Stability of Feynman-Kac Semigroup and SMC Parameters’ Tuning

François GIRAUD

PhD student, CEA-CESTA, INRIA Bordeaux

Tuesday, February 14th 2012
1 **Recalls and context of the analysis**
   - Feynman-Kac semigroup
   - SMC algorithm
   - Quantities of interest

2 **Parameters’ tuning**
   - Estimation of $q_{p,n}$ and $b_{p,n}$
   - Examples of tuning and consequences

3 **The particular case of Gibbs measures**
   - Associated FK structure
   - Dobrushin analysis
   - Parameters $\beta_p$ and $m_p$
   - Concentration inequality
Feynman-Kac formulae

Let us fix \((E, \mathcal{E})\) a measurable state space. A sequence of measures \((\eta_n)\) on \(E\) admits a FK structure associated with potentials \(g_n : E \to \mathbb{R}_+\) and Markov kernels \(M_n : E \times \mathcal{E} \to [0, 1]\) if \(\forall n\),

\[
\eta_n = \phi_n(\eta_{n-1})
\]

with, for all measure \(\mu\) on \(E\) and bounded function \(f : E \to \mathbb{R}\),

\[
\phi_n(\mu)(f) = \frac{\mu(g_n \times M_n.f)}{\mu(g_n)}
\]

or in a equivalent way :

\[
\phi_n(\mu) = \psi_{g_n}(\mu).M_n
\]

FK semigroup

\[
\phi_{p,n} = \phi_n \circ \ldots \circ \phi_{p+1}
\]

so that \(\phi_{p,n}(\eta_p) = \eta_n\).
Structure of $\phi_{p,n}$

The composed function $\phi_{p,n}$ admits a structure similar to that of each function $\phi_p$, involving some selecting function $g_{p,n}$ and some Markov kernel $M_{p,n}$, so that:

$$\phi_{p,n}(\mu)(f) = \frac{\mu(g_{p,n} \times M_{p,n} \cdot f)}{\mu(g_{p,n})}$$

that is to say:

$$\phi_{p,n}(\mu) = \psi_{g_{p,n}}(\mu) \cdot M_{p,n}$$

given by the backward formulae:

$$g_{n,n} = 1 \quad g_{p-1,n} = g_p \cdot M_{p} \cdot g_{p,n}$$
$$M_{n,n} = Id \quad M_{p-1,n} \cdot f = \frac{M_{p} \cdot (g_{p,n} \cdot M_{p,n} \cdot f)}{M_{p} \cdot g_{p,n}}$$
Sequential Monte Carlo (SMC)

SMC algorithm consists in approximating a theoretical FK sequence \((\eta_n)\) by a large cloud of random samples termed particles \((\zeta^n_k)_{1 \leq k \leq N} \in \mathcal{E}^N\) defining at each generation \(n\) the occupation distribution:

\[
\eta^n_N = \frac{1}{N} \sum_{k=1}^{N} \delta_{\zeta^n_k}
\]

We run from generation \((\zeta^{n-1}_k)\) to generation \((\zeta^n_k)\) through a selection step using positive function \(g_n\) on \(\mathcal{E}\), and a mutation step, using Markov kernel \(M_n\).

\[
\begin{aligned}
\begin{bmatrix}
\zeta^1_n \\
\vdots \\
\zeta^i_n \\
\vdots \\
\zeta^N_n
\end{bmatrix}
\end{aligned}
\xrightarrow{S_{n, \eta^n_N}}
\begin{aligned}
\begin{bmatrix}
\hat{\zeta}^1_n \\
\vdots \\
\hat{\zeta}^i_n \\
\vdots \\
\hat{\zeta}^N_n
\end{bmatrix}
\xrightarrow{M_{n+1}}
\begin{bmatrix}
\zeta^1_{n+1} \\
\vdots \\
\zeta^i_{n+1} \\
\vdots \\
\zeta^N_{n+1}
\end{bmatrix}
\end{aligned}
\]

with the selection Markov transition:

\[
S_{\eta^n_{n-1}}(x, dy) = \varepsilon g_n(x).\delta_x(dy) + (1 - \varepsilon g(x))\psi_{g^n_n}(\eta^n_{n-1})(dy)
\]
Natural decomposition of the error

\[ \eta_n^N - \eta_n = \sum_{p=0}^{n} \phi_{p,n}(\eta_p^N) - \phi_{p,n} \left( \phi_p(\eta_{p-1}^N) \right) \]

