ selon la distribution de probabilité $\eta_0(dx)$,
- étape de pondération : pour tout $i = 1, \cdots, N$, on pose

$$w_0^i \propto g_0(\xi_0^i) = 1_{(\xi_0^i \, \in \, A_0)} \ ,$$

et pour $k = 1, \cdots, n$

- étape de sélection : indépendamment pour tout $i = 1, \cdots, N$, on génère une variable aléatoire $\tau_k^i$ à valeurs dans $\{1, \cdots, N\}$ selon les poids $(w_{k-1}^1, \cdots, w_{k-1}^N)$, et on pose $\widehat{\xi}_{k-1}^i = \xi_{k-1}^{\tau_k^i}$,
- étape de mutation : indépendamment pour tout $i = 1, \cdots, N$, on génère une variable aléatoire $\xi_k^i$ selon la distribution de probabilité $Q_k(\widehat{\xi}_{k-1}^i, dx')$,
- étape de pondération : pour tout $i = 1, \cdots, N$, on pose

$$w_k^i \propto g_k(\xi_k^i) = 1_{(\xi_k^i \, \in \, A_k)} \ .$$

Dans cet algorithme, les particules se déplacent selon le modèle markovien, et sont multipliées ou au contraire éliminées selon qu'elles se trouvent ou non dans l'ensemble définissant la contrainte : en particulier

$$\langle \eta_k, g_k \rangle \approx \langle \eta_k^N, g_k \rangle = \frac{1}{N} \sum_{i=1}^N g_k(\xi_k^i) = \frac{1}{N} \sum_{i=1}^N 1_{(\xi_k^i \, \in \, A_k)} = \frac{|I_k^N|}{N} \ ,$$

où
$$I_k^N = \{i = 1 \cdots N \; : \; g_k(\xi_k^i) = 1\} = \{i = 1 \cdots N \; : \; \xi_k^i \in A_k\} \ ,$$
d'où l'interprétation en terme de la fraction des particules qui se trouvent dans l'ensemble définissant la contrainte.

□

(v) **Pour tout $k = 0, 1, \cdots, n$, soit $V_k$ une fonction mesurable définie sur $E$. Montrer comment calculer la probabilité**

$$P_n = \mathbb{P}[\sup_{k=0,1,\cdots,n} V_k(X_k) \leq c] \ ,$$

**et la distribution de probabilité conditionnelle**

$$\mu_n(dx) = \mathbb{P}[X_n \in dx \mid \sup_{k=0,1,\cdots,n} V_k(X_k) \leq c] \ .$$

Solution

Il suffit de remarquer que

$$\{\sup_{k=0,1,\cdots,n} V_k(X_k) \leq c\} = \{V_0(X_0) \leq c, \cdots, V_n(X_n) \leq c\} = \{X_0 \in A_0, \cdots, X_n \in A_n\} \ ,$$

de sorte que
$$P_n = \mathbb{P}[X_0 \in A_0, \cdots, X_n \in A_n] \ ,$$
et
$$\mu_n(dx) = \mathbb{P}[X_n \in dx \mid X_0 \in A_0, \cdots, X_n \in A_n] \ ,$$
où par définition $A_k = \{x \in E \; : \; V_k(x) \leq c\}$, pour tout $k = 0, 1, \cdots, n$. On se ramène donc ainsi au problème considéré au début de l'exercice, pour un choix particulier des ensembles définissant les contraintes.

□

(vi) **En déduire un algorithme particulaire pour l'approximation de la distribution de probabilité conditionnelle $\mu_n$ et de la probabilité $P_n$.**

Solution

L'algorithme SIR standard, aussi appelé algorithme *bootstrap*, pour ce problème propose une approximation sous la forme d'une distribution empirique pondérée

