

Séminaire Méthodes Particulières pour l'Estimation
et la Commande Optimale Stochastique
EDF R&D, Clamart, 23 mars 2004

Méthodes de Monte Carlo avec interaction pour l'inférence statistique des modèles de Markov cachés

François Le Gland
IRISA / INRIA Rennes

legland@irisa.fr

<http://www.irisa.fr/sigma2/>

travail en commun avec Natacha Caylus, Arnaud Guyader et Nadia Oudjane

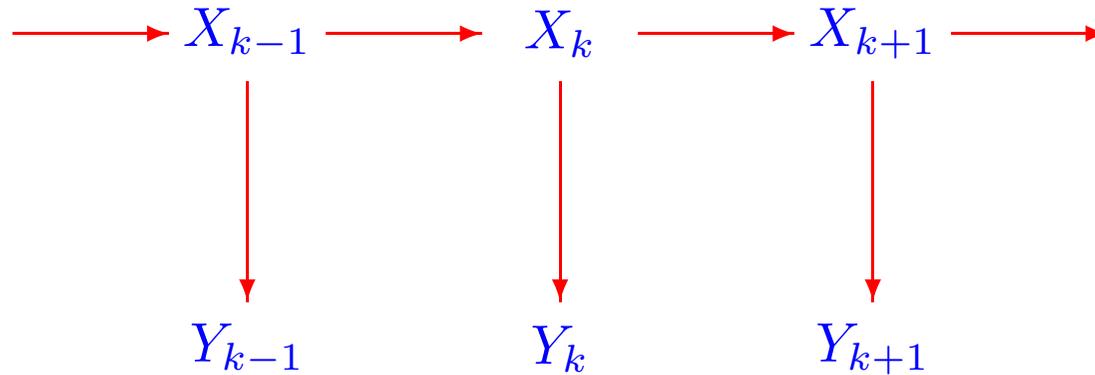
Plan

- modèles de Markov cachés (HMM) généraux
- filtre optimal et fonctions de contraste
- filtre particulaire : introduction, démo
- filtre particulaire, bootstrap vs. SIR
approximation particulaire (ponctuelle) des fonctions de contraste
- échantillonnage pondéré
approximation Monte Carlo régulière dans les modèles dominés
- hypothèses de régularité sur le modèle de Markov caché
- filtre dérivé
- filtre particulaire régulier
approximation particulaire (régulière) des fonctions de contraste

Plan

- modèles de Markov cachés (HMM) généraux
- filtre optimal et fonctions de contraste
- filtre particulaire : introduction, démo
- filtre particulaire, bootstrap vs. SIR
approximation particulaire (ponctuelle) des fonctions de contraste
- échantillonnage pondéré
approximation Monte Carlo régulière dans les modèles dominés
- hypothèses de régularité sur le modèle de Markov caché
- filtre dérivé
- filtre particulaire régulier
approximation particulaire (régulière) des fonctions de contraste

modèle de Markov caché (HMM)



états cachés : chaîne de Markov, noyau de transition

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = Q_k(x, dx')$$

espace d'état E

- fini, dénombrable
- espace euclidien, variété
- hybride continu / discret
- avec contraintes
- dépendant du temps
- trajectorien (croissant avec le temps)

$$\Xi_k = (X_0 \cdots X_k)$$

il suffit de savoir *simuler* une v.a. X_k de loi $Q_k(x, dx')$

exemple : équation différentielle stochastique (EDS)

$$dX'_t = b(X'_t) dt + \sigma(X'_t) dW'_t$$

échantillonnée aux instants $t_0 < \dots < t_k < \dots$

suite $X_k = X'_{t_k}$ chaîne de Markov

pas d'expression explicite pour le noyau de transition

$$Q_k(x, dx') = \mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = \mathbb{P}[X'_{t_k} \in dx' \mid X'_{t_{k-1}} = x]$$

mais facile de *simuler* la v.a. X'_{t_k} sachant $X'_{t_{k-1}} = x$:

discrétisation numérique de l'EDS entre les instants t_{k-1} et t_k

avec la condition initiale $X'_{t_{k-1}} = x$

observations : conditionnellement indépendantes sachant les états cachés

$$\mathbb{P}[Y_k \in dy' \mid X_k = x'] = g_k(x', y') \lambda_k^F(dy')$$

vraisemblance

$$\Psi_k(x') = g_k(x', Y_k)$$

adéquation entre l'état caché x' et l'observation Y_k

exemple : bruit blanc additif

$$Y_k = h_k(X_k) + V_k \quad \text{où} \quad V_k \sim q_k(v) dv$$

alors

$$\mathbb{P}[Y_k \in dy' \mid X_k = x'] = q_k(y' - h_k(x')) dy'$$

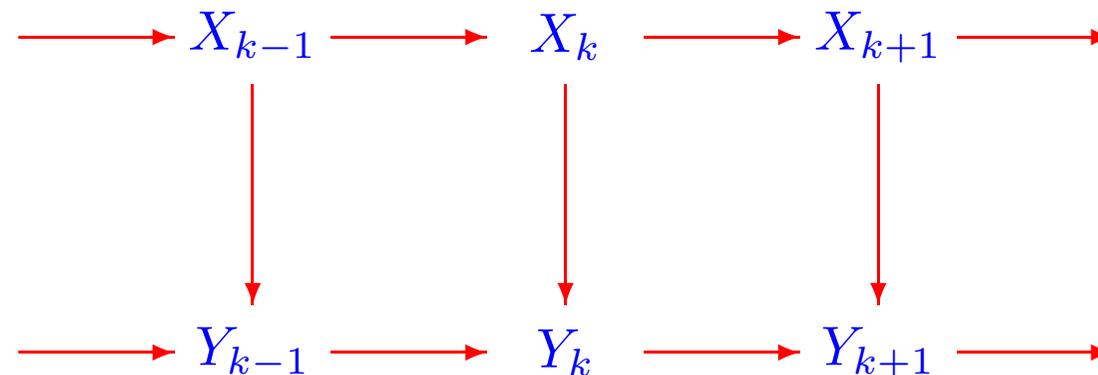
$$\Psi_k(x') = q_k(Y_k - h_k(x'))$$

et dans le cas de bruits gaussiens (variance identité)

$$\Psi_k(x') = \exp\left\{-\frac{1}{2} |Y_k - h_k(x')|^2\right\}$$

modèle : de + en + général

- états cachés : chaîne de Markov
observations : conditionnellement indépendantes sachant les états cachés
(v.a. indépendantes, dont les distributions dépendent des états cachés)
- états cachés : chaîne de Markov
observations : conditionnellement markoviennes sachant les états cachés
(chaîne de Markov, dont les transitions dépendent des états cachés)



