A Backward Particle Interpretation of Feynman-Kac Formulae

P. Del Moral

Centre INRIA de Bordeaux - Sud Ouest

Workshop on Filtering, Cambridge Univ., June 14-15th 2010

Preprints (with hyperlinks), joint works with A. Doucet & S.S. Singh:

- A Backward Particle Interpretation of Feynman-Kac Formulae `HAL-INRIA RR-7019 (2009)`, to appear in M2AN (2010).
- Forward Smoothing Using Sequential Monte Carlo `CUED/F-INFENG/TR.638`. Cambridge University, Engineering Dpt. (2009).

Outline

Summary

1. Introduction, motivations
   - Some notation
   - Feynman-Kac measures
   - Filtering and sensitivity analysis
2. Nonlinear Markov models
3. Some convergence results
4. Additive functionals

Some notation

E measurable state space, $\mathcal{P}(E)$ proba. on E, $\mathcal{B}(E)$ bounded meas. functions

- $(\mu, f) \in \mathcal{P}(E) \times \mathcal{B}(E) \quad \longrightarrow \quad \mu(f) = \int \mu(dx)\ f(x)$
- $M(x, dy)$ **integral operator over E**

$$M(f)(x) = \int M(x, dy) f(y)$$

$$[\mu M](dy) = \int \mu(dx) M(x, dy) \qquad (\Longrightarrow [\mu M](f) = \mu[M(f)]\ )$$

- **Bayes-Boltzmann-Gibbs transformation :** $G : E \rightarrow [0, \infty[$ with $\mu(G) > 0$

$$\Psi_G(\mu)(dx) = \frac{1}{\mu(G)}\ G(x)\ \mu(dx)$$

E finite $\Leftrightarrow$ Vector-Matrix notation $\mu = [\mu(1), \ldots, \mu(d)]$ and $f = [f(1), \ldots, f(d)]'$

Feynman-Kac measures

- Markov X_n with transitions $M_n(x_{n-1}, dx_n)$ on some state space E_n.
- Potential functions $G_n \ : \ x_n \in E_n \to G_n(x_n) \in [0, \infty[$

Feynman-Kac path measures:

$$d\mathbb{Q}_n := \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \leq p < n} G_p(X_p) \right\} d\mathbb{P}_n \quad \text{with} \quad \mathbb{P}_n := \mathrm{Law}(X_0, \ldots, X_n)$$

The *n*-time marginals: $\forall f_n \in \mathcal{B}(E_n)$

$$\eta_n(f_n) := \frac{\gamma_n(f_n)}{\gamma_n(1)} \quad \text{with} \quad \gamma_n(f_n) := \mathbb{E}\left(f_n(X_n) \prod_{0 \leq p < n} G_p(X_p) \right)$$

Updated measures $\longrightarrow$ $\widehat{\mathbb{Q}}_n, \widehat{\eta}_n$, and $\widehat{\gamma}_n$ w.r.t. the product $\prod_{0 \leq p \leq n}$

Example : Nonlinear filtering

$$\mathbb{P}\left((X_n, Y_n) \in d(x,y) | (X_{n-1}, Y_{n-1})\right) := M_n(X_{n-1}, dx)\ g_n(x,y)\ \lambda_n^Y(dy)$$

- Given the observation sequence $Y = y$ with $G_n(x_n) = g_n(x_n, y_n)$

$$\widehat{\eta}_n = \mathrm{Law}(X_n \mid \forall 0 \leq p \leq n \quad Y_p = y_p)$$

$$\widehat{\gamma}_n(1) = p_n(y_0, \ldots, y_n) \qquad \left(\text{density w.r.t.} \quad \otimes_{0 \leq p \leq n} \lambda_p^Y\right)$$

- In path space settings $\rightsquigarrow$ smoothing and path estimation

$$\widehat{\mathbb{Q}}_n = \mathrm{Law}((X_0, \ldots, X_n) \mid \forall 0 \leq p \leq n \quad Y_p = y_p)$$