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \, \delta_{\xi_k^i} \qquad \text{avec} \qquad \sum_{i=1}^N w_k^i = 1 \ ,$$

telle que

$$\mu_{k-1}^N \longrightarrow \eta_k^N = S^N(\mu_{k-1}^N\, Q_k) \longrightarrow \mu_k^N = g_k \cdot \eta_k^N \ ,$$

et que

$$P_k \approx P_k^N = \langle \eta_k^N, g_k \rangle \ P_{k-1}^N \ ,$$

ce qui donne pour $k = 0$

- étape de mutation : indépendamment pour tout $i = 1, \cdots, N$, on génère une variable aléatoire $\xi_0^i$ selon la distribution de probabilité $\eta_0(dx)$,
- étape de pondération : pour tout $i = 1, \cdots, N$, on pose

$$w_0^i \propto g_0(\xi_0^i) = 1_{(V_0(\xi_0^i) \leq c)} \ ,$$

et pour $k = 1, \cdots, n$

- étape de sélection : indépendamment pour tout $i = 1, \cdots, N$, on génère une variable aléatoire $\tau_k^i$ à valeurs dans $\{1, \cdots, N\}$ selon les poids $(w_{k-1}^1, \cdots, w_{k-1}^N)$, et on pose $\widehat{\xi}_{k-1}^i = \xi_{k-1}^{\tau_k^i}$,
- étape de mutation : indépendamment pour tout $i = 1, \cdots, N$, on génère une variable aléatoire $\xi_k^i$ selon la distribution de probabilité $Q_k(\widehat{\xi}_{k-1}^i, dx')$,
- étape de pondération : pour tout $i = 1, \cdots, N$, on pose

$$w_k^i \propto g_k(\xi_k^i) = 1_{(V_k(\xi_k^i) \leq c)} \ .$$

Dans cet algorithme, les particules se déplacent selon le modèle markovien, et sont multipliées ou au contraire éliminées selon que la fonction coût prend une valeur inférieure ou non au seuil : en particulier

$$\langle \eta_k, g_k \rangle \approx \langle \eta_k^N, g_k \rangle = \frac{1}{N} \sum_{i=1}^N g_k(\xi_k^i) = \frac{1}{N} \sum_{i=1}^N 1_{(V_k(\xi_k^i) \leq c)} = \frac{|I_k^N|}{N} \ ,$$

où

$$I_k^N = \{i = 1 \cdots N \ : \ g_k(\xi_k^i) = 1\} = \{i = 1 \cdots N \ : \ V_k(\xi_k^i) \leq c\} \ ,$$

d'où l'interprétation en terme de la fraction des particules pour lesquelles la fonction coût prend une valeur inférieure au seuil.

$\square$

## Problème :

L'objectif de ce problème est d'étudier un algorithme stochastique pour la maximisation par rapport au paramètre $\theta \in C$ d'une fonctionnelle intégrale du type

$$U(\theta) = \int_E u(\theta, x)\ \mu(\theta, dx)\ ,$$

où la fonction $u(\theta, x)$ est positive sur $C \times E$, et où $\mu(\theta, dx)$ est une distribution de probabilité sur $E$, pour tout $\theta \in C$. La difficulté du problème tient à ce qu'il n'existe pas en général de forme explicite pour la fonctionnelle $U(\theta)$. En revanche, on suppose qu'il est facile

- d'*évaluer* la fonction $u(\theta, x)$ pour toute valeur $(\theta, x) \in C \times E$,
- de *simuler* pour tout $\theta \in C$ une variable aléatoire à valeurs dans $E$ distribuée selon $\mu(\theta, dx)$,
- de *simuler* une variable aléatoire à valeurs dans $C$ distribuée selon $p(d\theta) = p(\theta)\, d\theta$, où $p(\theta)$ désigne aussi (avec un abus de notation sans conséquence) une densité de probabilité arbitraire sur $C$.

