École Nationale Supérieure de Techniques Avancées
Filière : Finance quantitative
Module : Commande des systèmes

Examen du cours B7–3
“Filtrage bayésien optimal et approximation particulaire”
Lundi 22 octobre 2007, 8:30 à 11:30

Exercice :

L'objectif de cet exercice est d'étudier une classe d'algorithmes de filtrage particulaire généralisés, connue sous le terme générique ABC (pour *approximate Bayesian computation*, en anglais), qui peut s'appliquer aux cas où

- il n'existe pas d'expression suffisamment explicite pour la fonction de vraisemblance,
- la distribution de probabilité de l'observation sachant l'état caché ne possède pas de densité (par rapport à une mesure ne dépendant pas de l'état caché), par exemple parce que le bruit d'observation n'apparaît pas de manière additive, c'est–à–dire qu'il n'existe simplement pas de fonction de vraisemblance,
- certaines composantes de l'observation sont observées exactement, sans bruit d'observation, ce qui peut s'interpréter comme une contrainte a posteriori sur l'état caché, etc.

On considère seulement le cas statique, dans l'une des deux situations suivantes

Modèle A La variable cachée X à valeurs dans E est distribuée selon $\mu(dx)$, il est facile de simuler une variable aléatoire selon la distribution $\mu(dx)$, et l'observation à valeurs dans $\mathbb{R}^d$ est reliée à l'état caché par la relation $Y = h(X)$.

Modèle B La variable cachée X à valeurs dans E et le bruit d'observation V à valeurs dans F sont indépendants, distribués selon $\mu(dx)$ et $\lambda(dv)$ respectivement, il est facile de simuler une variable aléatoire aussi bien selon la distribution $\mu(dx)$ que selon la distribution $\lambda(dv)$, et l'observation à valeurs dans $\mathbb{R}^d$ est reliée à l'état caché par la relation $Y = c(X, V)$.

Clairement, le modèle A est un cas particulier du modèle B.

(i) **Réciproquement, en posant $Z = (X, V)$, montrer que le modèle B se ramène au modèle A.**

On se contente donc dans la suite d'étudier le modèle A uniquement.

(ii) **Donner l'expression de la distribution de probabilité de l'observation Y sachant $X = x$. Que remarque–t–on ? Que peut–on dire qualitativement de la distribution conditionnelle de la variable cachée X sachant $Y = y$?**

(iii) **Montrer comment générer un échantillon $(X_i, Y_i, i = 1, \cdots, N)$ selon la distribution de probabilité jointe de la variable cachée et de l'observation.**

On définit une première approximation de la distribution de probabilité jointe de la variable cachée et de l'observation, sous la forme de la distribution empirique

$$\mathbb{P}[X \in dx, Y \in dy] \approx \frac{1}{N} \sum_{i=1}^{N} \delta_{(X_i, Y_i)}(dx, dy) \; .$$

On introduit ensuite une famille de noyaux régularisants : à partir d'une densité de probabilité $K(y)$ définie sur $\mathbb{R}^d$, on pose $K_\varepsilon(y) = \varepsilon^{-d} \; K(\varepsilon^{-1} \, y)$ pour tout $\varepsilon > 0$. On construit ainsi une seconde approximation de la distribution de probabilité jointe de la variable cachée et de l'observation, par convolution avec le noyau régularisant, c'est–à–dire

$$\mathbb{P}[X \in dx, Y \in dy] \approx \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}(dx) \; K_\varepsilon(y - Y_i) \; dy \; .$$

(iv) **En déduire par marginalisation une approximation de la distribution de probabilité de l'observation, puis une approximation de la distribution de probabilité conditionnelle de la variable cachée sachant l'observation.**

Pour interpréter le résultat obtenu, on considère le modèle perturbé suivant : la variable cachée X à valeurs dans E et la perturbation aléatoire U à valeurs dans $\mathbb{R}^d$ sont indépendantes, distribuées selon $\mu(dx)$ et selon $K(u) \, du$ respectivement, et l'observation est reliée à l'état caché par la relation $Y = h(X) + \varepsilon \, U$.

