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 $k = 1, \cdots, n$.

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 μ_n peuvent s'exprimer à l'aide du flot non–normalisé γ_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 ϕ mesurable bornée définie sur E.

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

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

(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

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

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

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 $c > 0$.

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).

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 θ et θ' 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 Ξ 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 $r(\theta, x_1, \cdots, x_n, \Xi)$.**

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

$$\begin{aligned} K_n^{\text{MH}}(y, dy') &= K_n^{\text{MH}}(\theta, x_1, \cdots, x_n, d\theta', dx'_1, \cdots, dx'_n) \\ &= [\,\min(1, r(\theta, x_1, \cdots, x_n, \theta'))\ K(\theta, \theta')\, d\theta' + p(\theta, x_1, \cdots, x_n)\ \delta_\theta(d\theta')\,] \\ &\qquad \delta_{x_1}(dx'_1) \cdots \delta_{x_n}(dx'_n)\ . \end{aligned}$$

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 Ξ. 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 ϕ définie sur $E_n = C \times E^n$.

(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).**

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**

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

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

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

est défini de E_{n-1} vers E_n.

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 $\mu_n(dy)$.**

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