

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

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.

- $(\mu, f) \in \mathcal{P}(E) \times \mathcal{B}(E) \longrightarrow \mu(f) = \int f(x)\mu(dx)$
- $M(x, dy)$ **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 [Time $n \in \mathbb{N}$, $K_n(x, dy)$ 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 $(X_n^i)_{i \geq 1}$

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

- **Time homogeneous models** $K_n = K \rightsquigarrow$ Occupation measures

$$\hat{\eta}_n := \frac{1}{n} \sum_{p=1}^n \delta_{X_p} \simeq_{n \uparrow \infty} \eta_\infty = \eta_\infty K$$

\rightsquigarrow Concentration inequalities for iid sequences or for Markov models

Nonlinear/nonhomogeneous Markov models

K_n, η collection of Markov transitions $\sim \eta$ probability meas.

$$\rightsquigarrow \eta_n = \eta_{n-1} K_{n, \eta_{n-1}} = \text{Law}(\bar{X}_n) \in \mathcal{P}(E_n)$$

i.e. :

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

McKean measures :

$$\begin{aligned} & \mathbb{P}((\bar{X}_0, \bar{X}_1, \dots, \bar{X}_n) \in d(x_0, x_1, \dots, x_n)) \\ &= \eta_0(dx_0) K_{1, \eta_0}(x_0, dx_1) \dots K_{n, \eta_{n-1}}(x_{n-1}, dx_n) \end{aligned}$$

Mean field particle interpretation

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

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

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

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

Rationale :

$$\begin{aligned} \eta_n^N &:= \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq_{N \uparrow \infty} \eta_n \implies K_{n+1, \eta_n^N} \simeq_{N \uparrow \infty} K_{n+1, \eta_n} \\ &\implies \xi_n^i \text{ almost iid copies of } \bar{X}_n \end{aligned}$$

Concentration/fluctuations properties : $\eta_n^N \simeq_{N \uparrow \infty} \eta_n$??

- 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

$$K_{n, \eta}(x, dy)$$

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



$$\bar{X}_{n+1} := d_{n+1}(\bar{X}_n, \eta_n) + W_{n+1} \quad \text{with} \quad \eta_n = \text{Law}(\bar{X}_n)$$

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 :

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

↓

$$K_{n+1, \eta}(x, dy) = \eta(1) \delta_x(dy) + \eta(-1) \delta_{-x}(dy)$$

Updating-prediction transformations

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

$$\eta_{n+1} = \Phi_{n+1}(\eta_n) := \Psi_{G_n}(\eta_n) M_{n+1} \quad (1)$$

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

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

\rightsquigarrow 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 :

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

Accept/Reject/Selection transition :

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

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

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

\hookrightarrow **FK particle models** \supset sequential Monte Carlo, population Monte Carlo, genetic algorithms, particle filters, pruning, spawning, reconfiguration, quantum-diffusion Monte carlo, go with the winner, etc.

- 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

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

↪ Mean field particle model

$$\begin{aligned} W_n^N &:= \sqrt{N} [\eta_n^N - \Phi_{n+1}(\eta_n^N)] := \frac{1}{\sqrt{N}} \sum_{1 \leq i \leq N} [\delta_{\xi_n^i} - K_{n+1, \eta_n^N}(\xi_{n-1}^i, \cdot)] \\ &\simeq W_n \perp \text{Gaussian fields} \end{aligned}$$



Stochastic perturbation formulation

$$\eta_n^N = \Phi_{n+1}(\eta_n^N) + \frac{1}{\sqrt{N}} W_n^N$$

A local transport formulation $\Phi_{p,n} = \Phi_n \circ \Phi_{n-1} \circ \dots \circ \Phi_{p+1}$

$$\begin{array}{ccccccc}
 \eta_0 & \rightarrow & \eta_1 = \Phi_1(\eta_0) & \rightarrow & \eta_2 = \Phi_{0,2}(\eta_0) & \rightarrow & \dots \rightarrow \Phi_{0,n}(\eta_0) \\
 \Downarrow & & & & & & \\
 \eta_0^N & \rightarrow & \Phi_1(\eta_0^N) & \rightarrow & \Phi_{0,2}(\eta_0^N) & \rightarrow & \dots \rightarrow \Phi_{0,n}(\eta_0^N) \\
 & & \Downarrow & & & & \\
 & & \eta_1^N & \rightarrow & \Phi_2(\eta_1^N) & \rightarrow & \dots \rightarrow \Phi_{1,n}(\eta_1^N) \\
 & & & & \Downarrow & & \\
 & & & & \eta_2^N & \rightarrow & \dots \rightarrow \Phi_{2,n}(\eta_2^N) \\
 & & & & & & \Downarrow \\
 & & & & & & \eta_{n-1}^N \rightarrow \Phi_n(\eta_{n-1}^N) \\
 & & & & & & \Downarrow \\
 & & & & & & \eta_n^N
 \end{array}$$

