École Nationale Supérieure de Techniques Avancées
module : Commande des Systèmes

examen du cours B7–1
“Filtrage bayésien et approximation particulaire”

mercredi 17 octobre 2012, 8:30 à 10:00

Problème

Si on essaye d'étudier le comportement asymptotique de l'algorithme SIR avec redistribution adaptative, on est rapidement confronté au problème que la *taille effective* de l'échantillon qui s'exprime très simplement au niveau algorithmique, en fonction des poids des particules, n'admet pas de contrepartie naturelle au niveau des distributions cibles. L'objectif de ce problème est de proposer une piste pour contourner cet obstacle.

On considère les distributions de Feynman–Kac non–normalisées et normalisées, définies sur l'espace E par

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

où $\{X_k \, , \, k = 0, 1, \cdots, n\}$ est une chaîne de Markov à valeurs dans E et caractérisée par

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

et où $g_k(x')$ est une fonction positive bornée pour tout $k = 0, 1, \cdots, n$. On définit ensuite les distributions non–normalisées et les noyaux positifs non–normalisés

$$\gamma_0(dx) = g_0(x) \, \eta_0(dx) \qquad \text{et} \qquad R_k(x, dx') = Q_k(x, dx') \, g_k(x') \; , \tag{2}$$

et

$$\gamma_0^\square(dx) = g_0^2(x) \, \eta_0(dx) \qquad \text{et} \qquad R_k^\square(x, dx') = Q_k(x, dx') \, g_k^2(x') \; , \tag{3}$$

respectivement, pour tout $k = 1, \cdots, n$. On rappelle que la distribution non–normalisée définie par la représentation probabiliste (1) vérifie la relation de récurrence $\gamma_k = \gamma_{k-1} \, R_k$, pour tout $k = 1, \cdots, n$.

L'approche proposée consiste à considérer un modèle étendu tel que

- la *taille effective* puisse être définie au niveau du modèle étendu, et pas seulement au niveau de l'algorithme, voir la question (ii),
- le modèle ordinaire défini par (1) puisse être vu comme un cas particulier du modèle étendu, voir la question (v),
- l'algorithme SIR avec redistribution adaptative pour le modèle ordinaire défini par (1) puisse s'interpréter comme un algorithme particulaire standard pour le modèle étendu, voir la question (vii).

Modèle de Markov non–linéaire

Sur l'espace produit $E \times [0, \infty)$, on définit *ex abrupto* la distribution de probabilité

$$\mu_0^{\mathrm{e}}(dx_0, dv_0) = \eta_0(dx_0)\ \delta_{g_0(x_0)}(dv_0)\ ,$$

et les noyaux positifs non–normalisés

$$R_k^{\mathrm{e}}(\mu, x, v, dx', dv') = g_{k-1}^{\mathrm{e}}(\mu, v)\ \underbrace{Q_k(x, dx')\ \delta_{f_k^{\mathrm{e}}(\mu, x, v, x')}(dv')}_{Q_k^{\mathrm{e}}(\mu, x, v, dx', dv')}\ ,$$

pour tout $k = 1, \cdots, n$, où

$$f_k^{\mathrm{e}}(\mu, x, v, x') = \begin{cases} v\ g_k(x')\ , & \text{if } \mu \notin D, \\ g_k(x')\ , & \text{if } \mu \in D, \end{cases} \qquad \text{et} \qquad g_{k-1}^{\mathrm{e}}(\mu, v) = \begin{cases} 1\ , & \text{if } \mu \notin D, \\ v\ , & \text{if } \mu \in D, \end{cases}$$

par définition, de sorte que l'identité

$$g_{k-1}^{\mathrm{e}}(\mu, v)\ f_k^{\mathrm{e}}(\mu, x, v, x') = v\ g_k(x')\ ,$$

est vérifiée pour tout $k = 1, \cdots, n$. Clairement

$$R_k^{\mathrm{e}}(\mu, x, v, dx', dv') = \\ = \mathbf{1}(\mu \notin D)\ \underbrace{Q_k(x, dx')\ \delta_{v\, g_k(x')}(dv')}_{Q_k^{\mathrm{imp}}(x, v, dx', dv')} + \mathbf{1}(\mu \in D)\ \underbrace{v\ Q_k(x, dx')\ \delta_{g_k(x')}(dv')}_{R_k^{\mathrm{red}}(x, v, dx', dv')}\ , \tag{4}$$

et on définit aussi

$$g_{k-1}^{\mathrm{red}}(x, v) = v \qquad \text{et} \qquad Q_k^{\mathrm{red}}(x, v, dx', dv') = Q_k(x, dx')\ \delta_{g_k(x')}(dv')\ ,$$

de sorte que

$$R_k^{\text{red}}(x,v,dx',dv') = g_{k-1}^{\text{red}}(x,v)\; Q_k^{\text{red}}(x,v,dx',dv')\ ,$$

pour tout $k=1,\cdots,n$. Comme d'habitude, ces noyaux positifs (normalisés ou non–normalisés) agissent à la fois sur les fonctions et sur les distributions (normalisées ou non–normalisées).

