5th Workshop on
Asymptotical Statistics of Stochastic Processes
Le Mans, January 6–8, 2005

Smooth interacting Monte Carlo approximation of contrast functions in HMM

François Le Gland
IRISA / INRIA Rennes

legland@irisa.fr
http://www.irisa.fr/aspi/

joint work with Natacha Caylus (Université de Rennes 1)
Arnaud Guyader (Université de Haute Bretagne, Rennes)
and Nadia Oudjane (Électricité de France R&D, Clamart)

Plan

- contrast functions in HMM
- interacting particle approximation
- global interacting particle approximation
- smooth interacting particle approximation
- linear tangent filter
- linear tangent filter : interacting particle approximation
- conclusion

statistical model, depending on a parameter

- Markov chain $\{X_k \, , \, k \geq 0\}$ on E, with Markov transition kernels

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

and initial probability distribution

$$\mathbb{P}[X_0 \in dx] = \mu_0(dx)$$

- observations $\{Y_k \, , \, k \geq 0\}$ with values in F,
satisfying the memoryless channel assumption, with

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

and likelihood function

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

conditional probability distributions

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k}] \qquad \text{and} \qquad \mu_{k|k-1}(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k-1}]$$

going from μ_{k-1} to μ_k : decomposition into prediction / correction steps

$$\mu_{k-1} \xrightarrow{\text{prediction}} \mu_{k|k-1} = \mu_{k-1}\, Q_k \xrightarrow{\text{correction}} \mu_k = \Psi_k \cdot \mu_{k|k-1}$$

where $\cdot$ denotes projective product, i.e.

$$\mu_k = \Psi_k \cdot \mu_{k|k-1} = \frac{\Psi_k\, \mu_{k|k-1}}{\mu_{k|k-1}(\Psi_k)}$$

or in a single step, introducing $R_k(x, dx') = Q_k(x, dx')\ \Psi_k(x')$

$$\mu_{k-1} \longrightarrow \mu_k = \bar{R}_k(\mu_{k-1}) = \frac{\mu_{k-1}\, R_k}{(\mu_{k-1}\, R_k)(E)}$$

indeed

$$(\mu_{k-1}\, R_k)(dx') = \underbrace{\int_E \mu_{k-1}(dx)\ Q_k(x, dx')}_{\mu_{k|k-1}(dx')}\ \Psi_k(x')$$

and

$$(\mu_{k-1}\, R_k)(E) = \int_E \mu_{k|k-1}(dx')\ \Psi_k(x') = \mu_{k|k-1}(\Psi_k)$$

log–likelihood function

$$\ell_n = \sum_{k=1}^n \log \mu_{k|k-1}(\Psi_k) = \sum_{k=1}^n \log(\mu_{k-1}\, R_k)(E)$$

Plan

- contrast functions in HMM
- interacting particle approximation
- global interacting particle approximation
- smooth interacting particle approximation
- linear tangent filter
- linear tangent filter : interacting particle approximation
- conclusion

recall that

$$\mu_{k-1} \longrightarrow \mu_k = \bar{R}_k(\mu_{k-1}) = \frac{\mu_{k-1}\, R_k}{(\mu_{k-1}\, R_k)(E)}$$

approximation : weighted empirical distribution

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

associated with particle system $\Sigma_k = \{(\xi_k^i, w_k^i)\, ,\, i = 1 \cdots N\}$

arbitrary importance decomposition of nonnegative kernel into

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

with

- Markov mutation kernel $P_k(x, dx')$, easy to simulate from
- selection function $W_k(x, x')$, easy to evaluate

a natural importance decomposition is available

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

example : sampling the solution of an SDE

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

at discrete time instants $t_0 < \cdots < t_k < \cdots$, yields a Markov chain $X_k = X'_{t_k}$

no explicit expression is available in general for the Markov kernel

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

but it is easy to simulate the r.v. X_k given $X_{k-1} = x$:

use any discretization scheme of the SDE between time instants t_{k-1} and t_k, with initial condition $X'_{t_{k-1}} = x$ $\square$

starting from particle approximation

$$\mu_{k-1} \approx \mu_{k-1}^N = \sum_{i=1}^N w_{k-1}^i \, \delta_{\xi_{k-1}^i}$$

and applying nonnegative kernel $R_k(x, dx')$ exactly, yields

$$\begin{aligned}(\mu_{k-1}^N \, R_k)(dx') &= \sum_{i=1}^N w_{k-1}^i \; R_k(\xi_{k-1}^i, dx') \\ &= \sum_{i=1}^N \underbrace{\frac{w_{k-1}^i}{\pi_k^i} \; W_k(\xi_{k-1}^i, x')}_{w_k^i(x')} \; \underbrace{\pi_k^i \; P_k(\xi_{k-1}^i, dx')}_{m_k^i(dx')}\end{aligned}$$

with arbitrary importance decomposition of discrete weights into

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

auxiliary particle idea of Pitt and Shephard : sample product space $E \times \{1 \cdots N\}$

