

Self Interacting Markov chains

P. Del Moral

Centre INRIA Bordeaux-Sud Ouest

Stochastic Analysis Seminar, Warwick University, december 08.

Some self-references :

- ↪ *A pair of joint works with Laurent Miclo : Stoch. Analysis and Appl. (2006) + Proc. Royal Soc. London A. (2003) ⊂ Toulouse Univ. (2002).*
- ↪ *Interacting MCMC, joint work with Brockwell & Doucet, UBC (2008).*
- ↪ *Fluctuations of Interacting MCMC, a series of joint works with Bercu & Doucet, HAL-INRIA preprints (2008).*

1 Introduction

- Self interacting processes in physics, biology and engineering
- Standard notation
- Self interacting processes
- A Toy model

2 Self Interacting sequences

3 Self interacting Markov chains

4 Interacting MCMC models

Self interacting processes in biology and engineering

- **Biology :**

- Genetic type reinforcement : "beneficial" interactions with the past.
- Historical and spatial interactions processes : ethology models, ant-and-bee systems, the particle swarm methods, and ant colony models, and others.

- **Stochastic engineering :**

- Re-initialization of random search models and stochastic algorithms.
- \rightsquigarrow *New class of stochastic algorithms* \subset Global optimization problems, interacting MCMC models for complex distribution flows, bayesian learning, filtering, and spectral analysis of Feynman-Kac-Schroedinger operators.

Some important questions

- Stochastic models & Rigorous stochastic analysis.
- \uparrow Application model areas.

Standard notation

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

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

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

$$[\mu M](dy) = \int \mu(dx) M(x, dy) \quad (\implies [\mu M](f) = \mu[M(f)])$$

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

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

Self interacting sequences

E-valued random sequence $(X_n)_{n \geq 0}$ s.t.

$$\mathbb{P}(X_n \in dx_n \mid X_0, \dots, X_{n-1}) = \Phi \left(\frac{1}{n} \sum_{0 \leq p < n} \delta_{X_p} \right) (dx_n)$$

with $\Phi : \nu \in \mathcal{P}(E) \rightarrow \Phi(\nu) \in \mathcal{P}(E)$ & $\text{Law}(X_0) = \eta \in \mathcal{P}(E)$.

⊂ Self Interacting Markov Chain models

E-valued random sequence $(X_n)_{n \geq 0}$ s.t.

$$\mathbb{P}(X_n \in dx_n \mid X_0, \dots, X_{n-1}) = K_{\frac{1}{n} \sum_{0 \leq p < n} \delta_{X_p}}(x_{n-1}, dx_n)$$

with

- A collection of Markov transitions $K_\eta(x, dy)$ with $\eta \in \mathcal{P}(E)$
- An initial distribution $\text{Law}(X_0) = \eta \in \mathcal{P}(E)$.

Some questions :

- 1 Motivations, application areas ?
- 2 Toy examples ?
- 3 \exists Conditions ? s.t. $\frac{1}{n} \sum_{0 \leq p < n} \delta_{X_p} \rightarrow_{n \rightarrow \infty} ?$
- 4 If any convergence \rightsquigarrow decay rates ?
- 5 Links with Vertex/Edge-Reinforced type Random Walks on \mathbb{Z}^d ?

Links with Vertex/Edge-Reinforced type Random Walks on $E = \mathbb{Z}^d$

- Expert in Oxford = Pierre Tarrès

⊕ *M. Benaim, P. Diaconis, R. Pemantle, S. Volkov,...*

- Non compact & "degenerate" integer lattice models

$$K_{\frac{1}{n} \sum_{0 \leq p < n} \delta_{x_p}}(x_{n-1}, x_n) \propto M(x_{n-1}, x_n) \times \left(a_n + \sum_{0 \leq p < n} 1_{x_p}(x_n) \right)$$

- Our framework :
 - More regular class of transitions.
 - Evolutionary & genetic mutation-selection models.
 - \rightsquigarrow Self-interacting quantum and diffusion Monte Carlo.
 - Stochastic sampling and MCMC application domain.