No explosion in time \( n \) under stability conditions on \( \phi_{p,n} \).
Dobrushin coefficient

To each Markov kernel $K$ on $E$, is associated its Dobrushin coefficient $\beta(K) \in [0,1]$ defined by:

$$\beta(K) = \text{Sup}\{K(x,A) - K(y,A); x, y \in E, A \in \mathcal{E}\}$$

or in an equivalent way:

$$\beta(K) = \text{Sup}\{\|\mu.K - \nu.K\|_{tv}; \mu, \nu \in \mathcal{P}(E), \mu \neq \nu\}$$

$\beta$ satisfies $\beta(K_1.K_2) \leq \beta(K_1).\beta(K_2)$, for any kernels $K_1, K_2$, and by definition, for any measures $\mu, \nu$ on $E$ and any Markov kernel $K$, one have:

$$\|\mu.K - \nu.K\|_{tv} \leq \beta(K).\|\mu - \nu\|_{tv}.$$
"calculable" quantities of interest

We denote:

\[ b_p = \beta(M_p) \quad q_p = \sup_{x,y \in E} \frac{g_p(x)}{g_p(y)} \]

"theoretical" quantities of interest

\[ b_{p,n} = \beta(M_{p,n}) \quad q_{p,n} = \sup_{x,y \in E} \frac{g_{p,n}(x)}{g_{p,n}(y)} \]

A total variation stability result

For any measures \( \mu, \nu \) on \( E \), and any integers \( p \leq n \):

\[ \| \phi_{p,n}(\mu) - \phi_{p,n}(\nu) \|_{tv} \leq q_{p,n} \cdot b_{p,n} \cdot \| \mu - \nu \|_{tv} \]

This simple result already highlights the product \( q_{p,n} \cdot b_{p,n} \) as being a quantity representing the semigroup’s stability between transformations \( p \) and \( n \).
\[ L^2 \text{ mean error} \]

For any function \( f \), uniformly bounded by 1 on \( E \):

\[
\left\| \eta^N_n(f) - \eta_n(f) \right\|_2 \leq \frac{1}{\sqrt{N}} \sum_{k=0}^{n} q_{k,n} b_{k,n}
\]

\[ A \text{ concentration inequality} \]

For any bounded by 1 function \( f \) and any \( \varepsilon > 0 \):

\[
-\frac{1}{N} \log \mathbb{P} \left( \left| \eta^N_n(f) - \eta_n(f) \right| \geq \frac{r_n}{N} + \varepsilon \right) \geq \frac{\varepsilon^2}{2} \left[ b^*_n \overline{\beta}_n + \frac{\sqrt{2} r_n}{\sqrt{N}} + \varepsilon \left( 2r_n + \frac{b^*_n}{3} \right) \right]^{-1}
\]

where \( r_n, \overline{\beta}_n \) and \( b^*_n \) are constants so that:

\[
\begin{aligned}
    r_n &\leq \sum_{p=0}^{n} 4q^3_{p,n} b_{p,n} \\
    \overline{\beta}_n^{-2} &\leq \sum_{p=0}^{n} 4q^2_{p,n} b^2_{p,n} \\
    b^*_n &\leq \sup_{0 \leq p \leq n} 2q_{p,n} b_{p,n}
\end{aligned}
\]
### Estimation of $q_{p,n}$

For any integers $p \leq n$:

$$q_{p,n} - 1 \leq \sum_{k=p+1}^{n} (q_k - 1)b_{p+1} \cdots b_{k-1}$$

### Estimation of $b_{p,n}$

For any integers $p \leq n$,

$$b_{p,n} \leq \prod_{k=p+1}^{n} b_k q_{k,n}$$
Recalls and context of the analysis