SIR (sampling / importance resampling) algorithm

going from Σ_{k-1} to Σ_k

- selection of particles with higher weights : independently for $i = 1 \cdots N$

$$\tau_k^i \sim (\pi_k^1 \cdots \pi_k^N) \qquad \text{with values in index set } \{1 \cdots N\}$$

- mutation using importance Markov kernel : independently for $i = 1 \cdots N$

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

- weighting according to importance weight : for $i = 1 \cdots N$

$$w_k^i = \frac{1}{c_k^N} \, \frac{w_{k-1}^{\tau_k^i}}{\pi_k^{\tau_k^i}} \, W_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i) \qquad \text{s.t.} \qquad \sum_{k=1}^{N} w_k^i = 1$$

discrete weights $(w_{k-1}^1 \cdots w_{k-1}^N)$ are used for selection and / or weighting particle approximation of log–likelihood function

$$\ell_n = \sum_{k=0}^{n} \log(\mu_{k-1}\, R_k)(E)$$

using SIR algorithm

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

alternatively, using simple bootstrap algorithm

$$\ell_n \approx \ell_n^N = \sum_{k=0}^{n} \log[\, \frac{1}{N} \sum_{i=1}^{N} \Psi_k(\xi_k^i)\,]$$

unfortunately, particle systems $\{\Sigma_k\, ,\, 0 \leq k \leq n\}$ depend on the parameter
$\longrightarrow$ pointwise, irregular, Monte Carlo approximation of log–likelihood function

Plan

- contrast functions in HMM
- interacting particle approximation
- global interacting particle approximation
- smooth interacting particle approximation
- linear tangent filter
- linear tangent filter : interacting particle approximation
- conclusion

define $R_k^0(x, dx')$, for some pivot (nominal) parameter value $\theta_0 \in \Theta$

Assumption AC$_0$

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

example : if

$$Q_k(x, dx') = q_k(x, x') \; Q_k^0(x, dx') \qquad (\star)$$

then

$$R_k(x, dx') = Q_k(x, dx') \; \Psi_k(x') = \underbrace{q_k(x, x') \, \frac{\Psi_k(x')}{\Psi_k^0(x')}}_{r_k(x, x')} \underbrace{Q_k^0(x, dx') \; \Psi_k^0(x')}_{R_k^0(x, dx')}$$

i.e. Assumption AC$_0$ holds

remark : assumption $(\star)$ is equivalent to

$$\mathbb{E}[\phi(X_k) \mid X_{k-1} = x] = \mathbb{E}^0[\phi(X_k)\, \Lambda_k \mid X_{k-1} = x] \qquad (\star\star)$$

for any test function ϕ, and for some (nonunique) r.v. Λ_k

indeed, if $(\star)$ holds, then

$$\begin{aligned} \mathbb{E}[\phi(X_k) \mid X_{k-1} = x] &= \int_E Q_k(x, dx')\, \phi(x') \\ &= \int_E Q_k^0(x, dx')\, q_k(x, x')\, \phi(x') \\ &= \mathbb{E}^0[\phi(X_k)\, q_k(X_{k-1}, X_k) \mid X_{k-1} = x] \end{aligned}$$

i.e. $(\star\star)$ holds with

$$\Lambda_k = q_k(X_{k-1}, X_k)$$

conversely, if $(\star\star)$ holds, then

$$\begin{aligned}
\int_E Q_k(x, dx')\, \phi(x') &= \mathbb{E}[\phi(X_k) \mid X_{k-1} = x] \\
&= \mathbb{E}^0[\phi(X_k)\, \Lambda_k \mid X_{k-1} = x] \\
&= \int_E \mathbb{E}^0[\Lambda_k \mid X_{k-1} = x, X_k = x']\, Q_k^0(x, dx')\, \phi(x')
\end{aligned}$$

i.e. $(\star)$ holds with

$$q_k(x, x') = \mathbb{E}^0[\Lambda_k \mid X_{k-1} = x, X_k = x']$$

even though no explicit expression is available in general for

$$q_k(x, x') = \mathbb{E}^0[\Lambda_k \mid X_{k-1} = x, X_k = x']$$

it may be easy to simulate jointly under $\mathbb{P}^0$ the r.v.'s X_k and Λ_k, given $X_{k-1} = x$

example (sampled SDE) :