La *région de test* qui déclenche la redistribution est définie *ex abrupto* comme

$$D = \{\mu \in \mathcal{P}(E\times[0,\infty))\ :\ \frac{\langle \mu, 1\otimes e^2\rangle}{\langle \mu, 1\otimes e\rangle^2} \geq c\}\ ,$$

où la fonction identité est notée $e(v)=v$ pour tout $v\geq 0$ et où $c\geq 1$ désigne un seuil à choisir, et la *statistique de test* est définie par

$$H(\mu) = \frac{\langle \mu, 1\otimes e^2\rangle}{\langle \mu, 1\otimes e\rangle^2}\ ,$$

pour tout $\mu \in \mathcal{P}(E\times[0,\infty))$.

On introduit enfin les distributions de Feynman–Kac non–normalisées et normalisées définies sur l'espace produit $E\times[0,\infty)$ par

$$\begin{aligned}\langle \gamma_n^{\text{e}}, F\rangle \ &=\ \int_E\int_0^\infty\cdots\int_E\int_0^\infty \mu_0^{\text{e}}(dx_0,dv_0)\ \prod_{k=1}^n R_k^{\text{e}}(\mu_{k-1}^{\text{e}},x_{k-1},v_{k-1},dx_k,dv_k)\ F(x_n,v_n)\\ \langle \mu_n^{\text{e}}, F\rangle \ &=\ \frac{\langle \gamma_n^{\text{e}}, F\rangle}{\langle \gamma_n^{\text{e}}, 1\rangle}\ ,\end{aligned} \tag{5}$$

respectivement.

(i) **Expliciter la relation de récurrence vérifiée par la distribution non–normalisée γ_k^{e}. En déduire la relation de récurrence vérifiée par la constante de normalisation $\langle \gamma_k^{\text{e}}, 1\rangle$ et par la distribution normalisée μ_k^{e}.**

Les distributions de Feynman–Kac définies en (5) sur l'espace produit $E\times[0,\infty)$ incluent comme cas particulier les distributions de Feynman–Kac définies en (1) sur l'espace E, et aussi d'autres distributions non–normalisées définies sur l'espace E, pour des choix appropriés des fonctions test F. On définit ainsi

$$\Gamma_k = \langle \gamma_k^{\text{e}}, 1\rangle \qquad \text{et} \qquad \langle \gamma_k^{(1)}, \phi\rangle = \langle \gamma_k^{\text{e}}, \phi\otimes e\rangle \qquad \text{et} \qquad \langle \gamma_k^{(2)}, \phi\rangle = \langle \gamma_k^{\text{e}}, \phi\otimes e^2\rangle\ , \tag{6}$$

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

(ii) **Montrer que la statistique de test $H(\mu_k^{\text{e}})$ qui déclenche la redistribution, évaluée pour la distribution μ_k^{e} définie sur l'espace $E\times[0,\infty)$, peut s'exprimer à l'aide des constantes de normalisation Γ_k, $\langle \gamma_k^{(1)}, 1\rangle$ et $\langle \gamma_k^{(2)}, 1\rangle$.**

(iii) **Montrer que**

$$Q_k^{\text{imp}}\,(\phi \otimes e) = (R_k\,\phi) \otimes e \qquad \textbf{et} \qquad Q_k^{\text{imp}}\,(\phi \otimes e^2) = (R_k^\square\,\phi) \otimes e^2 \ .$$

et que

$$R_k^{\text{red}}\,(\phi \otimes e) = (R_k\,\phi) \otimes e \qquad \textbf{et} \qquad R_k^{\text{red}}\,(\phi \otimes e^2) = (R_k^\square\,\phi) \otimes e \ ,$$

pour toute fonction ϕ définie sur E.

(iv) **En déduire que les distributions non–normalisées $\gamma_k^{(1)}$ et $\gamma_k^{(2)}$ et la constante de normalisation Γ_k définies en (6), vérifient les relations de récurrence *fermées* (c'est–à–dire ne faisant pas intervenir d'autres variables supplémentaires) suivantes**

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

pour tout $k = 1, \cdots, n$, avec la condition initiale $\gamma_0^{(1)} = \gamma_0$, et

$$\gamma_k^{(2)} = \mathbf{1}_{(\mu_{k-1}^{\text{e}} \not\in D)}\ \gamma_{k-1}^{(2)}\,R_k^\square + \mathbf{1}_{(\mu_{k-1}^{\text{e}} \in D)}\ \gamma_{k-1}\,R_k^\square \ ,$$

pour tout $k = 1, \cdots, n$, avec la condition initiale $\gamma_0^{(2)} = \gamma_0^\square$, et

$$\Gamma_k = \mathbf{1}_{(\mu_{k-1}^{\text{e}} \not\in D)}\ \Gamma_{k-1} + \mathbf{1}_{(\mu_{k-1}^{\text{e}} \in D)}\ \langle \gamma_{k-1}, 1 \rangle \ ,$$

pour tout $k = 1, \cdots, n$, avec la condition initiale $\Gamma_0 = 1$.