A Toy model [μ a fixed probability meas.]

$$\Phi(\eta) := \epsilon \eta + (1 - \epsilon) \mu \quad \text{with} \quad \epsilon \in [0, 1]$$

\Downarrow

$$\mathbb{P}(X_n \in dx_n \mid X_0, \dots, X_{n-1}) = \epsilon \frac{1}{n} \sum_{0 \leq p < n} \delta_{X_p}(dy) + (1 - \epsilon) \mu(dy)$$

\Downarrow

μ -i.i.d. sequences with ϵ -repetitions

\Downarrow

We expect : $\frac{1}{n} \sum_{0 \leq p < n} \delta_{X_p} \xrightarrow{n \rightarrow \infty} \mu$, as soon as $\epsilon \neq 1$

Mean value elementary estimate

$$S_n := \frac{1}{n+1} \sum_{0 \leq p \leq n} \delta_{X_p} = \frac{n}{n+1} S_{n-1} + \frac{1}{n+1} \delta_{X_{n+1}}$$

$$\Downarrow ([\mu(f) = 0])$$

$$\bar{S}_n(f) := \mathbb{E}(S_n(f)) = \left[\frac{n+\epsilon}{n+1} \right] \bar{S}_{n-1}(f) = a(n) \eta(f)$$

with

$$\frac{1}{2e^{n^{1-\epsilon}}} \leq a(n) := \left(\prod_{0 < p \leq n} \frac{p+\epsilon}{p+1} \right) \leq \frac{2}{n^{1-\epsilon}}$$

Explicit analytic expansion

$$\mathbb{E} \left([S_n(f) - \bar{S}_n(f)]^2 \right)^{\frac{1}{2}} \simeq \begin{cases} 1/\sqrt{n} & \text{if } \epsilon \in [0, \frac{1}{2}[\\ \sqrt{(\log n)/n} & \text{if } \epsilon = \frac{1}{2} \\ 1/n^{(1-\epsilon)} & \text{if } \epsilon \in]\frac{1}{2}, 1] \end{cases}$$

and

$$\|\bar{S}_n - \mu\|_{tv} \simeq 1/n^{(1-\epsilon)}$$

Note

$$\Phi(\eta) := \epsilon \eta + (1 - \epsilon) \mu \quad \text{with } \epsilon \in [0, 1]$$

↓

$\epsilon = \Phi$ -contraction coeff. :

$$[\Phi(\eta_1) - \Phi(\eta_2)] = \epsilon [\eta_1 - \eta_2]$$

- 1 Introduction
- 2 Self Interacting sequences
 - Description of the models
 - Evolutionary interaction models
 - A convergence theorem
 - Ground state energies
- 3 Self interacting Markov chains
- 4 Interacting MCMC models

Self Interacting sequences

$$\mathbb{P}(X_n \in dx_n \mid X_0, \dots, X_{n-1}) = \Phi \left(\frac{1}{n} \sum_{0 \leq p < n} \delta_{X_p} \right) (dx_n)$$

Genetic interaction model

$M(x, dy)$ Markov transition on E & a potential function $G : E \rightarrow [0, \infty[$

$$\Phi(\mu) := \Psi_G(\mu)M \quad \left(\text{with } \Psi_G(\mu)(dx) = \frac{1}{\mu(G)} G(x) \mu(dx) \right)$$

\Downarrow

$$\Phi \left(\frac{1}{n} \sum_{0 \leq p < n} \delta_{X_p} \right) := \sum_{0 \leq p < n} \frac{G(X_p)}{\sum_{0 \leq q < n} G(X_q)} M(X_p, \cdot)$$

\Downarrow

Genetic reinforced model : Past history selection \oplus M -free exploration

(H) : $\mathcal{B}(E) \supset \mathcal{F} \ni f, \|f\| \leq 1, d_{\mathcal{F}}(\mu, \eta) = \sup_{f \in \mathcal{F}} \|[\mu - \eta](f)\|$ complete and

$$|[\Phi(\eta) - \Phi(\mu)](f)| \leq \int |[\eta - \mu](g)| \Gamma_{\mu}(f, dg)$$

with

$$\beta(\Phi) := \sup_{\mu \in \mathcal{P}(E), f \in \mathcal{F}} \Gamma_{\mu}(f, 1) < 1$$