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

if $(b(x) - b^0(x))$ is in the range of $\sigma(x)$, and if $\sigma(x)$ has full (column) rank, then $\mathbb{P} \ll \mathbb{P}^0$, with Radon–Nikodym derivative

$$\frac{d\mathbb{P}}{d\mathbb{P}^0}\Big|_{\mathcal{F}_t} = \exp\{ \int_0^t [\, \sigma^+(X'_s)\, (b(X'_s) - b^0(X'_s))\,]^*\, dW'^{\,0}_s$$

$$- \tfrac{1}{2} \int_0^t |\sigma^+(X'_s)\, (b(X'_s) - b^0(X'_s))\, |^2\, ds \,\}$$

with pseudo–inverse $\sigma^+(x) = [\sigma^*(x)\, \sigma(x)]^{-1}\, \sigma^*(x)$, hence under $\mathbb{P}^0$

$$dX'_t = b_0(X'_t)\, dt + \sigma(X'_t)\, dW'^{\,0}_t$$

in addition

$$\Lambda_k \;=\; \exp\Big\{ \int_{t_{k-1}}^{t_k} [\,\sigma^+(X'_s)\,(b(X'_s) - b^0(X'_s))\,]^*\, dW'^{\,0}_s - \tfrac{1}{2} \int_{t_{k-1}}^{t_k} |\sigma^+(X'_s)\,(b(X'_s) - b^0(X'_s))\,|^2\, ds \;\Big\}$$

and it is easy to simulate jointly under $\mathbb{P}^0$ the r.v.'s $X_k = X'_{t_k}$ and Λ_k, given $X'_{t_{k-1}} = x$ $\Box$

recall that

$$\mu_{k-1} \longrightarrow \mu_k = \bar{R}_k(\mu_{k-1}) = \frac{\mu_{k-1}\, R_k}{(\mu_{k-1}\, R_k)(E)}$$

smooth approximation : weighted empirical distribution

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N u_k^i\, w_k^{0,i}\, \delta_{\xi_k^{0,i}} \qquad \text{with} \qquad \sum_{i=1}^N w_k^{0,i} = 1 \quad \text{and} \quad \sum_{i=1}^N u_k^i\, w_k^{0,i} = 1$$

associated with

- unique particle system $\Sigma_k^0 = \{(\xi_k^{0,i}, w_k^{0,i})\, ,\, i = 1 \cdots N\}$ defined using the pivot (nominal) value θ_0 only, i.e. independent of θ
- secondary weights $S_k = \{u_k^i\, ,\, i = 1 \cdots N\}$ depending on θ

unique source of randomness (from Monte Carlo simulation, including interaction)

can expect smooth approximation, if secondary weights are smooth (continuous, differentiable, etc.) w.r.t. parameter how is this possible ? starting from particle approximation

$$\mu_{k-1} \approx \mu_{k-1}^N = \sum_{i=1}^N u_{k-1}^i \, w_{k-1}^{0,i} \, \delta_{\xi_{k-1}^{0,i}}$$

and applying nonnegative kernel $R_k(x, dx')$ exactly, with decomposition

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

under Assumption AC_0, yields

$$\begin{aligned} (\mu_{k-1}^N \, R_k)(dx') &= \sum_{i=1}^N u_{k-1}^i \, w_{k-1}^{0,i} \; R_k(\xi_{k-1}^{0,i}, dx') \\ &= \sum_{i=1}^N \underbrace{u_{k-1}^i \; r_k(\xi_{k-1}^{0,i}, x')}_{r_k^i(x')} \; \underbrace{\frac{w_{k-1}^{0,i}}{\pi_k^{0,i}} \; W_k^0(\xi_{k-1}^{0,i}, x')}_{w_k^{0,i}(x')} \; \underbrace{\pi_k^{0,i} \; P_k^0(\xi_{k-1}^{0,i}, dx')}_{m_k^{0,i}(dx')} \end{aligned}$$

same auxiliary particle idea as above

smooth SIR (sampling / importance resampling) algorithm

going from $(\Sigma^0_{k-1}, S_{k-1})$ to (Σ^0_k, S_k)