- états cachés et observations : conjointement markoviens

hypothèse : de + en + faible

- savoir *simuler* une v.a. X_k sachant $X_{k-1} = x$
et *connaître explicitement / savoir calculer* la fonction de vraisemblance
(suppose un modèle dominé pour les distributions conditionnelles des observations)
- savoir *simuler* un couple de v.a. (X_k, Y_k) sachant $X_{k-1} = x$

exemples :

(i) bruit blanc additif, mais pas d'expression explicite pour la densité $q_k(v)$

$$Y_k = h_k(X_k) + V_k \quad \text{où} \quad V_k \sim q_k(v) dv$$

(ii) bruit blanc non-additif

$$Y_k = j_k(X_k, V_k)$$

(iii) observations parfaites

$$Y_k = h_k(X_k)$$

Plan

- modèles de Markov cachés (HMM) généraux
- filtre optimal et fonctions de contraste
- filtre particulaire : introduction, démo
- filtre particulaire, bootstrap vs. SIR
approximation particulaire (ponctuelle) des fonctions de contraste
- échantillonnage pondéré
approximation Monte Carlo régulière dans les modèles dominés
- hypothèses de régularité sur le modèle de Markov caché
- filtre dérivé
- filtre particulaire régulier
approximation particulaire (régulière) des fonctions de contraste

objectif : au vu des observations $Y_{0:k} = (Y_0 \cdots Y_k)$

- estimer (récursivement) l'état caché X_k
- identifier le modèle (paramètre inconnu θ dans le noyau markovien $Q_k^\theta(x, dx')$ et / ou dans la vraisemblance $\Psi_k^\theta(x')$)

filtre / prédicteur optimal (un pour chaque valeur du paramètre)

distribution de probabilité conditionnelle de l'état caché X_k

sachant les observations $Y_{0:k} = (Y_0 \cdots Y_k)$

$$\mu_k^\theta(dx) = \mathbb{P}^\theta[X_k \in dx \mid Y_{0:k}]$$

ou sachant les observations $Y_{0:k-1} = (Y_0 \cdots Y_{k-1})$

$$\mu_{k|k-1}^\theta(dx) = \mathbb{P}^\theta[X_k \in dx \mid Y_{0:k-1}]$$

approche bayésienne : loi *a priori* sur le paramètre θ

- MCMC : simulation d'un N -échantillon de la loi conditionnelle jointe des états $(X_0 \cdots X_n)$ et du paramètre θ sachant les observations $(Y_0 \cdots Y_n)$, puis marginalisation
- filtrage : chaîne de Markov augmentée : état X_k et paramètre θ
calcul récursif de la loi conditionnelle jointe de l'état (X_k, θ) sachant les observations $(Y_0 \cdots Y_n)$, puis marginalisation
mise en œuvre avec une approximation particulière

permet d'obtenir la loi *a posteriori* du paramètre θ ,
et pas seulement un estimateur ponctuel $\hat{\theta}$

au prix d'une augmentation de l'état

loi jointe des observations $Y_{0:n} = (Y_0 \cdots Y_n)$

$$\mathbb{P}^\theta[Y_{0:n} \in dy_{0:n}] = \prod_{k=0}^n \underbrace{\mathbb{P}^\theta[Y_k \in dy_k \mid Y_{0:k-1} = y_{0:k-1}]}_{p_k^\theta(y_{0:k}) \lambda_k^F(dy_k)}$$

avec

$$p_k^\theta(y_{0:k}) = \int_E g_k^\theta(x', y_k) \mathbb{P}^\theta[X_k \in dx' \mid Y_{0:k-1} = y_{0:k-1}]$$

fonction de log-vraisemblance pour l'estimation du paramètre θ

$$\begin{aligned} \ell_n^\theta &= \log \left[\prod_{k=0}^n p_k^\theta(Y_{0:k}) \right] = \sum_{k=0}^n \log p_k^\theta(Y_{0:k}) \\ &= \sum_{k=0}^n \log \int_E \underbrace{g_k^\theta(x', Y_k)}_{\Psi_k^\theta(x')} \underbrace{\mathbb{P}^\theta[X_k \in dx' \mid Y_{0:k-1}]}_{\mu_{k|k-1}^\theta(dx')} = \sum_{k=0}^n \log \langle \mu_{k|k-1}^\theta, \Psi_k^\theta \rangle \end{aligned}$$

équation d'évolution pour le filtre optimal

$$\mu_{k-1}^\theta \xrightarrow{\text{prédiction}} \mu_{k|k-1}^\theta = \mu_{k-1}^\theta Q_k^\theta \xrightarrow{\text{correction}} \mu_k^\theta = \Psi_k^\theta \cdot \mu_{k|k-1}^\theta$$

ou bien, en une seule étape

$$\mu_k^\theta = \bar{R}_k^\theta(\mu_{k-1}^\theta) \quad \text{avec} \quad \bar{R}_k^\theta(\mu) = \frac{\mu R_k^\theta}{\langle \mu R_k^\theta, 1 \rangle}$$

et avec le noyau non-normalisé

$$R_k^\theta(x, dx') = Q_k^\theta(x, dx') \Psi_k^\theta(x')$$

fonction de log-vraisemblance

$$\ell_n^\theta = \sum_{k=0}^n \log \langle \mu_{k|k-1}^\theta, \Psi_k^\theta \rangle = \sum_{k=0}^n \log \langle \mu_{k-1}^\theta Q_k^\theta, \Psi_k^\theta \rangle = \sum_{k=0}^n \log \langle \mu_{k-1}^\theta R_k^\theta, 1 \rangle$$

