Concentration $\leq$ for Mean Field Particle Models

P. Del Moral, E. Rio

Centre INRIA de Bordeaux - Sud Ouest

ICM, Indian Statistical Institute, Bangalore August 13-17, 2010.

• Concentration Inequalities for Mean Field Particle Models [HAL-INRIA RR-6901 (09)]. To appear in the Annals of Applied Probability (2010).
• DM, A. Doucet and S.S. Singh A Backward Particle Interpretation of Feynman-Kac Formulae HAL-INRIA RR-7019 (09). To appear in M2AN (2010)

Outline

Summary

Some notation

E measurable state space, $\mathcal{M}(E)$ & $\mathcal{P}(E)$ measures & probabilities on E
$\mathcal{B}(E)$ bounded meas. functions

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

$$\begin{aligned} M(f)(x) &= \int M(x, dy) f(y) \\ [\mu M](dy) &= \int \mu(dx) M(x, dy) \qquad (\Longrightarrow [\mu M](f) = \mu[M(f)]\ ) \end{aligned}$$

- **Composition:** $(M_1 M_2)(x, dz) = \int M_1(x, dy) M_2(y, dz)$
- **Boltzmann-Gibbs transformation :** $G : E \to [0, \infty[$ with $\mu(G) > 0$

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

Running example : Feynman-Kac integration models

- Markov chain X_n on some state spaces E_n, n=time index .
- Potential functions $G_n \ : \ x_n \in E_n \to G_n(x_n) \in [0,1]$

Feynman-Kac path measures:

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

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

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

Example: Filtering, Hidden Markov Chains, Bayesian Inference

Signal-observation type model

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

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

$$\eta_n = \mathrm{Law}(X_n \mid \forall 0 \leq p < n \quad Y_p = y_p) \quad \text{and} \quad \mathcal{Z}_{n+1} \propto p_n(y_0, \ldots, y_n)$$

- In path space settings

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

Other examples of conditioning:

$\subset$ rare event analysis, confinements problems, polymers, self-avoiding walks, ..

Ex.: $G_n = 1_{A_n} \Rightarrow \mathbb{Q}_n = \mathrm{Law}((X_0, \ldots, X_n) \mid \forall 0 \leq p < n \quad X_p \in A_p)$

3 Key observations

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

- **First important observation:**

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

- **Second important observation:**

$$\mathcal{Z}_n = \gamma_n(1) = \mathbb{E}\left(\prod_{0 \leq p < n} G_p(X_p)\right) = \prod_{0 \leq p < n} \eta_p(G_p)$$

Proof:

$$\gamma_n(1) = \gamma_{n-1}(G_{n-1}) = \eta_{n-1}(G_{n-1})\ \gamma_{n-1}(1)$$

3 Key observations

- **Third important observation:** (H) $M_n(x, dx') = H_n(x, x')\ \lambda_n(dx')$

$$
\begin{aligned}
\mathbb{Q}_n(d(x_0, \ldots, x_n)) &= \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \leq p < n} G_p(x_p) \right\} \mathbb{P}_n(d(x_0, \ldots, x_n)) \\
&= \eta_n(dx_n)\ M_{n,\eta_{n-1}}(x_n, dx_{n-1}) \ldots M_{1,\eta_0}(x_1, dx_0)
\end{aligned}
$$

with the backward Markov transitions :

$$
M_{n,\eta}(x, dx') := \frac{\eta(dx') G_{n-1}(x')\ H_n(x', x)}{\eta\left(G_{n-1}\ H_n(\cdot, x)\right)}
$$

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 $\epsilon_n \in [0,1]$

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

$$\Downarrow$$

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

Summary

1. Introduction, motivations
2. Mean field particle models
   - Nonlinear McKean distribution flows
   - Mean field particle interpretations
   - The 4 types of particle approximation measures
   - Some key advantages
3. Convergence analysis

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

- **Nonlinear transport formulae :**

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

- **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) :**

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

Other examples : Gaussian-McKean Vlasov type transitions

$$\overline{X}_{n+1} := d_n\left(\overline{X}_n, \eta_n\right) + \mathcal{N}(0, Q_n)$$

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 ($\forall 1 \leq i \leq N$)

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

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

$$
\begin{bmatrix} \xi_n^1 \\ \vdots \\ \xi_n^i \\ \vdots \\ \xi_n^N \end{bmatrix}
\xrightarrow{S_{n,\eta_n^N}}
\begin{bmatrix} \widehat{\xi}_n^1 & \xrightarrow{M_{n+1}} & \xi_{n+1}^1 \\ \vdots & & \vdots \\ \widehat{\xi}_n^i & \longrightarrow & \xi_{n+1}^i \\ \vdots & & \vdots \\ \widehat{\xi}_n^N & \longrightarrow & \xi_{n+1}^N \end{bmatrix}
$$

Accept/Reject-Selection : [Geometric clocks] [Confinement ex. : $G_n = 1_A$]

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

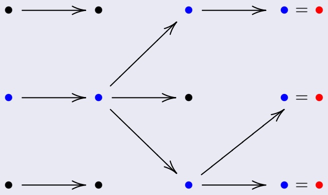
$\oplus$ Unbias particle normalizing Cts

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

$\supset$ *Particle filters, Diffusion Monte Carlo (DMC), Quantum Monte Carlo (QMC), Sequential Monte Carlo methods (SMC), ...*

