Concentration inequalities for mean field particle models

P. Del Moral & E. Rio

INRIA Centre Bordeaux-Sud Ouest & Versailles University

BoSanTouVal, Le Teich, June 2009.

Some References:

DM, Rio E., Concentration Inequalities for Mean Field Particle Models, HAL-INRIA 6901 (2009)

Rio E., Local invariance principles and their applications to density estimation. Probability Theory and related Fields, vol. 98, 21-45 (1994).

Feynman-Kac formulae: Genealogical and interacting particle systems, Springer (2004)

DM, Doucet, Jasra. SMC Samplers. JRSS B (2006).
1 Introduction
- Some notation
- iid sequences & Markov models
- Mean field particle models
  - Nonlinear type Markov models
  - Mean field particle interpretation
2 Some applications
- McKean diffusion type model
- A McKean model of gases
- Feynman-Kac models
3 Stochastic perturbation analysis
- Local sampling gaussian fields
- A local transport formulation
- First order analysis
4 Fluctuations and Concentration properties
- A functional fluctuation theorem
- Concentration inequalities
Outline

1 Introduction
- Some notation
- iid sequences & Markov models
- Mean field particle models
  - Nonlinear type Markov models
  - Mean field particle interpretation

2 Some applications

3 Stochastic perturbation analysis

4 Fluctuations and Concentration properties
Notation

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

- $(\mu, f) \in \mathcal{P}(E) \times \mathcal{B}(E) \quad \rightarrow \quad \mu(f) = \int f(x) \mu(dx)$

- $M(x, dy)$ integral operator on $E$

\[
M(f)(x) = \int M(x, dy) f(y)
\]
\[
[\mu M](dy) = \int \mu(dx) M(x, dy) \iff [\mu M](f) = \mu[M(f)]
\]

- Boltzmann-Gibbs transformation : $G : E \rightarrow [0, 1]$ with $\mu(G) > 0$

\[
\Psi_G(\mu)(dx) = \frac{1}{\mu(G)} G(x) \mu(dx)
\]
Linear Markov models \([\text{Time } n \in \mathbb{N}, K_n(x, dy) \text{ Markov transitions }]\)

Markov chain states \(X_n\) with transitions \(K_n\):

\[
\eta_n = \eta_{n-1} K_n = \text{Law}(X_n)
\]

- **Law of large numbers** \(\rightsquigarrow\) iid copies \((X_n^i)_{i \geq 1}\)

\[
\eta_n^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{X_n^i} \overset{N \uparrow \infty}{\sim} \eta_n
\]

- **Time homogeneous models** \(K_n \equiv K \overset{\sim}{\rightsquigarrow} \text{Occupation measures}\)

\[
\widehat{\eta}_n := \frac{1}{n} \sum_{p=1}^{n} \delta_{X_p} \overset{n \uparrow \infty}{\sim} \eta_{\infty} = \eta_{\infty} K
\]

\(\rightsquigarrow\) Concentration inequalities for iid sequences or for Markov models
Nonlinear/nonhomogeneous Markov models

\( K_n, \eta(x, dy) \) collection of Markov transitions \( \sim \eta \) probability meas.

\[ \sim \eta_n = \eta_{n-1} K_n, \eta_{n-1} = \text{Law}(X_n) \in \mathcal{P}(E_n) \]

i.e. :

\[ \mathbb{P}(X_n \in dx_n \mid X_{n-1}) = K_n, \eta_{n-1}(X_{n-1}, dx_n) \quad \text{with} \quad \text{Law}(X_{n-1}) = \eta_{n-1} \]

McKean measures :

\[ \mathbb{P}((X_0, X_1, \ldots, X_n) \in d(x_0, x_1, \ldots, x_n)) \]

\[ = \eta_0(dx_0)K_1, \eta_0(x_0, dx_1) \ldots K_n, \eta_{n-1}(x_{n-1}, dx_n) \]
Mean field particle interpretation

- **Objective**: Markov chain $\xi_n = (\xi_n^1, \ldots, \xi_n^N) \in E^N_n$ s.t.