- selection of particles with higher weights : independently for $i = 1 \cdots N$

$$\tau_k^{0,i} \sim (\pi_k^{0,1} \cdots \pi_k^{0,N}) \qquad \text{with values in index set } \{1 \cdots N\}$$

- mutation using importance Markov kernel : independently for $i = 1 \cdots N$

$$\xi_k^{0,i} \sim P_k^0(\xi_{k-1}^{0,\tau_k^{0,i}}, dx')$$

- weighting according to importance weight : for $i = 1 \cdots N$

$$w_k^{0,i} = \frac{1}{c_k^{0,N}} \frac{w_{k-1}^{0,\tau_k^{i,0}}}{\pi_k^{0,\tau_k^{i,0}}} W_k^0(\xi_{k-1}^{0,\tau_k^{i,0}}, \xi_k^{0,i}) \qquad \text{s.t.} \qquad \sum_{k=1}^N w_k^{0,i} = 1$$

- updating secondary weights : for $i = 1 \cdots N$

$$u_k^i = \frac{1}{b_k^N} u_{k-1}^{\tau_k^{0,i}} r_k(\xi_{k-1}^{0,\tau_k^{0,i}}, \xi_k^{0,i}) \qquad \text{s.t.} \qquad \sum_{k=1}^N u_k^i w_k^{0,i} = 1$$

particle approximation of log–likelihood function

$$\ell_n = \sum_{k=0}^{n} \log(\mu_{k-1}\, R_k)(E)$$

using SIR algorithm

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

$$= \sum_{k=0}^{n} \log[\, \frac{1}{N} \sum_{i=1}^{N} \frac{w_{k-1}^{0,\tau_k^{0,i}}}{\pi_k^{0,\tau_k^{0,i}}}\, W_k^0(\xi_{k-1}^{0,\tau_k^{0,i}}, \xi_k^{0,i}) \sum_{i=1}^{N} u_{k-1}^{\tau_k^{0,i}}\, r_k(\xi_{k-1}^{0,\tau_k^{0,i}}, \xi_k^{0,i})\, w_k^{0,i}\,]$$

$$= \sum_{k=0}^{n} \log[\, \frac{1}{N}\, c_k^{0,N}\,] + \sum_{k=0}^{n} \log b_k^N$$

particle systems $\{\Sigma_k^0\,,\, 0 \le k \le n\}$ do not depend on the parameter
$\longrightarrow$ global, possibly regular, Monte Carlo approximation of log–likelihood function

convergence result : under technical assumptions

- on the model, including Lipschitz continuity of $\theta \longmapsto r_k(x, x')$ in some sense
- on the importance decomposition of kernels

it holds

$$\sup_{\|\phi\|=1} \{ \, \mathbb{E}[\sup_{\theta \in \Theta} |\mu_n^N(\phi) - \mu_n(\phi)|^p \,] \, \}^{1/p} \leq \frac{C_{p,n}}{\sqrt{N}}$$

and

$$\{ \, \mathbb{E}[\sup_{\theta \in \Theta} |\ell_n^N - \ell_n|^p \,] \, \}^{1/p} \leq \frac{C_{p,n}}{\sqrt{N}}$$