(i) **Montrer que la distribution de probabilité**

$$p_n(d\theta) = p_n(\theta)\, d\theta \propto [U(\theta)]^n\ p(\theta)\, d\theta\ ,$$

**charge asymptotiquement l'ensemble**

$$C^* = \{\theta \in C\ :\ U(\theta) = U^*\} \qquad \textbf{où} \qquad U^* = \sup_{\theta \in C} U(\theta)\ ,$$

**quand $n$ tend vers l'infini (on supposera que $U^*$ est fini). On pourra montrer par exemple que**

$$p_n(\theta \in C\ :\ U(\theta) \leq U^*\, e^{-c}) \leq K(c)\, e^{-\frac{1}{2}\, n\, c}\ ,$$

**pour tout $c > 0$.**

Solution

Par définition

$$p_n(A(c)) = \frac{\displaystyle\int_C 1(\theta \in A(c))\ [U(\theta)]^n\ p(\theta)\, d\theta}{\displaystyle\int_C [U(\theta)]^n\ p(\theta)\, d\theta}\ ,$$

où $A(c) = \{\theta \in C \, : \, U(\theta) \leq U^* \, e^{-c}\}$. On remarque que

$$\int_C 1_{(\theta \in A(c))} \, [U(\theta)]^n \, p(\theta) \, d\theta \leq [U^*]^n \, e^{-n\,c} \, p(A(c)) \ ,$$

et

$$\int_C [U(\theta)]^n \, p(\theta) \, d\theta \geq \int_C 1_{(\theta \notin A(\frac{1}{2}\,c))} \, [U(\theta)]^n \, p(\theta) \, d\theta \geq [U^*]^n \, e^{-\frac{1}{2}\,n\,c} \, p(C \setminus A(\tfrac{1}{2}\,c)) \ ,$$

de sorte que

$$p_n(A(c)) \leq \frac{[U^*]^n \, e^{-n\,c} \, p(A(c))}{[U^*]^n \, e^{-\frac{1}{2}\,n\,c} \, p(C \setminus A(\frac{1}{2}\,c))} \leq e^{-\frac{1}{2}\,n\,c} \, \frac{p(A(c))}{1 - p(A(\frac{1}{2}\,c))} \ .$$

---
□

On se propose alors de générer un échantillon distribué (approximativement) selon $p_n(d\theta)$, qui sera donc concentré autour de l'ensemble $C^*$, pour $n$ assez grand.

(ii) **On considère la distribution de probabilité cible**

$$\mu_n(d\theta, dx_1, \cdots, dx_n) \propto \prod_{k=1}^n u(\theta, x_k) \; \mu(\theta, dx_k) \; p(\theta) \, d\theta \ ,$$

**sur l'espace produit $E_n = C \times E^n$. Montrer que la distribution de probabilité marginale obtenue en intégrant $\mu_n(d\theta, dx_1, \cdots, dx_n)$ par rapport à $x_1, \cdots, x_n$ est exactement la distribution de probabilité $p_n(d\theta)$ introduite à la question (i).**

Solution

Il suffit de remarquer que

$$\int_E \cdots \int_E \prod_{k=1}^n u(\theta, x_k) \; \mu(\theta, dx_k) = \prod_{k=1}^n \int_E u(\theta, x_k) \; \mu(\theta, dx_k) = [U(\theta)]^n \ ,$$

d'où la constante de normalisation

$$c_n = \int_C [U(\theta)]^n \, p(\theta) \, d\theta \ ,$$

pour la distribution de probabilité cible, et la distribution de probabilité marginale est définie par

$$\begin{aligned} \int_E \cdots \int_E \mu_n(d\theta, dx_1, \cdots, dx_n) \; &= \; \frac{1}{c_n} \int_E \cdots \int_E \prod_{k=1}^n u(\theta, x_k) \; \mu(\theta, dx_k) \; p(\theta) \, d\theta \\ &= \; \frac{1}{c_n} \; [U(\theta)]^n \; p(\theta) \, d\theta = p_n(d\theta) \ . \end{aligned}$$

$\square$

On suppose que pour tout $\theta, \theta' \in C$, il existe une fonction mesurable $g(\theta', \theta, x)$ strictement positive définie sur $C \times C \times E$, telle que

$$\mu(\theta', dx) = g(\theta', \theta, x)\ \mu(\theta, dx)\ ,$$

c'est–à–dire telle que les distributions de probabilité sur $E$ associées aux valeurs $\theta$ et $\theta'$ sont équivalentes. On suppose en outre qu'il est facile

- d'*évaluer* la fonction $g(\theta', \theta, x)$ pour toute valeur $(\theta', \theta, x) \in C \times C \times E$.