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

- **Particle approximation transitions** ($\forall 1 \leq i \leq N$)

  $$
  \xi_{n-1}^i \rightsquigarrow \xi_n^i \sim K_{n, \eta_{n-1}^N}(\xi_{n-1}^i, dx_n)
  $$
Discrete generation mean field particle model

Schematic picture: \( \xi_n \in E_n^N \sim \xi_{n+1} \in E_{n+1}^N \)

\[
\begin{array}{cccc}
\xi_n^1 & \xrightarrow{K_{n+1,\eta_n}} & \xi_{n+1}^1 \\
\vdots & & \vdots \\
\xi_i^1 & \xrightarrow{\delta_{\xi_i}^n} & \xi_i^1 \\
\vdots & & \vdots \\
\xi_n^N & \xrightarrow{\delta_{\xi_n}^N} & \xi_{n+1}^N \\
\end{array}
\]

Rationale:

\[
\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_i}^n \sim_{N \uparrow \infty} \eta_n \implies K_{n+1,\eta_n^N} \sim_{N \uparrow \infty} K_{n+1,\eta_n}
\]

\[
\implies \xi_i^1 \text{ almost iid copies of } \bar{X}_n
\]

Concentration/fluctuations properties: \( \eta_n^N \sim_{N \uparrow \infty} \eta_n \)
Outline

1 Introduction

2 Some applications
   - McKean diffusion type model
   - A McKean model of gases
   - Feynman-Kac models

3 Stochastic perturbation analysis

4 Fluctuations and Concentration properties
Nonlinear models with Gaussian transitions

McKean diffusion type model

\[ \eta_{n+1} := \eta_n K_{n+1, \eta_n} \]

with

\[ K_{n, \eta}(x, dy) = \frac{1}{\sqrt{(2\pi)^d \det(Q_n)}} \exp \left\{ -\frac{1}{2} (y - d_n(x, \eta))' Q_n^{-1} (y - d_n(x, \eta)) \right\} dy, \]

\[ \lvert \rvert \]

\[ \overline{X}_{n+1} := d_{n+1} (\overline{X}_n, \eta_n) + W_{n+1} \quad \text{with} \quad \eta_n = \text{Law}(\overline{X}_n) \]
A McKean model of gases

Collision-jump type model

\[ \eta_{n+1} := \eta_n K_{n+1, \eta_n} \]

with

\[ K_{n+1, \eta}(x, dy) = \int \nu_n(ds) \eta(du) a_n(s, u) M_{n+1}((s, x), dy) \]

Example:

\( \nu \) counting on \( \{-1, +1\} \), \( a_n(s, u) = 1_s(u) \), \( M_{n+1}((s, x), dy) = \delta_{sx}(dy) \)

\[ \Downarrow \]

\[ K_{n+1, \eta}(x, dy) = \eta(1) \delta_x(dy) + \eta(-1) \delta_{-x}(dy) \]
Updating-prediction transformations

$M_n(x, dy)$ Markov transitions and $G_n : E \rightarrow [0, 1]$

\[ \eta_{n+1} = \Phi_{n+1}(\eta_n) := \Psi G_n(\eta_n) M_{n+1} \] (1)

Markov chain $X_n$ with transitions $M_n$ and initial condition $X_0 \approx \eta_0$:

\[ (1) \iff \eta_n(f) \propto \mathbb{E} \left( f(X_n) \prod_{0 \leq p < n} G_p(X_p) \right) \]

~ Nonlinear flow $\eta_{n+1} := \eta_n K_{n+1, \eta_n}$

- **Nonlinear Markov models**:

  \[ K_{n+1, \eta_n}(x, dz) = \int S_{n, \eta_n}(x, dy) M_{n+1}(y, dz) \]

  \[ S_{n, \eta_n}(x, dy) := G_n(x) \delta_x(dy) + (1 - G_n(x)) \Psi G_n(\eta_n)(dy) \]
Mean field genetic type particle model:

\[
\begin{pmatrix}
\xi_1^n \\
\vdots \\
\xi_i^n \\
\vdots \\
\xi_N^n
\end{pmatrix}
\xrightarrow{S_{n,\eta_N^n}}
\begin{pmatrix}
\hat{\xi}_1^n \\
\vdots \\
\hat{\xi}_i^n \\
\vdots \\
\hat{\xi}_N^n
\end{pmatrix}
\xrightarrow{M_{n+1}}
\begin{pmatrix}
\xi_1^{n+1} \\
\vdots \\
\xi_i^{n+1} \\
\vdots \\
\xi_N^{n+1}
\end{pmatrix}
\]