Sensitivity analysis $\oplus$ Expected Maximization

$$\theta \mapsto \quad M_n(x_{n-1}, dx_n) = p_n^{\theta}(x_{n-1}, x_n)\ \lambda_n^X(dx_n) \quad \text{and} \quad g_n^{\theta}(x,y)\ \lambda_n^Y(dy)$$

$$\Downarrow$$

Parametric Feynman-Kac models : $(G_n, \mathbb{Q}_n, \gamma_n, \eta_n, \ldots) \rightsquigarrow (G_n^\theta, \mathbb{Q}_n^\theta, \gamma_n^\theta, \eta_n^\theta, \ldots)$

$$\frac{\partial}{\partial\theta} \log \widehat{\gamma}_n^\theta(1) = \widehat{\mathbb{Q}}_n^\theta(F_n^\theta) \quad \text{and} \quad \frac{\partial}{\partial\theta} \widehat{\mathbb{Q}}_n^\theta(\varphi_n) = \widehat{\mathbb{Q}}_n^\theta\left(F_n^\theta\ [\varphi_n - \widehat{\mathbb{Q}}_n^\theta(\varphi_n)]\right)$$

with the additive functional

$$\begin{aligned} F_n^\theta(x_0, \ldots, x_n) &:= \sum_{0\leq k\leq n} f_k^\theta(x_{k-1}, x_k) \\ f_k^\theta(x_{k-1}, x_p) &:= \frac{\partial \log G_p^\theta(x_k)}{\partial\theta} + \frac{\partial \log p_k^\theta(x_{k-1}, x_k)}{\partial\theta} \end{aligned}$$

Expected Maximization $\rightsquigarrow$ min entropy (additive fct. $F_n^{\theta'}$ without derivatives)

$$\theta' \mapsto \mathcal{E}_n(\theta, \theta') := \widehat{\mathbb{Q}}_n^\theta\left(\log\left(d\widehat{\mathbb{Q}}_n^\theta / d\widehat{\mathbb{Q}}_n^{\theta'}\right)\right) \text{"} \propto \text{"} - \widehat{\mathbb{Q}}_n^\theta(F_n^{\theta'})$$

3 Key observations

$$\eta_n(f_n) := \frac{\gamma_n(f_n)}{\gamma_n(1)} \quad \text{with} \quad \gamma_n(f_n) := \mathbb{E}\left(f_n(X_n) \prod_{0\leq p<n} G_p(X_p)\right)$$

- **Path space models:**

$$[\ X_n := (X'_0, \ldots, X'_n) \ \ \& \ \ G_n(X_n) := G'_n(X'_n)] \implies \eta_n = \mathbb{Q}_n$$

- **A multiplicative formula:**

$$\mathcal{Z}_n = \mathbb{E}\left(\prod_{0\leq p<n} G_p(X_p)\right) = \prod_{0\leq p<n} \eta_p(G_p)$$

Proof:

$$\mathcal{Z}_n := \gamma_n(1) = \gamma_{n-1}(G_{n-1}) = \eta_{n-1}(G_{n-1})\ \gamma_{n-1}(1)$$

3 Key observations

- **Path space models:**

$$M_n(X_{n-1}, dx_n) = \mathbb{P}\left(X_n \in dx_n \mid X_{n-1}\right) = H_n(X_{n-1}, x_n)\ \lambda_n(dx_n)$$
$$\Downarrow$$
$$\mathbb{Q}_n(d(x_0, \ldots, x_n)) = \eta_n(dx_n)\ M_{n,\eta_{n-1}}(x_n, dx_{n-1}) \ldots M_{1,\eta_0}(x_1, dx_0)$$

with the backward Markov transitions :

$$M_{n,\eta_{n-1}}(x_n, dx_{n-1}) \propto G_{n-1}(x_{n-1})\ H_n(x_{n-1}, x_n)\ \eta_{n-1}(dx_{n-1})$$

Summary

Flows of Feynman-Kac measures

- A two step correction prediction model

$$\eta_n \xrightarrow{\text{Updating-correction}} \widehat{\eta}_n = \Psi_{G_n}(\eta_n) \xrightarrow{\text{Prediction/Markov transport}} \eta_{n+1} = \widehat{\eta}_n M_{n+1}$$

- Selection nonlinear transport formulae

$$\Psi_{G_n}(\eta_n) = \eta_n S_{n,\eta_n}$$

with, for any $\epsilon_n = \epsilon_n(\eta_n) \in [0,1]$ s.t. $\epsilon_n G_n \leq 1$

$$S_{n,\eta_n}(x,.) := \epsilon_n G_n(x)\ \delta_x + (1 - \epsilon_n G_n(x))\ \Psi_{G_n}(\eta_n)$$

$$\Downarrow$$

$$\eta_{n+1} = \eta_n \left(S_{n,\eta_n} M_{n+1} \right) := \eta_n K_{n+1,\eta_n}$$

Nonlinear Markov chains $\eta_n = \mathrm{Law}(\overline{X}_n)$=Perfect sampling algorithm