(v) **Montrer que le modèle perturbé vérifie les hypothèses faites dans le cours, et que l'approximation obtenue en réponse à la question (iv) correspond à l'approximation par échantillonnage pondéré dans le modèle perturbé.**

On considère le cas particulier où $K(u)$ est la densité uniforme sur la boule unité de $\mathbb{R}^d$, et où il existe un noyau markovien $M(x, dx')$ réversible pour la distribution de probabilité $\mu(dx)$.

(vi) **En introduisant une suite décroissante $\varepsilon_1 > \cdots > \varepsilon_n > 0$ et en s'inspirant des exemples vus en cours et en travaux pratiques, proposer un algorithme particulaire pour l'approximation de la distribution conditionnelle de l'état caché sachant l'observation dans le modèle A.**

Problème :

L'objectif de ce problème est d'étudier une classe d'algorithmes d'approximation particulaire qui combine les algorithmes SIS et SIR, avec deux sortes de poids affectés à chaque particule

- des poids utilisés pour la redistribution, comme dans l'algorithme SIR,
- et des poids utilisés simplement pour la pondération, comme dans l'algorithme SIS.

On considère la distribution non–normalisée, et la distribution normalisée correspondante, définies sur l'ensemble E par

$$\langle \gamma_n, \phi \rangle = \int_E \cdots \int_E \phi(x_n)\, \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k) \qquad \text{et} \qquad \langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle} \ ,$$

où $\gamma_0(dx)$ est une mesure positive (non–normalisée) et où $R_k(x, dx')$ est un noyau positif (non–normalisé) pour tout $k = 1, \cdots, n$. En toute généralité, il est toujours possible de décomposer la mesure positive (non–normalisée)

$$\gamma_0(dx) = W_0(x)\, p_0(dx) \ ,$$

en termes d'une fonction positive et d'une distribution de probabilité, et de décomposer le noyau positif (non–normalisé)

$$R_k(x, dx') = W_k(x, x')\, P_k(x, dx') \ ,$$

en termes d'une fonction positive et d'un noyau markovien pour tout $k = 1, \cdots, n$.

Si on introduit une factorisation supplémentaire des fonctions positives

$$W_0(x) = W_0^{\rm imp}(x)\, W_0^{\rm red}(x) \qquad \text{et} \qquad W_k(x, x') = W_k^{\rm imp}(x, x')\, W_k^{\rm red}(x, x') \ ,$$

pour tout $k = 1, \cdots, n$, alors on obtient la décomposition suivante

$$\begin{aligned} \gamma_0(dx) &= W_0^{\rm imp}(x)\, W_0^{\rm red}(x)\, p_0(dx) \ , \\ R_k(x, dx') &= W_k^{\rm imp}(x, x')\, W_k^{\rm red}(x, x')\, P_k(x, dx') \ , \end{aligned} \tag{1}$$

pour tout $k = 1, \cdots, n$, et l'idée consiste à exploiter cette décomposition pour proposer une approximation particulaire de la forme

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

où

- les poids $(w_k^1, \cdots, w_k^N)$ sont utilisés pour la redistribution,
- et les poids d'importance $(u_k^1, \cdots, u_k^N)$ sont utilisés simplement pour la pondération.

Une motivation possible pour cette étude est l'approximation particulaire conjointe de distributions associées à un modèle de référence et à des modèles alternatifs, sous l'hypothèse suivante d'absolue continuité

$$\gamma_0(dx) = r_0(x)\, \gamma_0^0(dx) \qquad \text{et} \qquad R_k(x, dx') = r_k(x, x')\, R_k^0(x, dx') \ ,$$

pour tout $k = 1, \cdots, n$. Dans ce cas en effet, si on introduit une décomposition de la mesure positive (non–normalisée) de référence

$$\gamma_0^0(dx) = W_0^0(x)\, p_0^0(dx) \ ,$$

en termes d'une fonction positive et d'une distribution cde probabilité, et si on introduit une décomposition du noyau positif (non–normalisé) de référence

$$R_k^0(x, dx') = W_k^0(x, x')\, P_k^0(x, dx') \ ,$$

en termes d'une fonction positive et d'un noyau markovien pour tout $k = 1, \cdots, n$, alors on obtient la décomposition suivante