Accept/Reject/Selection transition:

\[
S_{n,\eta_N^n}(\xi_i^n, dx) := \epsilon_n G_n(\xi_i^n) \delta_{\xi_i^n}(dx) + (1 - \epsilon_n G_n(\xi_i^n)) \sum_{j=1}^{N} \frac{G_n(\xi_j^n)}{G_n(\xi_n^k)} \delta_{\xi_j^n}(dx)
\]

Ex. : \(G_n = 1_A, \ \epsilon_n = 1 \Rightarrow G_n(\xi_i^n) = 1_A(\xi_i^n)\)

\(\Leftarrow\textbf{FK particle models} \subset \textit{sequential Monte Carlo, population Monte Carlo, genetic algorithms, particle filters, pruning, spawning, reconfiguration, quantum-diffusion Monte Carlo, go with the winner, etc.}\)
Outline

1. Introduction
2. Some applications
3. Stochastic perturbation analysis
   - Local sampling gaussian fields
   - A local transport formulation
   - First order analysis
4. Fluctuations and Concentration properties
A stochastic perturbation formulation

Nonlinear transport equation

\[ \eta_{n+1} = \Phi_{n+1}(\eta_n) := \eta_n K_{n+1, \eta_n} \]

\( \rightsquigarrow \) Mean field particle model

\[ W_n^N := \sqrt{N} \left[ \eta_n^N - \Phi_{n+1}(\eta_n^N) \right] := \frac{1}{\sqrt{N}} \sum_{1 \leq i \leq N} \left[ \delta_{\xi_i} - K_{n+1, \eta_n^N}(\xi_i^{n-1}, \cdot) \right] \]

\( \cong \) \( W_n \perp \) Gaussian fields

\( \Uparrow \)

Stochastic perturbation formulation

\[ \eta_n^N = \Phi_{n+1}(\eta_n^N) + \frac{1}{\sqrt{N}} W_n^N \]
A local transport formulation \( \Phi_{p,n} = \Phi_n \circ \Phi_{n-1} \circ \ldots \circ \Phi_{p+1} \)

\[
\eta_0 \rightarrow \eta_1 = \Phi_1(\eta_0) \rightarrow \eta_2 = \Phi_{0,2}(\eta_0) \rightarrow \ldots \rightarrow \Phi_{0,n}(\eta_0)
\]

\[
\eta_0^N \rightarrow \Phi_1(\eta_0^N) \rightarrow \Phi_{0,2}(\eta_0^N) \rightarrow \ldots \rightarrow \Phi_{0,n}(\eta_0^N)
\]

\[
\eta_1^N \rightarrow \Phi_2(\eta_1^N) \rightarrow \ldots \rightarrow \Phi_{1,n}(\eta_1^N)
\]

\[
\eta_2^N \rightarrow \ldots \rightarrow \Phi_{2,n}(\eta_2^N)
\]

\[
\vdots
\]

\[
\eta_{n-1}^N \rightarrow \Phi_n(\eta_{n-1}^N)
\]

\[
\eta_n^N
\]

\[\eta_n^N - \eta_n = \sum_{q=0}^{n} \left[ \Phi_{q,n}(\eta_q^N) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N)) \right]\]
First order regularity properties $\supset$ previous examples

\[(H) \quad \Phi_n(\mu) - \Phi_n(\eta) = (\mu - \eta) \underbrace{D_{n,\eta}\Phi_n}_{\text{Integral operator}} + \underbrace{\mathcal{R}^{\Phi_n}(\mu, \eta)}_{\text{second order measure}}\]

$\Downarrow$

\[\Phi_{p,n}(\eta) - \Phi_{p,n}(\mu) = [\eta - \mu]D_\mu \Phi_{p,n} + \mathcal{R}^{\Phi_{p,n}}(\eta, \mu)\]