hence, almost surely as $N \to \infty$

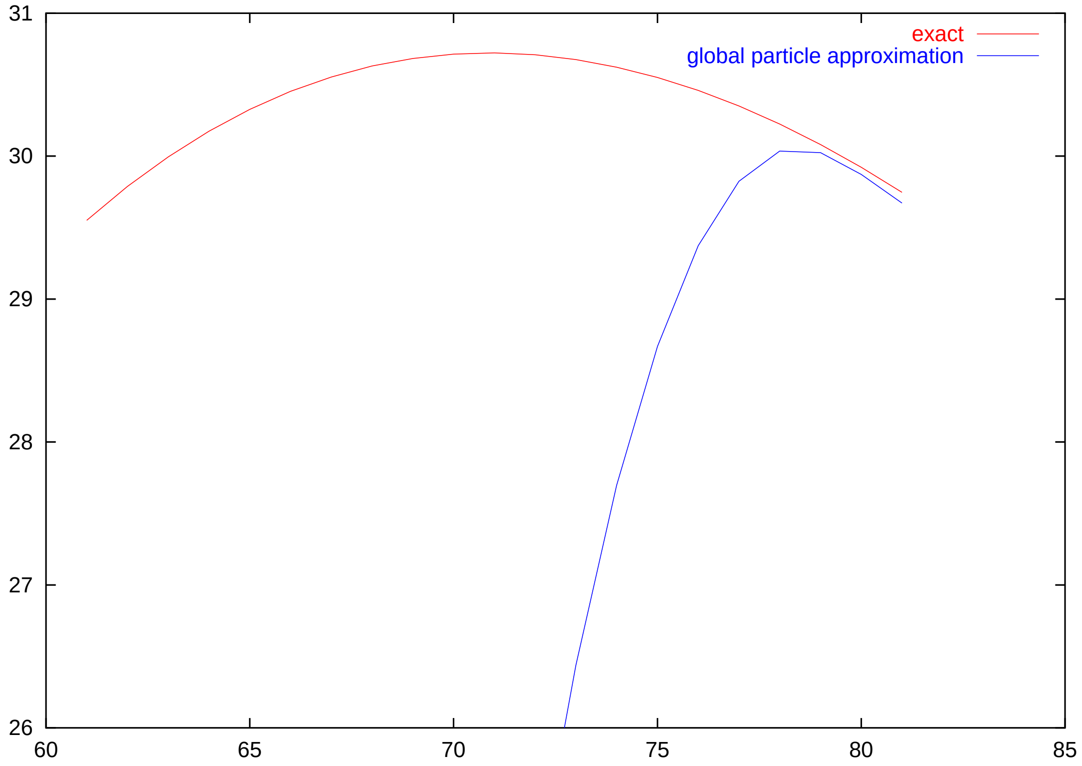
$$\sup_{\theta \in \Theta} |\ell_n^N - \ell_n| \longrightarrow 0$$

and

$$\widehat{\theta}_n^N = \operatorname*{argmax}_{\theta \in \Theta} \ell_n^N \longrightarrow \operatorname*{argmax}_{\theta \in \Theta} \ell_n = \widehat{\theta}_n$$

interacting particle implementation of the MCML algorithm of Geyer (1994, 1996)

numerical example



comments

- for values of parameter too much away from pivot (nominal) value, essentially weighted Monte Carlo approximation, without interaction : poor approximation, already noticed by Cappé, Douc, Moulines and Robert (2002)
- local (in the vicinity of pivot (nominal) value, if not global) vs. pointwise particle approximation
- can expect smooth approximation, if secondary weights are smooth (continuous, differentiable, etc.) w.r.t. parameter

Plan

- contrast functions in HMM
- interacting particle approximation
- global interacting particle approximation
- smooth interacting particle approximation
- linear tangent filter
- linear tangent filter : interacting particle approximation
- conclusion

nonnegative kernels $R_k(x, dx')$ differentiable w.r.t. parameter θ

Assumption AC

$$\frac{\partial}{\partial \theta} R_k(x, dx') = s_k(x, x') \, R_k(x, dx')$$

remark : if Assumption AC_0 holds, i.e. if

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

and if functions $r_k(x, x')$ are differentiable w.r.t. parameter θ

then Assumption AC holds, with

$$s_k(x, x') = \frac{\partial}{\partial \theta} \log r_k(x, x')$$

remark : recall that, if

$$Q_k(x, dx') = q_k(x, x') \, Q_k^0(x, dx') \qquad (\star)$$

then Assumption AC_0 holds, with

$$r_k(x, x') = q_k(x, x') \, \frac{\Psi_k(x')}{\Psi_k^0(x')}$$

hence if functions $q_k(x, x')$ and $\Psi_k(x')$ are differentiable w.r.t. parameter θ

then Assumption AC holds, with

$$s_k(x, x') = \frac{\partial}{\partial \theta} \log r_k(x, x') = \frac{\partial}{\partial \theta} \log q_k(x, x') + \frac{\partial}{\partial \theta} \log \Psi_k(x')$$

alternatively, if

$$\int_E Q_k(x, dx')\, \phi(x') = \mathbb{E}[\phi(X_k) \mid X_{k-1} = x] = \mathbb{E}^0[\phi(X_k)\, \Lambda_k \mid X_{k-1} = x] \quad (\star\star)$$

for any test function ϕ, and for some (nonunique) r.v. Λ_k, then

$$\begin{aligned} \int_E \frac{\partial}{\partial\theta} Q_k(x, dx')\, \phi(x') &= \mathbb{E}^0[\phi(X_k)\, \Lambda_k\, \frac{\partial}{\partial\theta} \log \Lambda_k \mid X_{k-1} = x] \\ &= \mathbb{E}[\phi(X_k)\, \frac{\partial}{\partial\theta} \log \Lambda_k \mid X_{k-1} = x] \\ &= \int_E \mathbb{E}[\frac{\partial}{\partial\theta} \log \Lambda_k \mid X_{k-1} = x, X_k = x']\, Q_k(x, dx')\, \phi(x') \end{aligned}$$

i.e.