\rightsquigarrow Key decomposition formula

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

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

$$\begin{aligned} V_n^N &:= \sqrt{N} [\eta_n^N - \eta_n] \\ &= \sum_{q=0}^n [\Phi_{q,n}(\eta_q^N) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))] \simeq \sum_{q=0}^n W_p^N \mathcal{D}_{p,n} \end{aligned}$$

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

$$\rightsquigarrow \text{osc}(\mathcal{D}_{p,n}(f)) \leq \underbrace{\beta(\mathcal{D}_{p,n})}_{\text{Dobrushin contraction coef.}} \quad \text{osc}(f) \leq e^{-\lambda(n-p)} \text{osc}(f)$$

- 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 $N \uparrow \infty$)

$$V_n^N = \sqrt{N} [\eta_n^N - \eta_n] \longrightarrow V_n := \sum_{q=0}^n W_q \mathcal{D}_{q,n}$$

with the independent centred gaussian fields W_n s.t.

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

↪ 2nd notation :

σ_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

$$\mathbb{P} (|\eta_n^N(f) - \eta_n(f)| > \epsilon) \leq (1 + c_0 \epsilon \sqrt{N}) e^{-N\epsilon^2/c_1(n)}$$

- (DM, A. Doucet, A. Jasra, Hal-INRIA pub. nb. 6700 (2008))

$$\mathbb{P} (|[\eta_n^N - \eta_n](f)| \geq \epsilon) \leq 6 e^{-N\epsilon^2/c_1(n)}$$

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 \rightsquigarrow Laplace estimates for the first **and** second order terms

$$\begin{aligned}\sqrt{N} [\eta_n^N - \eta_n] &= \sum_{q=0}^n [\Phi_{q,n}(\eta_q^N) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))] \\ &\simeq \sum_{q=0}^n \underbrace{W_p^N}_{\sim W_p \text{ independent}} \mathcal{D}_{p,n} + \text{second order measures}\end{aligned}$$

Proof idea :

- First order term \rightsquigarrow Bennett or Hoeffding's inequalities.
- Second order-Bias type term \rightsquigarrow "sharp" \mathbb{L}_{2p} -Kintchine's bounds.
- **First** \oplus **Second** order terms
 \rightsquigarrow Rio's lemma on the inverse of the Legendre transform of sums.

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

$$\bar{\sigma}_n^2 = \sum_{p=0}^n \sigma_p^2 \beta(\mathcal{D}_{p,n})^2 \leq \beta_n^2 = \sum_{p=0}^n \beta(\mathcal{D}_{p,n})^2 \quad \text{and} \quad b_n^* = \sup_{0 \leq p \leq n} \beta(\mathcal{D}_{p,n}).$$

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

$$[\eta_n^N - \eta_n](f_n) \leq \frac{r_n}{N} (1 + \epsilon_0^{-1}(x)) + \underbrace{\bar{\sigma}_n^2 b_n^* \epsilon_1^{-1} \left(\frac{x}{N \bar{\sigma}_n^2} \right)}_{\text{Bennett term}}$$

and

$$[\eta_n^N - \eta_n](f_n) \leq \frac{r_n}{N} (1 + \epsilon_0^{-1}(x)) + \underbrace{\sqrt{\frac{2x}{N}} \beta_n}_{\text{Hoeffding term}}$$

is greater than $1 - e^{-x}$, with $r_n \sim$ 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

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

↓

Corollary 1 [Bernstein type inequalities] :

$$\begin{aligned} & -\frac{1}{N} \log \mathbb{P} \left([\eta_n^N - \eta_n](f_n) \geq \frac{r_n}{N} + \lambda \right) \\ & \geq \frac{\lambda^2}{2} \left(\left(b_n^* \bar{\sigma}_n + \frac{\sqrt{2}r_n}{\sqrt{N}} \right)^2 + \lambda \left(2r_n + \frac{b_n^*}{3} \right) \right)^{-1} \end{aligned}$$

and

$$-\frac{1}{N} \log \mathbb{P} \left([\eta_n^N - \eta_n](f_n) \geq \frac{r_n}{N} + \lambda \right) \geq \frac{\lambda^2}{2} \left(\left(\beta_n + \frac{\sqrt{2}r_n}{\sqrt{N}} \right)^2 + 2r_n\lambda \right)^{-1}.$$

Feynman-Kac models s.t. Φ_n "stable"

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

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

Notation :

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

\Downarrow

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

as well as

$$\bar{\sigma}_n^2 \leq 4 \varpi_{2,2}(m) \sigma^2 \quad \text{and} \quad \beta_n^2 \leq 4 \varpi_{2,2}(m)$$

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

$$\begin{aligned} & [\eta_n^N - \eta_n](f_n) \\ & \leq \frac{4}{N} \varpi_{3,1}(m) (1 + \epsilon_0^{-1}(x)) + \frac{8\delta_m}{\epsilon_m} \varpi_{2,2}(m) \sigma^2 \epsilon_1^{-1} \left(\frac{x}{4\sigma^2 \varpi_{2,2}(m) N} \right) \end{aligned}$$

and

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

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