$\Downarrow$

\[V_{n}^N := \sqrt{N} \left[ \eta_{n}^N - \eta_n \right]\]

\[= \sum_{q=0}^{n} \left[ \Phi_{q,n}(\eta_{q}^N) - \Phi_{q,n}(\Phi_{q}(\eta_{q-1}^N)) \right] \simeq \sum_{q=0}^{n} W_{p}^N D_{p,n}\]

for some integral operator $D_{p,n}$ that enters the stability properties of $\Phi_{p,n}$

$\rightsquigarrow \quad \text{osc}(D_{p,n}(f)) \leq \beta(D_{p,n}) \quad \text{osc}(f) \leq e^{-\lambda(n-p)} \text{ osc}(f)$

Dobrushin contraction coef.
Outline

1. Introduction
2. Some applications
3. Stochastic perturbation analysis
4. Fluctuations and Concentration properties
   - A functional fluctuation theorem
   - Concentration inequalities
A functional Central Limit Theorem

We have the convergence in law of the random fields on $\mathcal{B}(E_n)$ (as $N \uparrow \infty$):

$$V_n^N = \sqrt{N} \left[ \eta^N_n - \eta_n \right] \rightarrow V_n := \sum_{q=0}^{n} W_p D_{p,n}$$

with the independent centred gaussian fields $W_n$ s.t.

$$\mathbb{E}(W_n(f) W_n(g)) = \eta_{n-1} K_{n,\eta_{n-1}} ([f - K_{n,\eta_{n-1}}(f)][g - K_{n,\eta_{n-1}}(g)]).$$

$\rightsquigarrow$ 2nd notation:

$\sigma_n^2$ uniform local variance parameter $f = g \in \mathcal{B}(E)$ s.t. $\text{osc}(f) \leq 1$
Concentration inequalities

- Feynman-Kac models:
  - (DM, FK Springer 2004) Using Kintchine’s type $\mathbb{L}_p$-estimates
    \[
    \mathbb{P} \left( |\eta_n^N(f) - \eta_n(f)| > \epsilon \right) \leq (1 + c_0 \epsilon \sqrt{N}) \, e^{-N\epsilon^2/c_1(n)}
    \]
  - (DM, A. Doucet, A. Jasra, Hal-INRIA pub. nb. 6700 (2008))
    \[
    \mathbb{P} \left( |[\eta_n^N - \eta_n](f)| \geq \epsilon \right) \leq 6 \, e^{-N\epsilon^2/c_1(n)}
    \]
  
with $\epsilon \in [0, 1/2]$ and $c_1(n)$ related to the Dobrushin contraction coefficient of

\[
P_{p,n}(f) := \frac{Q_{p,n}(f)}{Q_{p,n}(1)} \quad \text{with} \quad Q_{p,n}(f)(x_p) := \mathbb{E}_{p,x_p} \left( f(X_n) \prod_{p \leq p < n} G_q(X_p) \right)
\]

In addition: $\Phi_{p,n}$ "stable" $\Rightarrow c_1 := \sup_{n \geq 0} c_1(n) < \infty$. 
Some observations

- Rather crude uniform concentration inequalities $\sim \mathbb{L}_p$-bounds
- Refinements $\Rightarrow$ Laplace estimates for the first and second order terms

$$
\sqrt{N} \left[ \eta_n^N - \eta_n \right] = \sum_{q=0}^{n} \left[ \Phi_{q,n}(\eta_q^N) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N)) \right]
$$

$$
\simeq \sum_{q=0}^{n} W_p^N \quad D_{p,n} + \text{second order measures}
$$

$\sim W_p$ independent

Proof idea:

- First order term $\Rightarrow$ Bennett or Hoeffding’s inequalities.
- Second order-Bias type term $\Rightarrow$ "sharp" $\mathbb{L}_{2p}$-Kintchine’s bounds.
- First $\oplus$ Second order terms
  $\Rightarrow$ Rio’s lemma on the inverse of the Legendre transform of sums.
Concentration for general models [DM, Rio Hal-INRIA, 6901, (2009)]

Notation: \((\sigma_p \text{ uniform local variance, } \mathcal{D}_{p,n} \text{ first order integral operator})\)

\[
\overline{\sigma_n^2} = \sum_{p=0}^{n} \sigma_p^2 \beta(\mathcal{D}_{p,n})^2 \leq \beta_n^2 = \sum_{p=0}^{n} \beta(\mathcal{D}_{p,n})^2 \quad \text{and} \quad b_n^* = \sup_{0 \leq p \leq n} \beta(\mathcal{D}_{p,n}).
\]