$$\gamma_0(dx) = r_0(x)\, W_0^0(x)\, p_0^0(dx) \ ,$$

$$R_k(x, dx') = r_k(x, x')\, W_k^0(x, x')\, P_k^0(x, dx') \ ,$$

pour tout $k = 1, \cdots, n$, qui est clairement de la forme (1), et il est possible de proposer une approximation particulaire de la forme

$$\mu_k^N = \sum_{i=1}^N u_k^i\, w_k^{i,0}\, \delta_{\xi_k^{i,0}} \qquad \text{avec} \qquad \sum_{i=1}^N w_k^{i,0} = 1 \qquad \text{et} \qquad \sum_{i=1}^N u_k^i\, w_k^{i,0} = 1$$

où

- les positions $(\xi_k^{1,0}, \cdots, \xi_k^{N,0})$ des particules et les poids $(w_k^{1,0}, \cdots, w_k^{N,0})$ utilisés pour la redistribution dépendent du modèle de référence uniquement,
- les poids d'importance $(u_k^1, \cdots, u_k^N)$ utilisés simplement pour la pondération dépendent à la fois de modèle de référence et du modèle alternatif.

Bien entendu, les deux points de vue sont mathématiquement équivalents, à une identification près des différentes mesures positives et des différentes fonctions positives impliquées, et le choix de l'un ou l'autre point de vue dépend souvent de l'application.

Notation On utilisera souvent dans la suite les abus de notation $W_0^{\rm imp}(x,x') = W_0^{\rm imp}(x')$ et $W_0^{\rm red}(x,x') = W_0^{\rm red}(x')$.

Représentation trajectorielle

(i) **En utilisant la décomposition (1), montrer que**

$$\langle \gamma_n, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^n W_k^{\rm imp}(X_{k-1}, X_k) \prod_{k=0}^n W_k^{\rm red}(X_{k-1}, X_k)] \ ,$$

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

- **la distribution de probabilité initiale $p_0(dx)$,**
- **et les probabilités de transition $P_k(x, dx')$ pour tout $k = 0, 1, \cdots, n$.**

On pose $X_k^\bullet = (X_0, \cdots, X_k) = X_{0:k}$ pour tout $k = 0, 1, \cdots, n$, et on introduit la distribution non–normalisée, et la distribution normalisée correspondante, définies sur l'ensemble trajectoriel $E_{0:n} = E \times \cdots \times E$ par

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

respectivement, où $g_k^\bullet(x_{0:k}) = W_k^{\rm red}(x_{k-1}, x_k)$ pour tout $k = 0, 1, \cdots, n$.

(ii) **Montrer que la suite $\{X_k^\bullet \, , \, k = 0, 1, \cdots, n\}$ est une chaîne de Markov à valeurs dans l'ensemble des trajectoires dans E, caractérisée par**

- **la distribution de probabilité initiale $\eta_0^\bullet(dx_0) = p_0(dx_0)$,**
- **et les probabilités de transition**

$$Q_k^\bullet(x_{0:k-1}, dx'_{0:k}) = \delta_{x_{0:k-1}}(dx'_{0:k-1}) \; P_k(x_{k-1}, dx'_k) \ ,$$

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

(iii) **En déduire, en utilisant les résultats du cours, l'équation récurrente vérifiée par la suite $\{\gamma_k^\bullet \, , \, k = 0, 1, \cdots, n\}$ des distributions non–normalisées définies sur l'ensemble des trajectoires dans E.**

(iv) **Montrer que**

$$\langle \gamma_n, \phi \rangle = \langle \gamma_n^\bullet, T_n \, \phi \rangle \ ,$$

où la fonction $T_n \, \phi$ est définie sur l'ensemble trajectoriel $E_{0:n} = E \times \cdots \times E$ par

$$T_n \, \phi(x_{0:n}) = \phi(x_n) \prod_{k=0}^n W_k^{\rm imp}(x_{k-1}, x_k) \ .$$

En d'autres termes, la distribution non–normalisée $\gamma_n^\bullet$ définie sur l'ensemble trajectoriel $E_{0:n} = E \times \cdots \times E$ englobe comme cas particulier la distribution non–normalisée γ_n.

(v) **En déduire l'expression**

- **de la constante de normalisation $\langle \gamma_n, 1 \rangle$,**
- **et de la distribution normalisée μ_n,**

en fonction de la distribution non–normalisée $\gamma_n^\bullet$ définie sur l'ensemble trajectoriel $E_{0:n} = E \times \cdots \times E$.