Plan

- modèles de Markov cachés (HMM) généraux
- filtre optimal et fonctions de contraste
- **filtre particulaire : introduction, démo**
- filtre particulaire, bootstrap vs. SIR
approximation particulaire (ponctuelle) des fonctions de contraste
- échantillonnage pondéré
approximation Monte Carlo régulière dans les modèles dominés
- hypothèses de régularité sur le modèle de Markov caché
- filtre dérivé
- filtre particulaire régulier
approximation particulaire (régulière) des fonctions de contraste

filtrage particulaire : calcul numérique approché du filtre optimal, à l'aide d'un système de particules en interaction

les particules

- *explorent* l'espace d'état, en évoluant de manière indépendante comme le processus sous-jacent, i.e. selon le noyau markovien $Q_k(x, dx')$
- et *interagissent* sous l'effet d'un mécanisme de *sélection*, lié à l'adéquation de chaque particule avec l'observation courante, i.e. en fonction de la vraisemblance $\Psi_k(x')$

la combinaison des deux mécanismes d'*exploration* et de *sélection* permet de concentrer automatiquement les particules (i.e. la *puissance de calcul*) dans les régions d'intérêt de l'espace d'état

approximation particulière

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N \omega_k^i \delta_{\xi_k^i} \quad \text{avec} \quad \sum_{i=1}^N \omega_k^i = 1$$

distribution empirique pondérée associée à un N -échantillon, ou système de particules, caractérisé par l'ensemble

$$\Sigma_k = ((\xi_k^1, \omega_k^1) \cdots (\xi_k^N, \omega_k^N))$$

des positions et des poids des particules

algorithme décrit par le mécanisme qui permet de construire Σ_k à partir de Σ_{k-1}

repose sur la notion d'échantillonnage pondéré (présenté à partir de #23)

démo, © Fabien Campillo

<http://www.irisa.fr/sigma2/campillo/site-pf/contents/matlab/>

Plan

- modèles de Markov cachés (HMM) généraux
- filtre optimal et fonctions de contraste
- filtre particulaire : introduction, démo
- **filtre particulaire, bootstrap vs. SIR**
approximation particulaire (ponctuelle) des fonctions de contraste
- échantillonnage pondéré
approximation Monte Carlo régulière dans les modèles dominés
- hypothèses de régularité sur le modèle de Markov caché
- filtre dérivé
- filtre particulaire régulier
approximation particulaire (régulière) des fonctions de contraste

algorithme de base : filtre bootstrap (Gordon, Salmond and Smith, IEE Proceedings, Part F, 1993)

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N \omega_k^i \delta_{\xi_k^i} \quad \text{avec} \quad \sum_{i=1}^N \omega_k^i = 1$$

- sélection des particules de plus forts poids : indépendamment pour $i = 1 \dots N$

$$\tau_{k-1}^i \sim (\omega_{k-1}^1 \dots \omega_{k-1}^N) \quad \text{à valeurs dans les indices } 1 \dots N$$

- mutation selon le noyau markovien : indépendamment pour $i = 1 \dots N$

$$\xi_k^i \sim Q_k(\xi_{k-1}^{\tau_{k-1}^i}, dx')$$

- pondération selon la vraisemblance : pour $i = 1 \dots N$

$$\omega_k^i \propto \Psi_k(\xi_k^i)$$

généralisation : décomposition d'importance du noyau non-normalisé

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

avec

- un noyau de mutation $P_k(x, dx')$ markovien
a.c. par rapport à $R_k(x, dx')$ i.e. si $R_k(x, A) > 0$, alors $P_k(x, A) > 0$
- une fonction de sélection $W_k(x, x')$
dépendant de la transition (x, x') et pas seulement de l'état x'

et des poids de sélection

$$\omega_{k-1}^i = \frac{\omega_{k-1}^i}{\pi_k^i} \pi_k^i$$

avec

- une distribution discrète $(\pi_k^1 \cdots \pi_k^N)$
a.c. par rapport à $(\omega_{k-1}^1 \cdots \omega_{k-1}^N)$ i.e. si $\omega_{k-1}^i > 0$, alors $\pi_k^i > 0$

algorithme plus général : filtre SIR (sampling / importance resampling)

- sélection des particules de plus forts poids : indépendamment pour $i = 1 \dots N$

$$\tau_k^i \sim (\pi_k^1 \dots \pi_k^N) \quad \text{à valeurs dans les indices } 1 \dots N$$