Pour $y = (\theta, x_1, \cdots, x_n) \in E_n = C \times E^n$ fixé, on définit la variable aléatoire $Y = (\Theta, X_1, \cdots, X_n)$ à valeurs dans $E_n$, de la façon suivante :

- on génère une variable aléatoire $\Xi$ à valeurs dans $C$ distribuée selon $K(\theta, d\xi) = K(\theta, \xi)\, d\xi$, où $K(\theta, \xi)$ désigne aussi (avec un abus de notation sans conséquence) une densité de probabilité arbitraire sur $C$,
- on définit le rapport

$$r(y, \Xi) = r(\theta, x_1, \cdots, x_n, \Xi) = [\prod_{k=1}^{n} \frac{u(\Xi, x_k)}{u(\theta, x_k)}\ g(\Xi, \theta, x_k)\,]\ \frac{p(\Xi)\ K(\Xi, \theta)}{p(\theta)\ K(\theta, \Xi)}\ ,$$

- on pose

$$\Theta = \begin{cases} \Xi & \text{avec probabilité } \min(1, r(\theta, x_1, \cdots, x_n, \Xi)), \\ \\ \theta & \text{sinon,} \end{cases}$$

  et

$$(X_1, \cdots, X_n) = (x_1, \cdots, x_n)$$

(iii) **Dans le cas particulier où le noyau $K(\theta, d\xi)$ est réversible pour la distribution de probabilité $p(d\theta)$, donner une expression simplifiée du rapport de Metropolis–Hastings $r(\theta, x_1, \cdots, x_n, \Xi)$.**

Solution

D'après la propriété de réversibilité

$$p(\Xi)\ K(\Xi, \theta) = p(\theta)\ K(\theta, \Xi)\ ,$$

d'où l'expression simplifiée

$$r(\theta, x_1, \cdots, x_n, \Xi) = \prod_{k=1}^{n} \frac{u(\Xi, x_k)}{u(\theta, x_k)}\ g(\Xi, \theta, x_k)\ ,$$

pour le rapport de Metropolis–Hastings.

$\square$

(iv) **Montrer que la distribution de probabilité de la variable aléatoire $Y$ ainsi définie, est donnée par le noyau de Metropolis–Hastings**

$$
\begin{aligned}
K_n^{\mathrm{MH}}(y, dy') &= K_n^{\mathrm{MH}}(\theta, x_1, \cdots, x_n, d\theta', dx_1', \cdots, dx_n') \\
&= [\min(1, r(\theta, x_1, \cdots, x_n, \theta'))\; K(\theta, \theta')\, d\theta' + p(\theta, x_1, \cdots, x_n)\; \delta_\theta(d\theta')\,] \\
&\qquad \delta_{x_1}(dx_1') \cdots \delta_{x_n}(dx_n') \; .
\end{aligned}
$$

**où**

$$p(\theta, x_1, \cdots, x_n) = 1 - \int_C \min(1, r(\theta, x_1, \cdots, x_n, \xi))\; K(\theta, \xi)\, d\xi \; ,$$

**représente la probabilité de rejeter la proposition $\Xi$. On pourra par exemple vérifier que**

$$\mathbb{E}[\phi(Y)] = \mathbb{E}[\phi(\Theta, X_1, \cdots, X_n)] = K_n^{\mathrm{MH}}\, \phi(\theta, x_1, \cdots, x_n) \; ,$$