- **Nonlinear transport formulae :**

$$\eta_{n+1} = \eta_n K_{n+1,\eta_n}$$

with the collection of Markov probability transitions :

$$K_{n+1,\eta_n} = S_{n,\eta_n} M_{n+1}$$

- **Local transitions :**

$$\mathbb{P}(\overline{X}_n \in dx_n \mid \overline{X}_{n-1}) = K_{n,\eta_{n-1}}(\overline{X}_{n-1}, dx_n) \quad \text{avec} \quad \eta_{n-1} = \mathrm{Law}(\overline{X}_{n-1})$$

- **McKean measures (canonical process) :**

$$\mathbb{P}_n(d(x_0,\ldots,x_n)) = \eta_0(dx_0)\ K_{1,\eta_0}(x_0, dx_1)\ldots K_{n,\eta_{n-1}}(x_{n-1}, dx_n)$$

Sampling pb $\Rightarrow$ Mean field particle interpretations

- Markov Chain $\xi_n = (\xi_n^1, \ldots, \xi_n^N) \in E_n^N$ s.t.

$$\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq_{N \uparrow \infty} \eta_n$$

- Approximated local transitions ($\forall 1 \leq i \leq N$)

$$\xi_{n-1}^i \rightsquigarrow \xi_n^i \sim K_{n, \eta_{n-1}^N}(\xi_{n-1}^i, dx_n)$$

Schematic picture : $\xi_n \in E_n^N \rightsquigarrow \xi_{n+1} \in E_{n+1}^N$

$$\begin{array}{ccc}
\xi_n^1 & \xrightarrow{K_{n+1,\eta_n^N}} & \xi_{n+1}^1 \\
\vdots & & \vdots \\
\xi_n^i & \longrightarrow & \xi_{n+1}^i \\
\vdots & & \vdots \\
\xi_n^N & \longrightarrow & \xi_{n+1}^N
\end{array}$$

Rationale :

$$\eta_n^N \simeq_{N\uparrow\infty} \eta_n \implies K_{n+1,\eta_n^N} \simeq_{N\uparrow\infty} K_{n+1,\eta_n} \Longrightarrow \xi^i \sim \text{i.i.d. copies of } \overline{X}$$

$$\Downarrow$$

Particle McKean measures :

$$\frac{1}{N}\sum_{i=1}^N \delta_{(\xi_0^i,\ldots,\xi_n^i)} \longrightarrow_{N\uparrow\infty} \text{Law}(\overline{X}_0,\ldots,\overline{X}_n)$$

Some key advantages

- Mean field models=stochastic linearization/perturbation technique :

$$\eta_n^N = \eta_{n-1}^N K_{n,\eta_{n-1}^N} + \frac{1}{\sqrt{N}}\ W_n^N$$

avec $W_n^N \simeq W_n$ Centered Gaussian Fields $\perp$.

- $\eta_n = \eta_{n-1} K_{n,\eta_{n-1}}$ stable $\Rightarrow$ No propagation of local sampling errors

$$\Longrightarrow \text{Uniform control w.r.t. the time horizon}$$

- "No burning, no need to study the stability of MCMC models".
- Stochastic adaptive grid approximation
- Nonlinear system $\rightsquigarrow$ "positive-benefic interactions.
- Simple and natural sampling algorithm.

Feynman-Kac models $\Leftrightarrow$ Genetic type stochastic algo.

$$\left.\begin{array}{c}\xi_n^1\\ \vdots\\ \xi_n^i\\ \vdots\\ \xi_n^N\end{array}\right] \xrightarrow{S_{n,\eta_n^N}} \left[\begin{array}{ccc}\widehat{\xi}_n^1 & \xrightarrow{M_{n+1}} & \xi_{n+1}^1\\ \vdots & & \vdots\\ \widehat{\xi}_n^i & \longrightarrow & \xi_{n+1}^i\\ \vdots & & \vdots\\ \widehat{\xi}_n^N & \longrightarrow & \xi_{n+1}^N\end{array}\right]$$

Acceptance/Rejection-Selection : [Geometric type clocks]

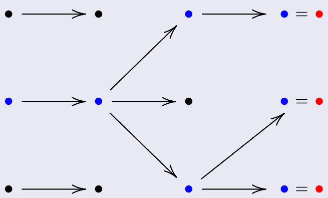
$$S_{n,\eta_n^N}(\xi_n^i, dx)$$

$$:= \epsilon_n G_n(\xi_n^i)\ \delta_{\xi_n^i}(dx) + \left(1 - \epsilon_n G_n(\xi_n^i)\right)\ \sum_{j=1}^N \frac{G_n(\xi_n^j)}{\sum_{k=1}^N G_n(\xi_n^k)} \delta_{\xi_n^j}(dx)$$