Parameters’ tuning

The particular case of Gibbs measures

Examples of tuning and consequences

**Theorem**

If \( q_p \) are bounded by constant \( M \), then \( b_p \leq \frac{a}{M+a} \) (for any \( a \in (0,1) \)) ensures \( L^p \) mean error bound:

\[
\| \eta_n^N(f) - \eta_n(f) \|_p \leq \frac{B_p}{\sqrt{N}} \frac{1}{1-a}
\]

and concentration inequality:

\[
P\left( |\eta_n^N(f) - \eta_n(f)| \geq \frac{r_1^* N + r_2^* y}{N^2} \right) \leq e^{-y}
\]
Theorem

If \( q_p \) tends to 1 (decreasingly), then if \( b_p \) satisfies \( \forall p \geq 1, \)

\[
\begin{align*}
b_p &\leq \frac{q_p^{\alpha} - 1}{q_p^{\alpha+1} - 1} \quad p \to +\infty \\
b_p &\leq \frac{a}{q_p^{\alpha+1}} \quad p \to +\infty
\end{align*}
\]

(with \( \alpha = \frac{a}{1-a} \)), we still have the following \( L^p \) mean error bound:

\[
\| \eta_n^N(f) - \eta_n(f) \|_p \leq \frac{B_p}{\sqrt{N}} \frac{1}{1-a}
\]

and the concentration inequality:

\[
P \left( \left| \eta_n^N(f) - \eta_n(f) \right| \geq \frac{r_3^* N + r_4^* y}{N^2} \right) \leq e^{-y}
\]
Let us fix a potential $V : E \rightarrow \mathbb{R}$ and a strictly increasing sequence of "temperatures" $\beta_n$ so that $\beta_n \xrightarrow{n \rightarrow +\infty} +\infty$.

- $\eta_n(dx) = \mu_{\beta_n}(dx) = \frac{1}{Z_{\beta_n}} e^{-\beta_n V(x)} m(dx)$ with $m$ a reference measure. (Gibbs measures)
- $g_p(x) = e^{(\beta_p - \beta_{p-1}) V(x)}$
- and then $q_p = e^{\text{osc}(V) \Delta_p}$, with $\Delta_p = \beta_p - \beta_{p-1}$
- $M_n$ satisfying $\eta_n.M_n = \eta_n$ (for example the Simulated-annealing kernel)

$\text{SMC} = \text{Interacting simulated annealing.}$
Simulated annealing kernel $K_{\beta}$, Dobrushin estimation

- temperature $\beta$;
- proposition kernel $K(x, dy)$

In general, this kernel doesn’t have any mixing property in the sense of Dobrushin. However, under the assumption $K^{k_0}(x, .) \geq \delta \nu(.)$ for some integer $k_0$, some measure $\nu$ and some $\delta > 0$, we can show that:

$$\beta(K_{\beta}^{k_0}) \leq \left(1 - \delta e^{-\beta \Delta V(k_0)}\right)$$

To obtain suitable mixing properties, we choose $M_p = K_{\beta_p}^{k_0.m_p}$. 
Tuning of $\beta_p$ and $m_p$

For all $b \in ]0, 1[$, condition $b \leq b_p \leq b$ is then turned into

$$(1 - \delta e^{-\beta_p \Delta V(k_0)})^{m_p} \leq b,$$

which can also be written:

$$m_p \geq \frac{\log(1/b) e^{\Delta V(k_0) \cdot \beta_p}}{\delta}.$$
Theoretical Gibbs measure’s concentration on $V$’s global minimizer

$\forall \varepsilon > 0, \forall \varepsilon' > 0$ so that $0 < \varepsilon' < \varepsilon$,

$$\eta_n (V \geq V_{\text{min}} + \varepsilon) \leq \frac{e^{-\varepsilon' \beta_n}}{m_{\varepsilon'}}$$

SMC’s concentration on $V$’s global minimizer

Taking $f = 1\{V \geq V_{\text{min}} + \varepsilon\}$ in the general concentration inequality and noticing that $\eta_n^N . f$ then designates the proportion $p_n^N(\varepsilon)$ of particles $(\zeta_n^i)$ satisfying $V(\zeta_n^i) \geq V_{\text{min}} + \varepsilon$, we obtain:

$$\mathbb{P}\left(p_n^N(\varepsilon) \geq \frac{e^{-\varepsilon' \beta_n}}{m_{\varepsilon'}} + \frac{r_i^* N + r_j^* y}{N^2}\right) \leq e^{-y}.$$
Thank you for your attention!