(v) **En déduire en particulier que la distribution non–normalisée $\gamma_k^{(1)}$ définie en (6) coïncide avec la distribution non–normalisée γ_k définie en (1), et que la distribution normalisée μ_k définie en (1) peut s'exprimer à l'aide de la distribution normalisée μ_k^{e} définie en (5), pour un choix approprié des fonctions test.**

Approximation particulaire

On recherche une approximation particulaire sous la forme

$$\mu_k^{\text{e}} \approx \mu_k^{\text{e},N} = \frac{1}{N}\ \sum_{i=1}^N \delta_{(\xi_k^i,\, v_k^i)} \ ,$$

et partant à l'instant $(k-1)$ d'une approximation particulaire sous la forme

$$\mu_{k-1}^{\text{e}} \approx \mu_{k-1}^{\text{e},N} = \frac{1}{N}\ \sum_{i=1}^N \delta_{(\xi_{k-1}^i,\, v_{k-1}^i)} \ ,$$

on obtient à l'instant k deux approximations particulaires différentes, selon que le test qui déclenche la redistribution est activé ou non : si $\mu_{k-1}^{e,N} \notin D$, alors

$$\mu_{k-1}^{e,N}\, Q_k^{\mathrm{imp}}(dx', dv') = \frac{1}{N} \sum_{i=1}^N \underbrace{Q_k(\xi_{k-1}^i, dx')\; \delta_{v_{k-1}^i\, g_k(x')}(dv')}_{m_k^i(dx', dv')} \;,$$

d'où l'approximation particulaire

$$\mu_k^{e,N} = \frac{1}{N} \sum_{i=1}^N \delta_{(\xi_k^i,\, v_k^i)} \;,$$

où indépendemment pour tout $i = 1, \cdots, N$ la variable aléatoire (ξ_k^i, v_k^i) est distribuée selon $m_k^i(dx', dv')$, ou autrement dit

$$\xi_k^i \sim Q_k(\xi_{k-1}^i, dx') \qquad \text{et} \qquad v_k^i = v_{k-1}^i\, g_k(\xi_k^i) \;,$$

et sinon, si $\mu_{k-1}^{e,N} \in D$, alors

$$g_{k-1}^{\mathrm{red}} \cdot \mu_{k-1}^{e,N} = \frac{\displaystyle\sum_{i=1}^N v_{k-1}^i\; \delta_{(\xi_{k-1}^i,\, v_{k-1}^i)}}{\displaystyle\sum_{i=1}^N v_{k-1}^i} = \sum_{i=1}^N w_{k-1}^i\; \delta_{(\xi_{k-1}^i,\, v_{k-1}^i)}$$

ce qui définit implicitement les nouveaux poids normalisés et

$$(g_{k-1}^{\mathrm{red}} \cdot \mu_{k-1}^{e,N})\, Q_k^{\mathrm{red}}(dx', dv') = \underbrace{\sum_{i=1}^N w_{k-1}^i\; Q_k(\xi_{k-1}^i, dx')\; \delta_{g_k(x')}(dv')}_{\bar{m}_k(dx', dv')} \;,$$

d'où l'approximation particulaire

$$\mu_k^{e,N} = \frac{1}{N} \sum_{i=1}^N \delta_{(\xi_k^i,\, v_k^i)} \;,$$

où indépendamment pour tout $i = 1, \cdots, N$ la variable aléatoire (ξ_k^i, v_k^i) est distribuée selon $\bar{m}_k(dx', dv')$, ou autrement dit

$$\xi_k^i \sim Q_k(\widehat{\xi}_{k-1}^i, dx') \qquad \text{et} \qquad v_k^i = g_k(\xi_k^i) \;,$$

où la variable aléatoire $\widehat{\xi}_{k-1}^i$ est sélectionnée (avec remplacement) au sein de la population courante $(\xi_{k-1}^1, \cdots, \xi_{k-1}^N)$ et selon les poids $(w_{k-1}^1, \cdots, w_{k-1}^N)$.

(vi) **Montrer que la statistique de test $H(\mu_{k-1}^{\mathrm{e},N})$ qui déclenche la redistribution, évaluée pour l'approximation particulaire $\mu_{k-1}^{\mathrm{e},N}$, coïncide avec la définition usuelle de la *taille effective* de l'échantillon.**

(vii) **Pour un choix approprié des fonctions test, montrer que l'approximation particulaire proposée pour le modèle étendu coïncide avec l'algorithme SIR avec redistribution adaptative pour le modèle ordinaire.**