Ex. : $G_n = 1_A \rightsquigarrow G_n(\xi_n^i) = 1_A(\xi_n^i)$

Interaction/branch. process $\hookrightarrow$ 4 types of occupation measures

$(N=3)$



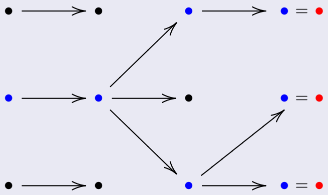
- Current population $\hookrightarrow \dfrac{1}{N}\sum_{i=1}^{N} \delta_{\xi_n^i \leftarrow i\text{-th individual at time } n} \simeq \eta_n$
- Genealogical tree $\hookrightarrow \dfrac{1}{N}\sum_{i=1}^{N} \delta_{(\xi_{0,n}^i,\xi_{1,n}^i,\ldots,\xi_{n,n}^i)\leftarrow i\text{-th ancestral line}} \simeq \mathbb{Q}_n$

$\oplus$ Unbias particle normalizing Constants

$$\mathcal{Z}_n^N := \prod_{0\leq p<n} \eta_p^N(G_p) \simeq \prod_{0\leq p<n} \eta_p(G_p) = \mathcal{Z}_n$$

Interaction/branch. process $\hookrightarrow$ 4 types of occupation measures

$(N = 3)$



- Complete genealogical tree $\simeq$ McKean measures

$$\frac{1}{N}\sum_{i=1}^{N} \delta_{(\xi_0^i,\xi_1^i,\ldots,\xi_n^i)} \simeq \eta_0(dx_0)K_{1,\eta_0}(x_0,dx_1)\ldots K_{n,\eta_{n-1}}(x_{n-1},dx_n)$$

- Backward Feynman-Kac path measures [$\rightsquigarrow$ elementary Matrices operations]

$$\begin{aligned}
\mathbb{Q}_n^N(d(x_0,\ldots,x_n)) &= \eta_n^N(dx_n)\ M_{n,\eta_{n-1}^N}(x_n,dx_{n-1})\ldots M_{1,\eta_0^N}(x_1,dx_0)\\
&\simeq \eta_n(dx_n)\ M_{n,\eta_{n-1}}(x_n,dx_{n-1})\ldots M_{1,\eta_0}(x_1,dx_0)\\
&= \mathbb{Q}_n(d(x_0,\ldots,x_n))
\end{aligned}$$

Equivalent Stochastic Algorithms :

- Genetic and evolutionary type algorithms.
- Spatial branching models.
- Sequential Monte Carlo methods.
- Population Monte Carlo models.
- Diffusion Monte Carlo (DMC), Quantum Monte Carlo (QMC), ...
- Some botanical names $\sim \neq$ application domain areas :

  particle filters, bootsrapping, selection, pruning-enrichment, reconfiguration, cloning, go with the winner, spawning, condensation, grouping, rejuvenations, harmony searches, biomimetics, splitting, ...

$$\Updownarrow$$

$1950 \leq$ [(Meta)Heuristics] $\leq 1996 \leq$ Feynman-Kac mean field particle model

Summary

"Asympt." theo. TCL,PGD, PDM $\oplus$ (n,N) fixed $\rightsquigarrow$ some examples :

- Empirical processes :

$$\sup_{n\geq 0}\sup_{N\geq 1}\sqrt{N}\ \mathbb{E}(\|\eta_n^N-\eta_n\|_{\mathcal{F}_n}^p)<\infty$$

- Concentration inequalities uniform w.r.t. time :

$$\sup_{n\geq 0}\mathbb{P}(|\eta_n^N(f_n)-\eta_n(f_n)|>\epsilon)\leq c\ \exp-(N\epsilon^2)/(2\ \sigma^2)$$

$+$ *Guionnet* $\sup_{n\geq 0}$ `(IHP 01)` & *Ledoux* $\sup_{\mathcal{F}_n}$ `(JTP 00)` & *Rio* `AAP 10`

- Propagations of chaos exansions *(+Patras,Rubenthaler* `(AAP 09)`) :

$$\begin{aligned}\mathbb{P}^N_{n,q} &:= \mathrm{Loi}(\xi_n^1,\ldots,\xi_n^q)\\ &\simeq \eta_n^{\otimes q}+\frac{1}{N}\ \partial^1\mathbb{P}_{n,q}+\ldots+\frac{1}{N^k}\ \partial^k\mathbb{P}_{n,q}+\frac{1}{N^{k+1}}\ \partial^{k+1}\mathbb{P}^N_{n,q}\end{aligned}$$

with $\sup_{N\geq 1}\|\partial^{k+1}\mathbb{P}^N_{n,q}\|_{\mathrm{tv}}<\infty$ & $\sup_{n\geq 0}\|\partial^1\mathbb{P}_{n,q}\|_{\mathrm{tv}}\leq c\ q^2$.