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

$(N = 3)$



- Current population $\hookrightarrow \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_n^i \leftarrow i\text{-th individual at time } n} \simeq \eta_n$
- Genealogical tree $\hookrightarrow \frac{1}{N}\sum_{i=1}^{N} \delta_{(\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i) \leftarrow i\text{-th ancestral line}} \simeq \mathbb{Q}_n$
- **Complete genealogical tree** $\hookrightarrow \frac{1}{N}\sum_{i=1}^{N} \delta_{(\xi_0^i, \xi_1^i, \ldots, \xi_n^i)} \simeq$ **McKean meas.**
- Forward particle approximation $\sim$ complete genealogical tree :

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

Some key advantages

- Mean field models=stochastic linearization/perturbation technique :

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

  with [theorem] $(W_n^N)_n \simeq (W_n)_n$ Indep. Centered Gaussian Fields.
- $\eta_n = \eta_{n-1} K_{n,\eta_{n-1}}$ 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 type models".
- PDE viewpoint : Stochastic adaptive grid approximation
- Nonlinear system $\rightsquigarrow$ "positive-benefic interactions.
- Simple and natural sampling algorithm.

Summary

Feynman-Kac models = rather complete analysis [LDP,CLT, Propagation of chaos] (FK Springer 2004 + refs)

- Example: Empirical processes $\mathbb{L}_p$-mean error estimates

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

- (New) Propagations of chaos *(+Patras & Rubenthaler* (AAP 10)*)*

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

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

- (New) Additive functional *(+ Doucet & Singh* M2AN 10*)* :

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

Concentration inequalities

- Large deviation principles

  [joint works with D. Dawson (Springer 05), A. Guionnet (SPA 98,IHP 01), T. Zajic (Bernoulli 03)]

$$\rightsquigarrow \sup_{n\geq 0} \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(\sup_{p\leq n} \frac{1}{2} \sum_{1\leq k\leq d} |\eta_p^N(A_k) - \eta_p(A_k)| \geq \epsilon \right) \leq -\epsilon^2/\beta(\Phi)$$

  with $\beta(\Phi)$ related to the contraction properties of the sg $\Phi_{p,n}(\eta_p) = \eta_n$.

- Non asymptotic estimates

  [joint wortks with M. Ledoux (JTP 00), A. Doucet & S.Singh *(Feynman-Kac path space models & Additive functionals)* (M2AN-10)]

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

Concentration inequalities $\sim$ stability prop. semigroups $\Phi_{p,n}(\eta_p) = \eta_n$

[+ E. Rio (HAL-INRIA 09 & AAP 10) $\supset$ FK and McKean-Vlasov type models]

$$\Phi_{p,n}(\eta) - \Phi_{p,n}(\mu) = [\eta - \mu] \underbrace{D_\mu \Phi_{p,n}}_{\text{1st order op.}} + \underbrace{\mathcal{R}_{p,n}(\eta,\mu)}_{\text{2nd order } (\eta-\mu)^{\otimes 2}(\ldots)}$$

[Theo:] $\forall x \geq 0$, the probability of the next events is $\geq 1 - e^{-x}$

$$(\eta_n^N - \eta_n)(f) \leq \frac{r_n}{N} \left(1 + \epsilon_0^{-1}(x)\right) + \overline{\sigma}_n^2 \; b_n^\star \; \epsilon_1^{-1}\left(\frac{x}{N\overline{\sigma}_n^2}\right) \quad \text{[Bennett } (r=0)]$$

$$(\eta_n^N - \eta_n)(f) \leq \frac{r_n}{N} \left(1 + \epsilon_0^{-1}(x)\right) + \sqrt{\frac{2x}{N}} \; \beta_n \quad \text{[Hoeffding } (r=0)]$$

with

$$\epsilon_0(\lambda) = \frac{1}{2}\left(\lambda - \log\left(1+\lambda\right)\right) , \quad \epsilon_1(\lambda) = (1+\lambda)\log\left(1+\lambda\right) - \lambda$$

Cts : $r_n \sim$ bias second order & $\overline{\sigma}_n, \beta_n, b_n^\star \sim$ Dobrushin coef. 1st order operators.

Note : Crude bounds $\epsilon_0^{-1}(x) \leq 2x + 2\sqrt{x}$, $\epsilon_1^{-1}(x) \leq x/3 + \sqrt{2x} \rightsquigarrow$ [Bernstein]

Application : *Time homogeneous Feynman-Kac models* (M, G)

$$(H)\ \exists m\ :\quad M^m(x,\cdot) \geq \epsilon\ M^m(y,\cdot) \quad \text{and} \quad \delta := \sup \prod_{0\leq p<m} \left(G(x_p)/G(y_p)\right) < \infty$$

[Corollary:]
$\forall x \geq 0$, $\forall$ time horizon n, the probability of the next events is $\geq 1 - e^{-x}$

$$\begin{aligned} (\eta_n^N - \eta_n)(f) &\leq \frac{4a}{N}\left(1 + \epsilon_0^{-1}(x)\right) + 8a\ \epsilon_1^{-1}\left(\frac{x}{4bN}\right) \\ (\eta_n^N - \eta_n)(f) &\leq \frac{4a}{N}\left(1 + \epsilon_0^{-1}(x)\right) + 2\sqrt{\frac{2bx}{N}} \end{aligned}$$

with $a \leq m\ (\delta/\epsilon)^5$ and $b \leq m\ (\delta/\epsilon)^4$.