**pour toute fonction mesurable bornée $\phi$ définie sur $E_n = C \times E^n$.**

Solution

On vérifie que

$$
\begin{aligned}
\mathbb{E}[\phi(Y) \mid \Xi] &= \mathbb{E}[\phi(\Theta, X_1, \cdots, X_n) \mid \Xi] \\
&= \phi(\Xi, x_1, \cdots, x_n)\; \min(1, r(\theta, x_1, \cdots, x_n, \Xi)) \\
&\qquad + \phi(\theta, x_1, \cdots, x_n)\; (1 - \min(1, r(\theta, x_1, \cdots, x_n, \Xi))) \; ,
\end{aligned}
$$

d'où

$$
\begin{aligned}
\mathbb{E}[\phi(Y)] &= \int_C \phi(\theta', x_1, \cdots, x_n)\; \min(1, r(\theta, x_1, \cdots, x_n, \theta'))\; K(\theta, \theta')\, d\theta' \\
&\qquad + \phi(\theta, x_1, \cdots, x_n)\; (1 - \int_C \min(1, r(\theta, x_1, \cdots, x_n, \theta'))\; K(\theta, \theta')\, d\theta'\,) \\
&= \int_C \phi(\theta', x_1, \cdots, x_n)\; [\min(1, r(\theta, x_1, \cdots, x_n, \theta'))\; K(\theta, \theta')\, d\theta' \\
&\qquad + p(\theta, x_1, \cdots, x_n)\; \delta_\theta(d\theta')\,] \\
&= \int_C \int_E \cdots \int_E \phi(\theta', x_1', \cdots, x_n')\; [\min(1, r(\theta, x_1, \cdots, x_n, \theta'))\; K(\theta, \theta')\, d\theta' \\
&\qquad + p(\theta, x_1, \cdots, x_n)\; \delta_\theta(d\theta')\,]\; \delta_{x_1}(dx_1') \cdots \delta_{x_n}(dx_n') \\
&= K_n^{\mathrm{MH}}\, \phi(\theta, x_1, \cdots, x_n) \; ,
\end{aligned}
$$

pour toute fonction mesurable bornée $\phi$ définie sur $E_n = C \times E^n$.

$\square$

(v) **Montrer que le noyau de Metropolis–Hastings introduit à la question (iv) laisse invariante la distribution de probabilité cible introduite à la question (ii).**

Solution

On rappelle l'expression

$$K_n^{\mathrm{MH}}\, \phi(\theta, x_1, \cdots, x_n) \;=\; \int_C \phi(\theta', x_1, \cdots, x_n)\;\min(1, r(\theta, x_1, \cdots, x_n, \theta'))\; K(\theta, \theta')\, d\theta'$$

$$+\, \phi(\theta, x_1, \cdots, x_n)\; p(\theta, x_1, \cdots, x_n)\;,$$

obtenue dans la réponse à la question (iv) pour toute fonction mesurable bornée $\phi$ définie sur $E_n = C \times E^n$, de sorte que

$$\begin{aligned}
\langle \mu_n\, K_n^{\mathrm{MH}}, \phi \rangle \;&=\; \langle \mu_n, K_n^{\mathrm{MH}}\, \phi \rangle \\
&=\; \int_C \int_C \int_E \cdots \int_E \phi(\theta', x_1, \cdots, x_n)\;\min(1, r(\theta, x_1, \cdots, x_n, \theta')) \\
&\qquad\qquad \mu_n(d\theta, dx_1, \cdots, dx_n)\; K(\theta, \theta')\, d\theta' \\
&\quad +\, \int_C \int_E \cdots \int_E \phi(\theta, x_1, \cdots, x_n)\, p(\theta, x_1, \cdots, x_n)\; \mu_n(d\theta, dx_1, \cdots, dx_n) \\
&=\; I_{\mathrm{move}} + I_{\mathrm{stay}}\;.
\end{aligned}$$

On remarque que

$$\begin{aligned}
\mu_n(d\theta, dx_1, \cdots, dx_n)\; K(\theta, \theta')\, d\theta' \;&=\; \frac{1}{c_n}\, \prod_{k=1}^n u(\theta, x_k)\, \mu(\theta, dx_k)\; p(\theta)\, d\theta\; K(\theta, \theta')\, d\theta' \\
&=\; \prod_{k=1}^n \frac{u(\theta, x_k)}{u(\theta', x_k)}\, g(\theta, \theta', x_k)\, \frac{p(\theta)\, K(\theta, \theta')}{p(\theta')\, K(\theta', \theta)} \\
&\qquad \frac{1}{c_n}\, \prod_{k=1}^n u(\theta', x_k)\, \mu(\theta', dx_k)\; p(\theta')\, d\theta'\; K(\theta', \theta)\, d\theta \\
&=\; r(\theta', x_1, \cdots, x_n, \theta)\; \mu_n(d\theta', dx_1, \cdots, dx_n)\; K(\theta', \theta)\, d\theta\;,
\end{aligned}$$