- mutation selon le noyau d'importance : indépendamment pour $i = 1 \dots N$

$$\xi_k^i \sim P_k(\xi_{k-1}^{\tau_k^i}, dx')$$

- pondération selon le rapport d'importance : pour $i = 1 \dots N$

$$\omega_k^i \propto \frac{\omega_{k-1}^{\tau_k^i}}{\pi_k^{\tau_k^i}} W_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i)$$

utilisation des poids discrets $(\omega_{k-1}^1 \dots \omega_{k-1}^N)$ pour sélection et / ou pondération

approximation particulière de la fonction de log-vraisemblance

$$l_n^\theta = \sum_{k=0}^n \langle \mu_k^\theta |_{k-1}, \Psi_k^\theta \rangle = \sum_{k=0}^n \log \langle \mu_{k-1}^\theta | R_k^\theta, 1 \rangle$$

avec le filtre bootstrap

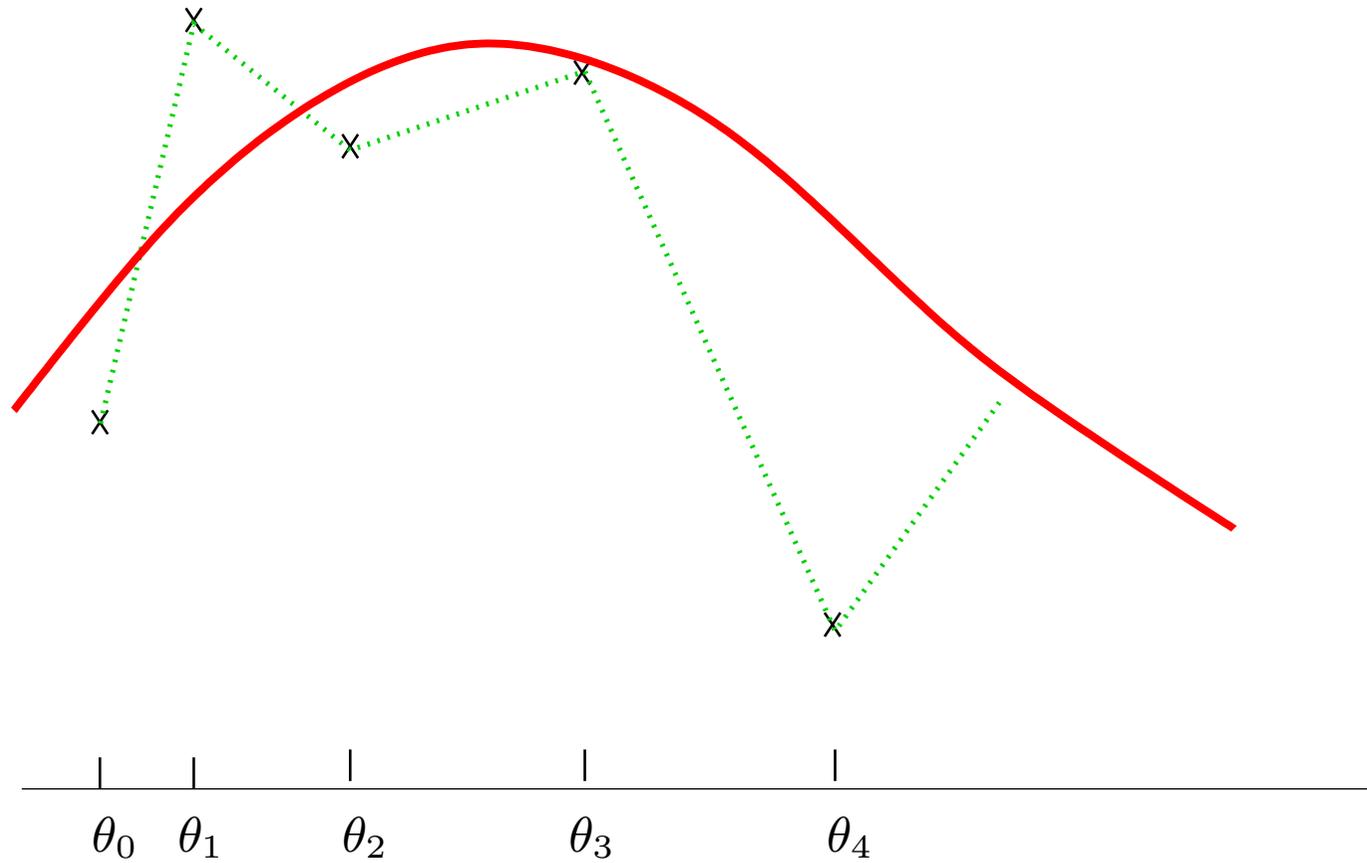
$$l_n^\theta \approx \sum_{k=0}^n \log \left[\sum_{i=1}^N \Psi_k(\xi_k^i) \right]$$

avec le filtre SIR

$$l_n^\theta \approx \sum_{k=0}^n \log \left[\sum_{i=1}^N \frac{\omega_{k-1}^{\tau_k^i}}{\pi_k^{\tau_k^i}} W_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i) \right]$$

les systèmes de particules Σ_k dépendent du paramètre θ

→ approximation Monte Carlo ponctuelle, irrégulière, de la fonction de log-vraisemblance



aller directement au #23

quelques cas particuliers (choix des noyaux de mutation
et des poids de sélection)

► filtre bootstrap

$$P_k(x, dx') = Q_k(x, dx') \quad W_k(x, x') = \Psi_k(x') \quad \pi_k^i = \omega_{k-1}^i$$

d'où le nouveau poids

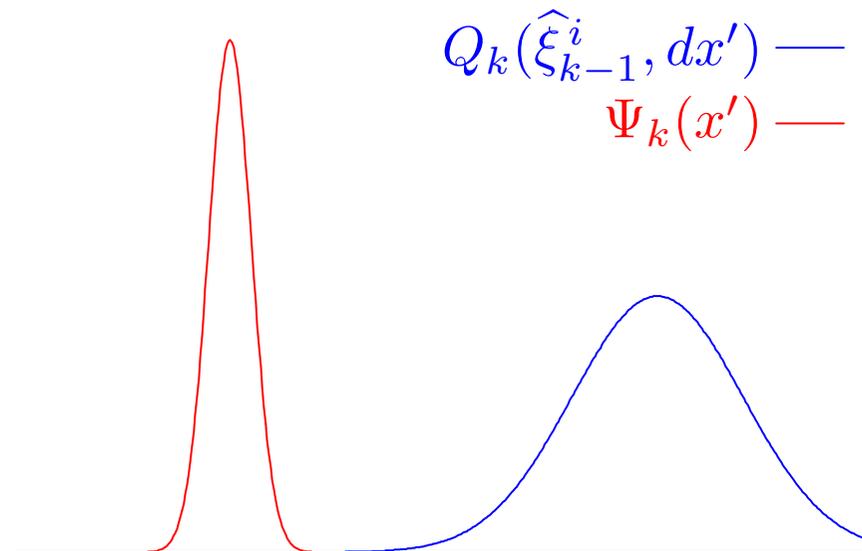
$$\omega_k^i \propto \Psi_k(\xi_k^i)$$

pour tout $i = 1 \cdots N$

simuler des particules en aveugle peut entraîner une *dégénérescence des poids*, e.g. en cas de mauvais recouvrement entre distributions de probabilité a priori et fonction de vraisemblance, càd si