On recherche une approximation particulaire pondérée de la forme

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

où $\xi_k^{\bullet,i} = (\xi_{0,k}^i, \cdots, \xi_{k,k}^i)$ est une particule dans l'ensemble trajectoriel $E_{0:k} = E \times \cdots \times E$ dont la position terminale est notée $\xi_k^i = \xi_{k,k}^i$ pour tout $i = 1, \cdots, N$.

(vi) **Décrire, en utilisant les résultats du cours, l'approximation particulaire de type bootstrap pour la distribution $\mu_k^\bullet$ définie sur l'ensemble trajectoriel $E_{0:k} = E \times \cdots \times E$. En déduire, en utilisant les expressions obtenues en réponse à la question (iv), une approximation particulaire de type bootstrap pour la distribution μ_k.**

Représentation en terme d'une fonctionnelle multiplicative

En partant plutôt de l'hypothèse d'absolue continuité

$$\gamma_0(dx) = W_0^{\text{imp}}(x)\, \gamma_0^0(dx) \qquad \text{et} \qquad R_k(x, dx') = W_k^{\text{imp}}(x, x')\, R_k^0(x, dx') \ , \tag{2}$$

ce qui définit implicitement, au vu de la décomposition (1)

$$\gamma_0^0(dx) = W_0^{\text{red}}(x)\, p_0(dx) \qquad \text{et} \qquad R_k^0(x, dx') = W_k^{\text{red}}(x, x')\, P_k(x, dx') \ , \tag{3}$$

pour tout $k = 1, \cdots, n$, on introduit la distribution non–normalisée, et la distribution normalisée correspondante, définies sur l'ensemble produit $E \times [0, \infty)$ par

$$\langle \gamma_n^{\text{e}}, F \rangle = \int_E \cdots \int_E F(x_n, \prod_{k=0}^n W_k^{\text{imp}}(x_{k-1}, x_k))\ \gamma_0^0(dx_0) \prod_{k=1}^n R_k^0(x_{k-1}, dx_k) \ ,$$

et

$$\langle \mu_n^{\text{e}}, F \rangle = \frac{\langle \gamma_n^{\text{e}}, F \rangle}{\langle \gamma_n^{\text{e}}, 1 \rangle} \ ,$$

respectivement.

(vii) **Montrer que**

$$\gamma_0^{\rm e}(dx,dv) = \gamma_0^0(dx)\ \delta_{W_0^{\rm imp}(x)}(dv)\ ,$$

à l'instant initial.

Pour le modèle de référence, on introduit la distribution non–normalisée, et la distribution normalisée correspondante, définies sur E par

$$\langle \gamma_n^0, \phi\rangle = \int_E \cdots \int_E \phi(x_n)\ \gamma_0^0(dx_0)\ \prod_{k=1}^n R_k^0(x_{k-1},dx_k) \qquad \text{et} \qquad \langle \mu_n^0, \phi\rangle = \frac{\langle \gamma_n^0, \phi\rangle}{\langle \gamma_n^0, 1\rangle}\ ,$$

respectivement. Pour tout $v \in [0,\infty)$, on pose $e_0(v) \equiv 1$ et $e(v) = v$.

(viii) **Montrer que**

$$\langle \gamma_n^0, \phi\rangle = \langle \gamma_n^{\rm e}, \phi \otimes e_0\rangle\ ,$$

où la fonction $F = \phi \otimes e_0$ est définie sur l'ensemble produit $E \times [0,\infty)$ par $F(x,v) = \phi(x)$, et montrer que

$$\langle \gamma_n, \phi\rangle = \langle \gamma_n^{\rm e}, \phi \otimes e\rangle\ ,$$

où la fonction $F = \phi \otimes e$ est définie sur l'ensemble produit $E \times [0,\infty)$ par $F(x,v) = \phi(x)\ v$.

En d'autres termes, la distribution non–normalisée $\gamma_n^{\rm e}$ définie sur l'ensemble produit $E \times [0,\infty)$ englobe simultanément les deux distributions non–normalisées γ_n^0 et γ_n associées respectivement au modèle de référence et au modèle alternatif.