et d'autre part

$$r(\theta, x_1, \cdots, x_n, \theta')\; r(\theta', x_1, \cdots, x_n, \theta) = \prod_{k=1}^{n} g(\theta', \theta, x_k)\; g(\theta, \theta', x_k) = 1 \ ,$$

de sorte que

$$\min(1, r(\theta, x_1, \cdots, x_n, \theta'))\; r(\theta', x_1, \cdots, x_n, \theta') = \min(1, r(\theta', x_1, \cdots, x_n, \theta)) \ .$$

On en déduit que

$$\begin{aligned}
I_{\text{move}} &= \int_C \int_C \int_E \cdots \int_E \phi(\theta', x_1, \cdots, x_n)\; \min(1, r(\theta, x_1, \cdots, x_n, \theta')) \\
&\qquad\qquad \mu_n(d\theta, dx_1, \cdots, dx_n)\; K(\theta, \theta')\, d\theta' \\
&= \int_C \int_C \int_E \cdots \int_E \phi(\theta', x_1, \cdots, x_n)\; \min(1, r(\theta, x_1, \cdots, x_n, \theta')) \\
&\qquad\qquad r(\theta', x_1, \cdots, x_n, \theta)\; \mu_n(d\theta', dx_1, \cdots, dx_n)\; K(\theta', \theta)\, d\theta \\
&= \int_C \int_C \int_E \cdots \int_E \phi(\theta', x_1, \cdots, x_n)\; \min(1, r(\theta', x_1, \cdots, x_n, \theta)) \\
&\qquad\qquad \mu_n(d\theta', dx_1, \cdots, dx_n)\; K(\theta', \theta)\, d\theta \\
&= \int_C \int_E \cdots \int_E \phi(\theta', x_1, \cdots, x_n)\; \mu_n(d\theta', dx_1, \cdots, dx_n) \\
&\qquad\qquad \int_C \min(1, r(\theta', x_1, \cdots, x_n, \theta))\; K(\theta', \theta)\, d\theta \\
&= \int_C \int_E \cdots \int_E \phi(\theta', x_1, \cdots, x_n)\; (1 - p(\theta', x_1, \cdots, x_n))\; \mu_n(d\theta', dx_1, \cdots, dx_n) \\
&= \langle \mu_n, \phi \rangle - I_{\text{stay}} \ ,
\end{aligned}$$

et finalement

$$\langle \mu_n\, K_n^{\text{MH}}, \phi \rangle = I_{\text{move}} + I_{\text{stay}} = \langle \mu_n, \phi \rangle \ .$$

$\square$

On cherche maintenant à exprimer la distribution de probabilité cible $\mu_n(dy')$, définie sur $E_n = C \times E^n$, en fonction de la distribution de probabilité cible $\mu_{n-1}(dy)$, définie sur $E_{n-1} = C \times E^{n-1}$.

(vi) **Montrer que**

$$\begin{aligned}
\mu_n(dy') &= \mu_n(d\theta', dx'_1, \cdots, dx'_n)\\
&\propto u(\theta', x'_n)\ \mu(\theta', dx'_n)\ \mu_{n-1}\, K^{\mathrm{MH}}_{n-1}(d\theta', dx'_1, \cdots, dx'_{n-1})\\
&\propto g_n(y')\ \mu_{n-1}\, Q_n(dy')\ ,
\end{aligned}$$

**où on donnera l'expression de la fonction de sélection $g_n(y')$ définie sur $E_n$, et où le noyau de transition**

$$\begin{aligned}
Q_n(y, dy') &= Q_n(\theta, x_1, \cdots, x'_{n-1}, d\theta', dx'_1, \cdots, dx'_n)\\
&= \mu(\theta', dx'_n)\ K^{\mathrm{MH}}_{n-1}(\theta, x_1, \cdots, x_{n-1}, d\theta', dx'_1, \cdots, dx'_{n-1})\ ,
\end{aligned}$$