$$Q_k \Psi_k(\hat{\xi}_{k-1}^i) = \int_E Q_k(\hat{\xi}_{k-1}^i, dx') \Psi_k(x') \approx 0$$

pour la plupart des $i = 1 \dots N$



► mutation guidée par les observations

$$P_k(x, dx') = \widehat{Q}_k(x, dx') = \frac{Q_k(x, dx') \Psi_k(x')}{Q_k \Psi_k(x)}$$

$$W_k(x, x') = \widehat{\Psi}_k(x) = Q_k \Psi_k(x)$$

dépend de l'observation future Y_k et ne dépend pas de x' , donc peut être utilisé dès l'étape de sélection

$$\pi_k^i = \widehat{\Psi}_k(\xi_{k-1}^i) / \left[\sum_{j=1}^N \widehat{\Psi}_k(\xi_{k-1}^j) \right]$$

d'où le nouveau poids

$$\omega_k^i \propto \omega_{k-1}^{\tau_k^i}$$

pour tout $i = 1 \cdots N$

bonne idée, mais il n'est en général pas facile

- de simuler des v.a. selon le noyau $\hat{Q}_k(x, dx')$,
- ni de calculer la fonction de sélection $\hat{\Psi}_k(x)$

sauf dans certains cas particuliers, e.g. bruits mélanges de v.a. gaussiennes et fonction d'observation linéaire

► particules auxiliaires (Pitt et Shephard, 1999)

$$P_k(x, dx') = Q_k(x, dx') \quad W_k(x, x') = \Psi_k(x') \quad \pi_k^i \propto \Psi_k^i$$

où Ψ_k^i est une approximation de la mesure de recouvrement

$$Q_k \Psi_k(\xi_{k-1}^i) = \int_E Q_k(\xi_{k-1}^i, dx') \Psi_k(x') \approx \Psi_k^i$$

d'où le nouveau poids

$$\omega_k^i \propto \omega_{k-1}^{\tau_k^i} \frac{\Psi_k(\xi_k^i)}{\Psi_k^{\tau_k^i}}$$

pour tout $i = 1 \dots N$

Plan

- modèles de Markov cachés (HMM) généraux
- filtre optimal et fonctions de contraste
- filtre particulaire : introduction, démo
- filtre particulaire, bootstrap vs. SIR
approximation particulaire (ponctuelle) des fonctions de contraste
- **échantillonnage pondéré**
approximation Monte Carlo régulière dans les modèles dominés
- hypothèses de régularité sur le modèle de Markov caché
- filtre dérivé
- filtre particulaire régulier
approximation particulaire (régulière) des fonctions de contraste

approximation Monte Carlo

$$\mu(dx) = \frac{r(x) m(dx)}{\int_E r(x') m(dx')} \propto r(x) m(dx)$$

densité $r(x)$ connue à une constante multiplicative près
s'il est possible de simuler directement selon $\mu(dx)$, alors

$$\langle \mu, \phi \rangle = \int_E \phi(x) \mu(dx) \approx \frac{1}{N} \sum_{i=1}^N \phi(x_i)$$

i.e.

$$\mu \approx \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \quad \text{avec} \quad x_i \sim \mu(dx)$$

indépendamment pour tout $i = 1 \dots N$

échantillonnage pondéré : s'il est plus facile de simuler selon $m(dx)$, alors

$$\langle \mu, \phi \rangle = \int_E \phi(x) \mu(dx) = \frac{\int_E \phi(x) r(x) m(dx)}{\int_E r(x) m(dx)} \approx \frac{\sum_{i=1}^N \phi(\xi_i) r(\xi_i)}{\sum_{i=1}^N r(\xi_i)}$$

i.e.

$$\mu \approx \sum_{i=1}^N \frac{r(\xi_i)}{\sum_{j=1}^N r(\xi_j)} \delta_{\xi_i} \quad \text{avec} \quad \xi_i \sim m(dx)$$

indépendamment pour tout $i = 1 \dots N$

application : approximation Monte Carlo régulière dans un modèle dominé

$$\mu^\theta(dx) \propto r^\theta(x) m(dx)$$

pour chaque valeur du paramètre θ

$$\mu^\theta \approx \frac{1}{N} \sum_{i=1}^N \delta_{x_i^\theta} \quad \text{avec} \quad x_i^\theta \sim \mu^\theta(dx)$$

v.s.