$$\frac{\partial}{\partial\theta} Q_k(x, dx') = I_k(x, x')\, Q_k(x, dx')$$

with

$$I_k(x, x') = \mathbb{E}[\frac{\partial}{\partial\theta} \log \Lambda_k \mid X_{k-1} = x, X_k = x']$$

and

$$\frac{\partial}{\partial\theta} R_k(x, dx') \;=\; [\frac{\partial}{\partial\theta} Q_k(x, dx') + Q_k(x, dx') \, \frac{\partial}{\partial\theta} \log \Psi_k(x')] \; \Psi_k(x')$$

$$= \; [I_k(x, x') + \frac{\partial}{\partial\theta} \log \Psi_k(x')] \; R_k(x, dx')$$

i.e. Assumption AC holds, with

$$s_k(x, x') = I_k(x, x') + \frac{\partial}{\partial\theta} \log \Psi_k(x')$$

even though no explicit expression is available in general for

$$I_k(x, x') = \mathbb{E}[\frac{\partial}{\partial\theta} \log \Lambda_k \mid X_{k-1} = x, X_k = x']$$

it may be easy to simulate jointly under $\mathbb{P}$ the r.v.'s X_k and $\dfrac{\partial}{\partial\theta} \log \Lambda_k$, given $X_{k-1} = x$

example (sampled SDE) :

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

recall that, if $(b(x) - b^0(x))$ is in the range of $\sigma(x)$, and if $\sigma(x)$ has full (column) rank, then

$$\Lambda_k \;=\; \exp\Big\{ \int_{t_{k-1}}^{t_k} [\,\sigma^+(X'_s)\,(b(X'_s) - b^0(X'_s))\,]^*\,dW'^{\,0}_s$$

$$- \tfrac{1}{2} \int_{t_{k-1}}^{t_k} |\sigma^+(X'_s)\,(b(X'_s) - b^0(X'_s))\,|^2\,ds \;\Big\}$$

therefore

$$\frac{\partial}{\partial \theta} \log \Lambda_k = \int_{t_{k-1}}^{t_k} [\, \sigma^+(X'_s)\, \frac{\partial}{\partial \theta} b(X'_s)\,]^*\, dW'^{\,0}_s$$

$$- \int_{t_{k-1}}^{t_k} [\, \sigma^+(X'_s)\, \frac{\partial}{\partial \theta} b(X'_s)\,]^*\, [\, \sigma^+(X'_s)\, (b(X'_s) - b^0(X'_s))\,]\, ds$$

$$= \int_{t_{k-1}}^{t_k} [\, \sigma^+(X'_s)\, \frac{\partial}{\partial \theta} b(X'_s)\,]^*\, [dW'^{\,0}_s - \sigma^+(X'_s)\, (b(X'_s) - b^0(X'_s))\, ds\,]$$

$$= \int_{t_{k-1}}^{t_k} [\, \sigma^+(X'_s)\, \frac{\partial}{\partial \theta} b(X'_s)\,]^*\, dW'_s$$

and it is easy to simulate jointly under $\mathbb{P}$ the r.v.'s $X_k = X'_{t_k}$ and $\dfrac{\partial}{\partial \theta} \log \Lambda_k$, given $X'_{t_{k-1}} = x$ $\square$

recall that

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

and

$$u_k^i = \frac{1}{b_k^N} \, u_{k-1}^{\tau_k^{0,i}} \, r_k(\xi_{k-1}^{0,\tau_k^{0,i}}, \xi_k^{0,i}) \qquad \text{s.t.} \qquad \sum_{k=1}^N u_k^i \, w_k^{0,i} = 1$$

under Assumption AC, secondary weights are differentiable w.r.t. parameter θ, hence particle approximation is differentiable w.r.t. parameter θ, with derivative

$$\dot{\mu}_k^N = \frac{\partial}{\partial \theta} \mu_k^N = \sum_{i=1}^N \rho_k^i \, w_k^{0,i} \, \delta_{\xi_k^{0,i}} \qquad \text{with} \qquad \sum_{k=1}^N \rho_k^i \, w_k^{0,i} = 0$$

where (using logarithmic derivatives)