Theorem: \(\forall x \geq 0\) the probability of each of the following pair of events

\[
[\eta_n^N - \eta_n](f_n) \leq \frac{r_n}{\mathcal{N}} \left( 1 + \epsilon_0^{-1}(x) \right) + \overline{\sigma_n^2} \ b_n^* \ \epsilon_1^{-1} \left( \frac{x}{N\overline{\sigma_n^2}} \right)
\]

is greater than \(1 - e^{-x}\), with \(r_n \sim \text{second order remainder measures}\) and

\[
\epsilon_0(\lambda) = \frac{1}{2} (\lambda - \log (1 + \lambda)) , \quad \epsilon_1(\lambda) = (1 + \lambda) \log (1 + \lambda) - \lambda
\]
Some “direct” consequences

Notice that \((\epsilon_0, \epsilon_1) = (\alpha_0^*, \alpha_1^*)\) with

\[
\alpha_0(t) := -t - \frac{1}{2} \log (1 - 2t) \quad \text{and} \quad \alpha_1(t) := e^t - 1 - t
\]

\(\downarrow\)

**Corollary 1 [Bernstein type inequalities]**:

\[
- \frac{1}{N} \log \mathbb{P} \left( [\eta_n^N - \eta_n](f_n) \geq \frac{r_n}{\sqrt{N}} + \lambda \right)
\]

\[
\geq \frac{\lambda^2}{2} \left( \left( b_n^* \sigma_n + \frac{\sqrt{2} r_n}{\sqrt{N}} \right)^2 + \lambda \left( 2r_n + \frac{b_n^*}{3} \right) \right)^{-1}
\]

and

\[
- \frac{1}{N} \log \mathbb{P} \left( [\eta_n^N - \eta_n](f_n) \geq \frac{r_n}{\sqrt{N}} + \lambda \right) \geq \frac{\lambda^2}{2} \left( \left( \beta_n + \frac{\sqrt{2} r_n}{\sqrt{N}} \right)^2 + 2r_n \lambda \right)^{-1}.
\]
Feynman-Kac models s.t. $\Phi_n$ ”stable”

Time homogeneous models s.t. $\sup_{x,y} G(x)/G(y) < \infty$ and

$$(M)_m \ \exists m \geq 1 \ \exists \epsilon_m > 0 \ \text{s.t.} \ \forall (x, y) \in E^2 \ \ M^m(x, \cdot) \geq \epsilon_m \ M^m(y, \cdot).$$

Notation :

$$\delta_m := \sup_{0 \leq p < m} \prod (G(x_p)/G(y_p)) \ \text{and} \ \varpi_{k,l}(m) \leq m \ \delta_{m-1} \ \delta_m^k/\epsilon_m^{k+2}$$

$$\Downarrow$$

$$r_n \leq 4 \ \varpi_{3,1}(m) \ \text{and} \ b_n^* \leq 2 \ \delta_m/\epsilon_m$$

as well as

$$\bar{\sigma}_n^2 \leq 4 \ \varpi_{2,2}(m) \ \sigma^2 \ \text{and} \ \beta_n^2 \leq 4 \ \varpi_{2,2}(m)$$
Corollary 2 [Uniform estimates w.r.t. time horizon]:

For any $n \geq 0$, and any $x \geq 0$ the probability of each of the following pair of events

$$[\eta_n^N - \eta_n](f_n)$$

$$\leq \frac{4}{N} \varpi_{3,1}(m) \left(1 + \epsilon_0^{-1}(x)\right) + \frac{8\delta_m}{\epsilon_m} \varpi_{2,2}(m) \sigma^2 \epsilon_1^{-1} \left(\frac{x}{4\sigma^2 \varpi_{2,2}(m) N}\right)$$

and

$$[\eta_n^N - \eta_n](f_n) \leq \frac{4}{N} \varpi_{3,1}(m) \left(1 + \epsilon_0^{-1}(x)\right) + 2\sqrt{\frac{2\varpi_{2,2}(m)x}{N}}$$

is greater than $1 - e^{-x}$. 