$$\mu^\theta \approx \sum_{i=1}^N \frac{r^\theta(\xi_i)}{\sum_{j=1}^N r^\theta(\xi_j)} \delta_{\xi_i} \quad \text{avec} \quad \xi_i \sim m(dx)$$

indépendamment pour tout $i = 1 \dots N$

si $\theta \mapsto r^\theta(x)$ est régulière (continue, dérivable, etc.), alors

$$\theta \mapsto \mu^\theta(dx) = \frac{r^\theta(x) m(dx)}{\int_E r^\theta(x') m(dx')} \quad \text{et} \quad \theta \mapsto \mu_N^\theta = \sum_{i=1}^N \frac{r^\theta(\xi_i)}{\sum_{j=1}^N r^\theta(\xi_j)} \delta_{\xi_i}$$

sont régulières, avec pour dérivées les mesures signées de masse nulle

$$\frac{\partial \mu^\theta}{\partial \theta}(dx) = \frac{\frac{\partial r^\theta}{\partial \theta}(x) m(dx)}{\int_E r^\theta(x') m(dx')} - a^\theta \mu^\theta(dx)$$

et

$$\frac{\partial \mu_N^\theta}{\partial \theta} = \sum_{i=1}^N \frac{\frac{\partial r^\theta}{\partial \theta}(\xi_i)}{\sum_{j=1}^N r^\theta(\xi_j)} \delta_{\xi_i} - a_N^\theta \mu_N^\theta$$

respectivement (a^θ et a_N^θ constantes de normalisation)

Plan

- modèles de Markov cachés (HMM) généraux
- filtre optimal et fonctions de contraste
- filtre particulaire : introduction, démo
- filtre particulaire, bootstrap vs. SIR
approximation particulaire (ponctuelle) des fonctions de contraste
- échantillonnage pondéré
approximation Monte Carlo régulière dans les modèles dominés
- hypothèses de régularité sur le modèle de Markov caché
- filtre dérivé
- filtre particulaire régulier
approximation particulaire (régulière) des fonctions de contraste

noyau markovien $Q_k^\theta(x, dx')$ paramétré par θ dans un voisinage de θ_0 ,
avec $Q_k^{\theta_0}(x, dx') = Q_k(x, dx')$ pour $\theta = \theta_0$

Hypothèse AC₀ : représentation probabiliste

$$Q_k^\theta \phi(x) = \mathbb{E}^\theta[\phi(X_k) \mid X_{k-1} = x] = \mathbb{E}[\phi(X_k) \Lambda_k^\theta \mid X_{k-1} = x]$$

où

$$\mathbb{P}[X_k \in dx', \Lambda_k^\theta \in d\lambda' \mid X_{k-1} = x] = L_k^\theta(x, dx', d\lambda')$$

et facile de **simuler** une v.a. (X_k, Λ_k^θ) de loi $L_k^\theta(x, dx', d\lambda')$

exemple : EDS échantillonnée, dépendant d'un paramètre

$$dX'_t = b^\theta(X'_t) dt + \sigma(X'_t) dW'_t{}^\theta \quad \text{sous } \mathbb{P}^\theta \quad (1)$$

$$dX'_t = b(X'_t) dt + \sigma(X'_t) dW'_t \quad \text{sous } \mathbb{P} = \mathbb{P}^{\theta_0} \quad (2)$$

dans le cas simple où $\sigma(x) \equiv I$, l'Hypothèse AC_0 est vérifiée avec

$$\Lambda_k^\theta = \exp\left\{ \int_{t_{k-1}}^{t_k} [b^\theta(X'_s) - b(X'_s)]^* dW'_s - \frac{1}{2} \int_{t_{k-1}}^{t_k} |b^\theta(X'_s) - b(X'_s)|^2 ds \right\}$$

facile de *simuler* sous $\mathbb{P} = \mathbb{P}^{\theta_0}$ la v.a. $(X'_{t_k}, \Lambda_k^\theta)$ sachant $X'_{t_{k-1}} = x$:

discrétisation numérique de l'EDS (2) entre les instants t_{k-1} et t_k

avec la condition initiale $X'_{t_{k-1}} = x$

et approximation numérique de l'intégrale stochastique

sous l'Hypothèse AC_0 , nécessairement

$$Q_k^\theta(x, dx') = J_k^\theta(x, x') Q_k(x, dx')$$

avec

$$J_k^\theta(x, x') = \mathbb{E}[\Lambda_k^\theta \mid X_{k-1} = x, X_k = x']$$

inclut les modèles dominés de chaînes de Markov

on fait ici l'hypothèse plus forte

$$Q_k^\theta(x, dx') = J_k^\theta(x, x') Q_k(x, dx')$$

avec une expression *explicite* pour $J_k^\theta(x, x')$

Hypothèse AC : représentation probabiliste pour le noyau dérivé

$$\Gamma_k^\theta \phi(x) = \frac{\partial}{\partial \theta} Q_k^\theta \phi(x) = \mathbb{E}^\theta[\phi(X_k) \Xi_k^\theta \mid X_{k-1} = x]$$

où

$$\mathbb{P}^\theta[X_k \in dx', \Xi_k^\theta \in ds' \mid X_{k-1} = x] = K_k^\theta(x, dx', ds')$$

et facile de *simuler* une v.a. (X_k, Ξ_k^θ) de loi $K_k^\theta(x, dx', d\lambda')$

sous l'Hypothèse AC₀, et si $\theta \longmapsto \Lambda_k^\theta$ est dérivable, alors

$$\Gamma_k^\theta \phi(x) = \mathbb{E}[\phi(X_k) \frac{\partial \log \Lambda_k^\theta}{\partial \theta} \Lambda_k^\theta \mid X_{k-1} = x] = \mathbb{E}^\theta[\phi(X_k) \frac{\partial \log \Lambda_k^\theta}{\partial \theta} \mid X_{k-1} = x]$$

i.e. l'Hypothèse AC est vérifiée, avec $\Xi_k^\theta = \frac{\partial \log \Lambda_k^\theta}{\partial \theta}$

exemple : EDS échantillonnée, dépendant d'un paramètre (suite)
 dans le cas simple où $\sigma(x) \equiv I$, l'Hypothèse AC est vérifiée avec

$$\begin{aligned}\Xi_k^\theta &= \int_{t_{k-1}}^{t_k} \left[\frac{\partial b^\theta}{\partial \theta}(X'_s) \right]^* dW'_s - \int_{t_{k-1}}^{t_k} \left[\frac{\partial b^\theta}{\partial \theta}(X'_s) \right]^* [b^\theta(X'_s) - b(X'_s)] ds \\ &= \int_{t_{k-1}}^{t_k} \left[\frac{\partial b^\theta}{\partial \theta}(X'_s) \right]^* dW'_s\end{aligned}$$