$$\rho_k^i = \frac{\partial}{\partial \theta} u_k^i \;=\; [\, \frac{\rho_{k-1}^{\tau_k^{0,i}}}{u_{k-1}^{\tau_k^{0,i}}} + s_k(\xi_{k-1}^{0,\tau_k^{0,i}}, \xi_k^{0,i}) - a_k^N \,] \, u_k^i$$

at pivot (nominal) value θ_0, it holds

$$\dot{\mu}_k^{0,N} = \frac{\partial}{\partial\theta} \mu_k^N \Big|_{\theta=\theta_0} = \sum_{i=1}^N \rho_k^{0,i} \, w_k^{0,i} \, \delta_{\xi_k^{0,i}} \qquad \text{s.t.} \qquad \sum_{k=1}^N \rho_k^{0,i} \, w_k^{0,i} = 0$$

where, taking $u_k^{0,i} \equiv 1$ into account

$$\rho_k^{0,i} = \frac{\partial}{\partial\theta} u_k^i \Big|_{\theta=\theta_0} = \rho_{k-1}^{0,\tau_k^{0,i}} + s_k^0(\xi_{k-1}^{0,\tau_k^{0,i}}, \xi_k^{0,i}) - a_k^{0,N}$$

questions

- is the optimal filter differentiable w.r.t. parameter (e.g. under Assumption AC) ?
- does the derivative of the smooth particle approximation provide a reasonable approximation of the linear tangent filter ?

Plan

- contrast functions in HMM
- interacting particle approximation
- global interacting particle approximation
- smooth interacting particle approximation
- linear tangent filter
- linear tangent filter : interacting particle approximation
- conclusion

recall that

$$\mu_{k-1} \longrightarrow \mu_k = \bar{R}_k(\mu_{k-1}) = \frac{\mu_{k-1}\, R_k}{(\mu_{k-1}\, R_k)(E)}$$

if Assumption AC holds, i.e. if

$$\frac{\partial}{\partial\theta} R_k(x, dx') = s_k(x, x')\, R_k(x, dx')$$

then

$$(\mu_{k-1}\, \frac{\partial}{\partial\theta} R_k)(dx') = \int_E \mu_{k-1}(dx)\, s_k(x, x')\, R_k(x, dx')$$

and $\{\mu_k\, ,\, k \geq 0\}$ is differentiable w.r.t. parameter, with derivative $\{\dot{\mu}_k\, ,\, k \geq 0\}$ given by linear tangent equation

$$(\mu_{k-1}, \dot{\mu}_{k-1}) \longrightarrow \dot{\mu}_k = \frac{\dot{\mu}_{k-1}\, R_k + \mu_{k-1}\, \dfrac{\partial}{\partial\theta} R_k}{(\mu_{k-1}\, R_k)(E)} - a_k\, \mu_k$$

with constant a_k s.t. signed measure $\dot{\mu}_k$ has zero total mass

Plan

- contrast functions in HMM
- interacting particle approximation
- global interacting particle approximation
- smooth interacting particle approximation
- linear tangent filter
- linear tangent filter : interacting particle approximation
- conclusion

by induction $\dot{\mu}_k \ll \mu_k$

joint approximation : weighted empirical distributions

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

$$\dot{\mu}_k \approx \dot{\mu}_k^N = \sum_{i=1}^N \rho_k^i \, w_k^i \, \delta_{\xi_k^i} \qquad \text{with} \qquad \sum_{i=1}^N \rho_k^i \, w_k^i = 0$$

associated with

- unique particle system $\Sigma_k = \{(\xi_k^i, w_k^i) \, , \, i = 1 \cdots N\}$
- signed weights $S'_k = \{\rho_k^i \, , \, i = 1 \cdots N\}$

starting from particle approximations

$$\mu_{k-1} \approx \mu^N_{k-1} = \sum_{i=1}^N w^i_{k-1}\, \delta_{\xi^i_{k-1}}$$

$$\dot{\mu}_{k-1} \approx \dot{\mu}^N_{k-1} = \sum_{i=1}^N \rho^i_{k-1}\, w^i_{k-1}\, \delta_{\xi^i_{k-1}}$$

and applying nonnegative kernels $R_k(x, dx')$ and $\dfrac{\partial}{\partial \theta} R_k(x, dx')$ exactly, yields

$$\begin{aligned}(\mu^N_{k-1}\, R_k)(dx') &= \sum_{i=1}^N w^i_{k-1}\, R_k(\xi^i_{k-1}, dx') \\ &= \sum_{i=1}^N \underbrace{\frac{w^i_{k-1}}{\pi^i_k}\, W_k(\xi^i_{k-1}, x')}_{w^i_k(x')}\, \underbrace{\pi^i_k\, P_k(\xi^i_{k-1}, dx')}_{m^i_k(dx')}\end{aligned}$$