**est défini de $E_{n-1}$ vers $E_n$.**

Solution

Il suffit de remarquer que

$$\begin{aligned}
\mu_n(d\theta', dx'_1, \cdots, dx'_n) &= \frac{1}{c_n} \prod_{k=1}^{n} u(\theta', x'_k)\ \mu(\theta', dx'_k)\ p(\theta')\, d\theta'\\
&= \frac{1}{c_n}\ u(\theta', x'_n)\ \mu(\theta', dx'_n) \prod_{k=1}^{n-1} u(\theta', x'_k)\ \mu(\theta', dx'_k)\ p(\theta')\, d\theta'\\
&= \frac{c_{n-1}}{c_n}\ u(\theta', x'_n)\ \mu(\theta', dx'_n)\ \mu_{n-1}(d\theta', dx'_1, \cdots, dx'_{n-1})\\
&= \frac{c_{n-1}}{c_n}\ u(\theta', x'_n)\ \mu(\theta', dx'_n)\ \mu_{n-1}\, K^{\mathrm{MH}}_{n-1}(d\theta', dx'_1, \cdots, dx'_{n-1})\ ,
\end{aligned}$$

compte tenu que le noyau de Metropolis–Hastings $K^{\mathrm{MH}}_{n-1}$ laisse la distribution de probabilité $\mu_{n-1}$ invariante. Par définition

$$\begin{aligned}
\mu_{n-1}\, Q_n(dy') &= \int_{E_{n-1}} \mu_{n-1}(dy)\, Q_n(y, dy')\\
&= \int_C \int_E \cdots \int_E \mu_{n-1}(d\theta, dx_1, \cdots, dx_{n-1})\\
&\qquad\qquad \mu(\theta', dx'_n)\, K^{\mathrm{MH}}_{n-1}(\theta, x_1, \cdots, x_{n-1}, d\theta', dx'_1, \cdots, dx'_{n-1})\\
&= \mu(\theta', dx'_n)\ \mu_{n-1}\, K^{\mathrm{MH}}_{n-1}(d\theta', dx'_1, \cdots, dx'_{n-1})\ ,
\end{aligned}$$

d'où finalement

$$\mu_n(dy') = \frac{c_{n-1}}{c_n}\, g_n(y')\, \mu_{n-1}\, Q_n(dy')\ ,$$

avec $g_n(y') = g_n(\theta', x'_1, \cdots, x'_n) = u(\theta', x'_n)$. On en déduit que

$$\mu_{n-1} \longrightarrow \eta_n = \mu_{n-1}\, Q_n \longrightarrow \mu_n = g_n \cdot \eta_n\ .$$

□

On dénotera par $\{Y_n\, ,\, n \geq 1\}$ la chaîne de Markov à espace d'état dépendant du temps, défini par les noyaux de transition

$$Q_n(y, dy') = \mathbb{P}[Y_n \in dy' \mid Y_{n-1} = y]\ .$$

(vii) **Déduire de la question précédente un algorithme particulaire pour l'approximation de la distribution de probabilité cible $\mu_n(dy)$.**

Solution

L'algorithme SIR standard, aussi appelé algorithme *bootstrap*, pour ce problème propose une approximation sous la forme d'une distribution empirique trajectorielle pondérée