facile de *simuler* sous \mathbb{P}^θ la v.a. (X'_{t_k}, Ξ_k^θ) sachant $X'_{t_{k-1}} = x$:

discrétisation numérique de l'EDS (1) entre les instants t_{k-1} et t_k

avec la condition initiale $X'_{t_{k-1}} = x$

et approximation numérique de la deuxième intégrale stochastique

sous l'Hypothèse AC, nécessairement

$$\Gamma_k^\theta(x, dx') = I_k^\theta(x, x') Q_k^\theta(x, dx')$$

avec

$$I_k^\theta(x, x') = \mathbb{E}^\theta[\Xi_k^\theta \mid X_{k-1} = x, X_k = x']$$

on fait ici l'hypothèse plus forte qu'il existe une expression *explicite* pour $I_k^\theta(x, x')$
 sous l'Hypothèse AC₀, on a vu que nécessairement

$$Q_k^\theta(x, dx') = J_k^\theta(x, x') Q_k(x, dx')$$

de sorte que

$$\Gamma_k^\theta(x, dx') = \frac{\partial \log J_k^\theta}{\partial \theta}(x, x') \underbrace{J_k^\theta(x, x') Q_k(x, dx')}_{Q_k^\theta(x, dx')}$$

i.e.

$$I_k^\theta(x, x') = \frac{\partial \log J_k^\theta}{\partial \theta}(x, x')$$

Plan

- modèles de Markov cachés (HMM) généraux
- filtre optimal et fonctions de contraste
- filtre particulaire : introduction, démo
- filtre particulaire, bootstrap vs. SIR
approximation particulaire (ponctuelle) des fonctions de contraste
- échantillonnage pondéré
approximation Monte Carlo régulière dans les modèles dominés
- hypothèses de régularité sur le modèle de Markov caché
- **filtre dérivé**
- filtre particulaire régulier
approximation particulaire (régulière) des fonctions de contraste

rappel : équation d'évolution pour le filtre optimal

$$\mu_{k-1}^\theta \xrightarrow{\text{prédiction}} \mu_{k|k-1}^\theta = \mu_{k-1}^\theta Q_k^\theta \xrightarrow{\text{correction}} \mu_k^\theta = \frac{\Psi_k^\theta \mu_{k|k-1}^\theta}{\langle \mu_{k|k-1}^\theta, \Psi_k^\theta \rangle}$$

équation d'évolution pour le filtre dérivé

$$w_{k-1}^\theta \xrightarrow{\quad} w_{k|k-1}^\theta = w_{k-1}^\theta Q_k^\theta + \mu_{k-1}^\theta \Gamma_k^\theta$$

$$\xrightarrow{\quad} w_k^\theta = \frac{\Psi_k^\theta (w_{k|k-1}^\theta + S_k^\theta \mu_{k|k-1}^\theta)}{\langle \mu_{k|k-1}^\theta, \Psi_k^\theta \rangle} - a_k^\theta \mu_k^\theta$$

avec

$$S_k^\theta(x') = \frac{\partial \log \Psi_k^\theta(x')}{\partial \theta}$$

ou bien, en une seule étape

$$w_k^\theta = \bar{W}_k^\theta(\mu_{k-1}^\theta, w_{k-1}^\theta) = \frac{\mu_{k-1}^\theta T_k^\theta + w_{k-1}^\theta R_k^\theta}{\langle \mu_{k-1}^\theta R_k^\theta, 1 \rangle} - a_k^\theta \mu_k^\theta$$

avec

$$\mu T_k^\theta(dx') = \int_E \mu(dx) [I_k^\theta(x, x') + S_k^\theta(x')] R_k^\theta(x, dx')$$

$$w R_k^\theta(dx') = \int_E w(dx) R_k^\theta(x, dx')$$

quel schéma d'approximation particulière pour $w_k^\theta = \frac{\partial \mu_k^\theta}{\partial \theta}$?

Plan

- modèles de Markov cachés (HMM) généraux
- filtre optimal et fonctions de contraste
- filtre particulaire : introduction, démo
- filtre particulaire, bootstrap vs. SIR
approximation particulaire (ponctuelle) des fonctions de contraste
- échantillonnage pondéré
approximation Monte Carlo régulière dans les modèles dominés
- hypothèses de régularité sur le modèle de Markov caché
- filtre dérivé
- **filtre particulaire régulier**
approximation particulaire (régulière) des fonctions de contraste

approximation particulaire régulière

$$\mu_k^\theta \approx \mu_k^{N,\theta} = \sum_{i=1}^N u_k^{i,\theta} \omega_k^i \delta_{\xi_k^i} \quad \text{avec} \quad \sum_{i=1}^N u_k^{i,\theta} \omega_k^i = 1$$

pour tout θ , et avec

$$u_k^{i,\theta_0} = \dots = u_k^{i,\theta_0} = 1 \quad \text{pour } \theta = \theta_0$$

distribution empirique pondérée associée à un unique N -échantillon, caractérisé par l'ensemble (ne dépendant que de θ_0 , et pas de θ)

$$\Sigma_k = ((\xi_k^1, \omega_k^1) \cdots (\xi_k^N, \omega_k^N))$$

des positions et des poids des particules

et par un ensemble de poids supplémentaires $(u_k^{1,\theta} \cdots u_k^{N,\theta})$ dépendant de θ

filtre SIR régulier

version avec interaction de l'approche MCML (Monte Carlo maximum likelihood)
(Geyer, JRSS, Series B, 1994)

sous l'Hypothèse AC_0 , et pour une décomposition d'importance donnée, on a

$$\begin{aligned}
 R_k^\theta(x, dx') &= Q_k^\theta(x, dx') \Psi_k^\theta(x') \\
 &= \underbrace{J_k^\theta(x, x') \frac{\Psi_k^\theta(x')}{\Psi_k(x')}}_{r_k^\theta(x, x')} \underbrace{Q_k(x, dx') \Psi_k(x')}_{R_k(x, dx')} = P_k(x, dx') \underbrace{r_k^\theta(x, x') W_k(x, x')}_{W_k^\theta(x, x')}
 \end{aligned}$$