and

$$(\dot{\mu}_{k-1}^N \, R_k + \mu_{k-1}^N \, \frac{\partial}{\partial\theta} R_k)(dx')$$

$$= \sum_{i=1}^N w_{k-1}^i \, [\, \rho_{k-1}^i + s_k(\xi_{k-1}^i, x') \,] \, R_k(\xi_{k-1}^i, dx')$$

$$= \sum_{i=1}^N \underbrace{\frac{w_{k-1}^i}{\pi_k^i} \, W_k(\xi_{k-1}^i, x')}_{w_k^i(x')} \, \underbrace{[\, \rho_{k-1}^i + s_k(\xi_{k-1}^i, x') \,]}_{r_k^i(x')} \, \underbrace{\pi_k^i \; P_k(\xi_{k-1}^i, dx')}_{m_k^i(dx')}$$

under Assumption AC

same auxiliary particle idea as above

joint SIR (sampling / importance resampling) algorithm

going from (Σ_{k-1}, S'_{k-1}) to (Σ_k, S'_k)

- selection of particles with higher weights : independently for $i = 1 \cdots N$

$$\tau_k^i \sim (\pi_k^1 \cdots \pi_k^N) \qquad \text{with values in index set } \{1 \cdots N\}$$

- mutation using importance Markov kernel : independently for $i = 1 \cdots N$

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

- weighting according to importance weight : for $i = 1 \cdots N$

$$w_k^i = \frac{1}{c_k^N} \frac{w_{k-1}^{\tau_k^i}}{\pi_k^{\tau_k^i}} W_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i) \qquad \text{s.t.} \qquad \sum_{k=1}^{N} w_k^i = 1$$

- updating signed weights : for $i = 1 \cdots N$

$$\rho_k^i = \rho_{k-1}^{\tau_k^i} + s_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i) - a_k^N \qquad \text{s.t.} \qquad \sum_{k=1}^{N} \rho_k^i \, w_k^i = 0$$

particle approximation of score function

recall expression for log–likelihood function

$$\ell_n = \sum_{k=0}^{n} \log(\mu_{k-1}\, R_k)(E)$$

hence

$$\frac{\partial}{\partial \theta} \ell_n = \sum_{k=0}^{n} \frac{(\dot{\mu}_{k-1}\, R_k + \mu_{k-1}\, \dfrac{\partial}{\partial \theta} R_k)(E)}{(\mu_{k-1}\, R_k)(E)}$$

using SIR algorithm

$$\ell_n \approx \ell_n^N = \sum_{k=0}^{n} \log[\, \frac{1}{N} \sum_{i=1}^{N} \frac{w_{k-1}^{\tau_k^i}}{\pi_k^{\tau_k^i}}\; W_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i)\,] = \sum_{k=0}^{n} \log[\, \frac{1}{N}\, c_k^N\,]$$

and

$$\frac{\partial}{\partial \theta} \ell_n \approx \frac{\partial}{\partial \theta} \ell_n^N = \sum_{k=0}^{n} \sum_{i=1}^{N} w_k^i\, [\, \rho_{k-1}^{\tau_k^i} + s_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i)\,] = \sum_{k=0}^{n} a_k^N$$

Plan

- contrast functions in HMM
- interacting particle approximation
- global interacting particle approximation
- smooth interacting particle approximation
- linear tangent filter
- linear tangent filter : interacting particle approximation
- conclusion

conclusion

- two different approaches to obtain interacting particle approximation of linear tangent filter
- more general version, under weaker form of Assumptions AC_0 and AC (no explicit expression for $r_k(x,x')$ and $s_k(x,x')$)

references and some related work

- particle and Monte Carlo approximation of log–likelihood function : Sørensen (PhD thesis 2000, 2003), Hürzeler and Künsch (2001)
- pointwise interacting particle approximation of linear tangent filter : Cérou, Le Gland and Newton (2001)
- interacting particle approximation of recursive MLE : Guyader, Le Gland and Oudjane (2003), Doucet and Tadić (2003)
- global interacting particle approximation of linear tangent filter : Caylus, Guyader, Le Gland and Oudjane (2004)
- interacting particle (or MCMC) implementation of fully Bayesian approach : dominant approach, with many contributions, including recently Papavasiliou (PhD thesis 2002, 2003), Rossi (PhD thesis 2004)