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i\, \delta_{(\theta_k^i, \xi_1^i, \cdots, \xi_k^i)} \qquad \text{avec} \qquad \sum_{i=1}^N w_k^i = 1\ ,$$

telle que

$$\mu_{k-1}^N \longrightarrow \eta_k^N = S^N(\mu_{k-1}^N\, Q_k) \longrightarrow \mu_k^N = g_k \cdot \eta_k^N\ ,$$

ce qui donne pour $k = 1$

- étape de mutation : indépendamment pour tout $i = 1, \cdots, N$, on génère une variable aléatoire $(\theta_1^i, \xi_1^i)$ selon la distribution de probabilité $\mu(\theta, dx)\, p(\theta)\, d\theta$, c'est–à–dire que
  - on génère d'abord une variable aléatoire $\theta_1^i$ selon la distribution de probabilité $p(\theta)\, d\theta$,
  - et on génère ensuite une variable aléatoire $\xi_1^i$ selon la distribution de probabilité $\mu(\theta_1^i, dx)$,
- étape de pondération : pour tout $i = 1, \cdots, N$, on pose

$$w_1^i \propto u(\theta_1^i, \xi_1^i)\ ,$$

et pour $k = 2, \cdots, n$

- étape de sélection : indépendamment pour tout $i = 1, \cdots, N$, on génère une variable aléatoire $\tau_k^i$ à valeurs dans $\{1, \cdots, N\}$ selon les poids $(w_{k-1}^1, \cdots, w_{k-1}^N)$, et on pose $(\widehat{\theta}_{k-1}^i, \widehat{\xi}_1^i, \cdots, \widehat{\xi}_{k-1}^i) = (\theta_{k-1}^{\tau_k^i}, \xi_1^{\tau_k^i}, \cdots, \xi_{k-1}^{\tau_k^i})$,
- étape de mutation : indépendamment pour tout $i = 1, \cdots, N$, on génère une variable aléatoire $(\theta_k^i, \xi_1^i, \cdots, \xi_k^i)$ selon la distribution de probabilité $Q_k(\widehat{\theta}_{k-1}^i, \widehat{\xi}_1^i, \cdots, \widehat{\xi}_{k-1}^i, d\theta', dx'_1, \cdots, dx'_k)$, c'est–à–dire que
  - on génère d'abord une variable aléatoire $\Xi_k^i$ selon la distribution de probabilité $K(\widehat{\theta}_{k-1}^i, \xi)\, d\xi$, et on pose
  $$\theta_k^i = \begin{cases} \Xi_k^i & \text{avec probabilité } \min(1, r(\widehat{\theta}_{k-1}^i, \widehat{\xi}_1^i, \cdots, \widehat{\xi}_{k-1}^i, \Xi_k^i)), \\ \\ \widehat{\theta}_{k-1}^i & \text{sinon,} \end{cases}$$
  - on pose $(\xi_1^i, \cdots, \xi_{k-1}^i) = (\widehat{\xi}_1^i, \cdots, \widehat{\xi}_{k-1}^i)$,
  - et on génère ensuite une variable aléatoire $\xi_k^i$ selon la distribution de probabilité $\mu(\theta_k^i, dx)$,
- étape de pondération : pour tout $i = 1, \cdots, N$, on pose
$$w_k^i \propto g_k(\theta_k^i, \xi_1^i, \cdots, \xi_k^i) = u(\theta_k^i, \xi_k^i) \ .$$

Dans cet algorithme, les particules *paramètre* se déplacent selon le noyau de Metropolis–Hastings, les particules *variable d'intégration* se déplacent selon la distribution de probabilité par rapport à laquelle on intègre, et les couples *(paramètre, variable d'intégration)* sont multipliées ou au contraire éliminées en fonction de la valeur prise par la fonction intégrée.

---

□

(viii) **En déduire un algorithme particulaire permettant de simuler un échantillon distribué (approximativement) selon $p_n(d\theta)$.**

Solution

D'après la réponse à la question (ii), la distribution de probabilité $p_n(d\theta)$ est exactement la distribution de probabilité marginale obtenue en intégrant $\mu_n(d\theta, dx_1, \cdots, dx_n)$ par rapport à $x_1, \cdots, x_n$. En utilisant l'approximation particulaire

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \ \delta_{(\theta_k^i, \xi_1^i, \cdots, \xi_k^i)} \qquad \text{avec} \qquad \sum_{i=1}^N w_k^i = 1 \ ,$$

introduite en réponse à la question (vii), on obtient donc l'approximation particulaire

$$p_n \approx p_n^N = \sum_{i=1}^N w_k^i \ \delta_{\theta_k^i} \ .$$

---

□