- sélection des particules de plus forts poids : indépendamment pour $i = 1 \dots N$

$$\tau_k^i \sim (\pi_k^1 \dots \pi_k^N) \quad \text{à valeurs dans les indices } 1 \dots N$$

- mutation selon le noyau d'importance : indépendamment pour $i = 1 \dots N$

$$\xi_k^i \sim P_k(\xi_{k-1}^{\tau_k^i}, dx')$$

- pondération selon le rapport d'importance : pour $i = 1 \dots N$

$$\omega_k^i \propto \frac{\omega_{k-1}^{\tau_k^i}}{\pi_k^{\tau_k^i}} W_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i)$$

et pour chaque valeur θ du paramètre

$$u_k^{i,\theta} = \frac{u_{k-1}^{\tau_k^i, \theta} r_k^\theta(\xi_{k-1}^{\tau_k^i}, \xi_k^i)}{\sum_{i=1}^N \omega_k^i u_{k-1}^{\tau_k^i, \theta} r_k^\theta(\xi_{k-1}^{\tau_k^i}, \xi_k^i)}$$

si en outre l'application $\theta \longmapsto r_k^\theta(x, x')$ est dérivable, alors l'approximation particulière $\mu_k^{N, \theta}$ est aussi dérivable : en effet

$$\frac{\partial \mu_k^{N, \theta}}{\partial \theta} = \sum_{i=1}^N \frac{\partial u_k^{i, \theta}}{\partial \theta} \omega_k^i \delta_{\xi_k^i}$$

compte tenu que

$$\theta \longmapsto u_k^{i, \theta} = \frac{u_{k-1}^{\tau_k^i, \theta} r_k^\theta(\xi_{k-1}^{\tau_k^i}, \xi_k^i)}{\sum_{i=1}^N \omega_k^i u_{k-1}^{\tau_k^i, \theta} r_k^\theta(\xi_{k-1}^{\tau_k^i}, \xi_k^i)}$$

est dérivable, avec pour dérivée

$$\frac{\partial u_k^{i, \theta}}{\partial \theta} = \frac{\frac{\partial}{\partial \theta} [u_{k-1}^{\tau_k^i, \theta} r_k^\theta(\xi_{k-1}^{\tau_k^i}, \xi_k^i)]}{\sum_{i=1}^N \omega_k^i u_{k-1}^{\tau_k^i, \theta} r_k^\theta(\xi_{k-1}^{\tau_k^i}, \xi_k^i)} - a_k^{N, \theta} u_k^{i, \theta}$$

où $a_k^{N, \theta}$ est déterminé de façon unique par la condition $\sum_{i=1}^N \omega_k^i \frac{\partial u_k^{i, \theta}}{\partial \theta} = 0$

en particulier

$$w_k^N = \frac{\partial \mu_k^{N,\theta}}{\partial \theta} \Big|_{\theta=\theta_0} = \sum_{i=1}^N \rho_k^i \omega_k^i \delta_{\xi_k^i}$$

avec

$$\rho_k^i = \frac{\partial u_k^{i,\theta}}{\partial \theta} \Big|_{\theta=\theta_0} = \rho_{k-1}^{\tau_k^i} + I_k^{\theta_0}(\xi_{k-1}^{\tau_k^i}, \xi_k^i) + S_k^{\theta_0}(\xi_k^i) - a_k^N$$

où a_k^N est déterminé de façon unique par la condition $\sum_{i=1}^N \rho_k^i \omega_k^i = 0$

- sélection des particules de plus forts poids : indépendamment pour $i = 1 \dots N$

$$\tau_k^i \sim (\pi_k^1 \dots \pi_k^N) \quad \text{à valeurs dans les indices } 1 \dots N$$

- mutation selon le noyau d'importance : indépendamment pour $i = 1 \dots N$

$$\xi_k^i \sim P_k(\xi_{k-1}^{\tau_k^i}, dx')$$

- pondération selon le rapport d'importance : pour $i = 1 \dots N$

$$\omega_k^i \propto \frac{\omega_{k-1}^{\tau_k^i}}{\pi_k^{\tau_k^i}} W_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i)$$

et

$$\rho_k^i = \rho_{k-1}^{\tau_k^i} + I_k^{\theta_0}(\xi_{k-1}^{\tau_k^i}, \xi_k^i) + S_k^{\theta_0}(\xi_k^i) - a_k^N$$

où a_k^N est déterminé de façon unique par la condition $\sum_{i=1}^N \rho_k^i \omega_k^i = 0$

approximation particulaire de la fonction de log-vraisemblance

$$\ell_n^\theta = \sum_{k=0}^n \langle \mu_k^\theta |_{k-1}, \Psi_k^\theta \rangle = \sum_{k=0}^n \log \langle \mu_{k-1}^\theta R_k^\theta, 1 \rangle$$

avec le filtre SIR régulier

$$\ell_n^\theta \approx \sum_{k=0}^n \log \left[\sum_{i=1}^N \frac{\omega_{k-1}^{\tau_k^i}}{\pi_k^{\tau_k^i}} W_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i) r_k^\theta(\xi_{k-1}^{\tau_k^i}, \xi_k^i) \right]$$

le système de particules Σ_k ne dépend que de θ_0 et pas de θ

→ approximation Monte Carlo globale, régulière, de la fonction de log-vraisemblance

approximation particulaire de la fonction score