(ix) **En déduire l'expression**

- **des deux constantes de normalisation $\langle \gamma_n^0, 1\rangle$ et $\langle \gamma_n, 1\rangle$,**
- **et des deux distributions normalisées μ_n^0 et μ_n,**

associées au modèle de référence et au modèle alternatif, en fonction de la distribution non–normalisée $\gamma_n^{\rm e}$ définie sur l'ensemble produit $E \times [0,\infty)$.

On introduit le noyau positif (non–normalisé)

$$R_k^{\rm e}(x,v,dx',dv') = R_k^0(x,dx')\ \delta_{v\, W_k^{\rm imp}(x,x')}(dv')\ ,$$

défini sur l'ensemble produit $E \times [0,\infty)$ pour tout $k = 1,\cdots,n$.

(x) **Montrer que**

$$\langle \gamma_n^{\rm e}, F \rangle = \int_E \int_0^\infty \cdots \int_E \int_0^\infty F(x_n, v_n)\; \gamma_0^{\rm e}(dx_0, dv_0)\; \prod_{k=1}^n R_k^{\rm e}(x_{k-1}, v_{k-1}, dx_k, dv_k)\ ,$$

ce qui justifie l'interprétation de la distribution $\gamma_n^{\rm e}$ comme une distribution de Feynman–Kac non–normalisée.

(xi) **En déduire, en utilisant les résultats du cours, l'équation récurrente vérifiée par la suite $\{\gamma_k^{\rm e}\,,\, k = 0, 1, \cdots, n\}$ des distributions non–normalisées définies sur l'ensemble produit $E \times [0, \infty)$.**

(xii) **En revenant à la définition, et en utilisant la décomposition (3), montrer que**

$$\langle \gamma_n^{\rm e}, F \rangle = \mathbb{E}[F(X_n, \prod_{k=0}^n W_k^{\rm imp}(X_{k-1}, X_k))\; \prod_{k=0}^n W_k^{\rm red}(X_{k-1}, X_k)\,]\ ,$$

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

- **la distribution de probabilité initiale $p_0(dx)$,**
- **et les probabilités de transition $P_k(x, dx')$, pour tout $k = 1, \cdots, n$.**

On pose

$$M_0 = W_0^{\rm imp}(X_0) \qquad \text{et} \qquad M_k = W_k^{\rm imp}(X_{k-1}, X_k)\, M_{k-1}\ ,$$

pour tout $k = 1, \cdots, n$, ce qui fait de la suite $\{M_k\,,\, k = 0, 1, \cdots, n\}$ une fonctionnelle multiplicative associée à la chaîne de Markov $\{X_k\,,\, k = 0, 1, \cdots, n\}$.

(xiii) **En déduire une nouvelle représentation**

$$\langle \gamma_n^{\rm e}, F \rangle = \mathbb{E}[F(X_n, M_n)\; \prod_{k=0}^n W_k^{\rm red}(X_{k-1}, X_k)\,]\ ,$$

de la distribution non–normalisée $\gamma_n^{\rm e}$, où conjointement $\{(X_k, M_k)\,,\, k = 0, 1, \cdots, n\}$ forme une nouvelle chaîne de Markov à valeurs dans l'ensemble produit $E \times [0, \infty)$, caractérisée par

- **la distribution de probabilité initiale $p_0^{\rm e}(dx, dv) = p_0(dx)\, \delta_{W_0^{\rm imp}(x)}(dv)$,**
- **et les probabilités de transition**

$$P_k^{\rm e}(x, v, dx', dv') = P_k(x, dx')\; \delta_{v\, W_k^{\rm imp}(x, x')}(dv')\ ,$$

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

On recherche une approximation particulaire pondérée de la forme

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

où (ξ_k^i, v_k^i) est une particule dans l'ensemble produit $E \times [0,\infty)$ pour tout $i = 1, \cdots, N$.

(xiv) **Décrire, en utilisant les résultats du cours, l'approximation particulaire de type bootstrap pour la distribution μ_k^{e} définie sur l'ensemble produit $E \times [0,\infty)$. En déduire, en utilisant les expressions obtenues en réponse à la question (viii), une approximation particulaire de type bootstrap pour les deux distributions μ_k^0 et μ_k.**