Unnormalized models

Un-bias particle approximation measures

$$\gamma_n^N(f_n) := \eta_n^N(f_n) \prod_{0\leq p<n} \eta_p^N(G_p)$$

- **Asymptotic theorms** : fluctuations & deviations
  + A. Guionnet (`AAP 99`, `SPA 98`*), + L. Miclo* (`SP 2000`), *+ D. Dawson*
- Non asymptotic theory : bias and variance estimates
  1. Taylor type expansion *(+Patras & Rubenthaler* (`AAP 09`) :

$$\mathbb{E}\left((\gamma_n^N)^{\otimes q}(F)\right) =: \mathbb{Q}_{n,q}^N(F) = \gamma_n^{\otimes q}(F) + \sum_{1\leq k\leq (q-1)(n+1)} \frac{1}{N^k}\, \partial^k \mathbb{Q}_{n,q}(F)$$

  2. *Variance estimates* (+Cerou & Guyader `Hal-INRIA 08` & IPH 2010) *:*

$$\mathbb{E}\left(\left[\gamma_n^N(f_n) - \gamma_n(f_n)\right]^2\right) \leq\ c\ \frac{n}{N}\ \times\ \gamma_n(1)^2$$

Summary

Additive funct. *(+ Doucet & Singh* `Hal-INRIA 09` *& M2AN 2010)*

- Path space models $\mathbb{P}_n := \mathrm{Law}(X_0, \ldots, X_n)$

$$d\mathbb{Q}_n := \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \leq p < n} G_p(X_p) \right\} d\mathbb{P}_n$$

- Hyp. : $M_n(x_{n-1}, dx_n) = H_n(x_{n-1}, x_n)\ \lambda_n(dx_n)$

$$\Rightarrow \mathbb{Q}_n(d(x_0, \ldots, x_n)) = \eta_n(dx_n)\ M_{n,\eta_{n-1}}(x_n, dx_{n-1}) \ldots M_{1,\eta_0}(x_1, dx_0)$$

with the backward transitions :

$$M_{p+1,\eta}(x, dx') \propto G_p(x')\ H_{p+1}(x', x)\ \eta(dx')$$

- Particle estimates $\sim$ complete genealogical tree :

$$\mathbb{Q}_n^N(d(x_0, \ldots, x_n)) = \eta_n^N(dx_n)\ M_{n,\eta_{n-1}^N}(x_n, dx_{n-1}) \ldots M_{1,\eta_0^N}(x_1, dx_0)$$

2 type of path-space estimates

- Complete genealogical tree $\Longrightarrow$ McKean meas. $\oplus$ FK-Path space

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{(\xi_0^i,\ldots,\xi_n^i)} \simeq_N \mathrm{Loi}(\overline{X}_0,\ldots,\overline{X}_n) \quad \& \quad \mathbb{Q}_n^N \simeq_N \mathbb{Q}_n$$

- Simple genealogical tree $\Longrightarrow$ FK-Path space

$$\eta_n^N = \frac{1}{N}\sum_{i=1}^{N}\delta_{(\xi_{0,n}^i,\xi_{1,n}^i,\ldots,\xi_{n,n}^i)} \simeq_N \mathbb{Q}_n = \eta_n$$

Main problem :

Path degeneracy w.r.t. time horizon (as any genetic ancestral tree)

$$\Downarrow$$

Roughly : Uniform estimates $\rightsquigarrow$ linear estimates w.r.t. the time horizon

Some non asymptotic estimates

Additive functional :

$$F_n(x_0,\ldots,x_n) = \frac{1}{n+1}\sum_{0\leq p\leq n} f_p(x_p)$$

- Bias estimate+uniform $\mathbb{L}_p$-bounds + variance $(1/N^2)$

$$N\ \mathbb{E}\left([(\mathbb{Q}_n^N-\mathbb{Q}_n)(F_n)]^2\right) \leq c \times \left(\underbrace{1/n}_{\text{bias term}} + 1/N\right)$$

- Uniform exponential concentration

$$\frac{1}{N}\log \sup_{n\geq 0}\ \mathbb{P}\left(\left|[\mathbb{Q}_n^N-\mathbb{Q}_n](F_n)\right| \geq \frac{b}{\sqrt{N}}+\epsilon\right) \leq -\epsilon^2/(2b^2)$$