Example

Genetic type (G, M) -transformations :

$$\Phi(\mu) := \Psi_G(\mu)M \quad \text{with} \quad \mathcal{F} := \{f : \text{osc}(f) \leq 1\}$$

$$\beta(\Phi) \leq \left[2 \left(\frac{G_{\max}}{G_{\min}} \right) - 1 \right] \beta(M)$$

with the Dobrushin coefficient :

$$\beta(M) := \sup_{x, y} \|M(x, \cdot) - M(y, \cdot)\|_{tv} := \sup_{f \in \mathcal{F}} \text{osc}(M(f))$$

Th 1 : [DM & Miclo [2002]]

(H) :

$$\exists! \mu = \Phi(\mu)$$

and the three types of decay rates :

$$\mathbb{E} \left([S_n(f) - \mu]^2 \right)^{\frac{1}{2}} \leq \begin{cases} \frac{1}{\sqrt{n}} & \text{if } \beta(\Phi) \in [0, \frac{1}{2}[\\ \sqrt{\frac{(\log n)}{n}} & \text{if } \beta(\Phi) = \frac{1}{2} \\ \frac{1}{n^{(1-\beta(\Phi))}} & \text{if } \beta(\Phi) \in]\frac{1}{2}, 1] \end{cases}$$

Remarks :

- $\rightsquigarrow \mathbb{L}_p$ -convergence rates (without the log scales) in DM & Miclo [2003].
- **Alternative approach** \rightsquigarrow **Measure valued processes**

$$\eta_n = \Phi(\eta_{n-1}) \xrightarrow{n \rightarrow \infty} \mu = \Phi(\mu)$$

\Downarrow

Mean Field particle approximations = Spatial interaction processes

$$(X_n^{(1,N)}, X_n^{(2,N)}, \dots, X_n^{(N,N)}) \text{ "iid" } \sim \Phi \left(\frac{1}{N} \sum_{1 \leq i \leq N} \delta_{X_{n-1}^{(i,N)}} \right)$$

[Top eigenvalues and ground states] (ν -reversible $M \oplus$ some reg.)

$$\Phi(\eta)(f) = \frac{\eta(Q(f))}{\eta(Q(1))} \quad \text{with} \quad Q(f) = G M(f)$$

\Downarrow

Top eigenvalues $\lambda \oplus$ Top eigenvector $Q(h) = \lambda h$

$$\mu(dx) = \Psi_h(\nu)M \quad \text{and} \quad \Psi_G(\mu) = \Psi_h(\nu)$$

$$\begin{aligned} \text{Proof: } \Phi(\mu)(f) \propto \mu(GM(f)) &= \nu(hM[GM(f)]) \\ &= \nu(Q(h)M(f)) \propto \nu(hM(f)) \propto \mu(f) \end{aligned}$$

- \rightsquigarrow **Note** : \supset Mean field particle model = QMC & DMC approximations
- \rightsquigarrow Self interacting genetic model \simeq Self-interacting QMC & DMC

- 1 Introduction
- 2 Self Interacting sequences
- 3 Self interacting Markov chains
 - Description of the models
 - A convergence theorem
- 4 Interacting MCMC models

Self interacting Markov chains

$$\mathbb{P}(X_n \in dx_n \mid X_0, \dots, X_{n-1}) = K_{\frac{1}{n} \sum_{0 \leq p < n} \delta_{X_p}}(x_{n-1}, dx_n)$$

with

- A collection of Markov transitions $K_\eta(x, dy)$ with $\eta \in \mathcal{P}(E)$
- An initial distribution $\text{Law}(X_0) = \eta \in \mathcal{P}(E)$.

Example : Genetic type mutation/selection transitions

$$K_\eta(x, dy) := \epsilon M(x, dy) + (1 - \epsilon) \eta$$

or

$$K_\eta(x, dy) := \epsilon M(x, dy) + (1 - \epsilon) \Psi_G(\eta)$$

and more generally

$$K_\eta(x, dy) := \epsilon(\eta) M(x, dy) + (1 - \epsilon(\eta)) \Phi(\eta)$$

(H') : $\mathcal{F} \subset \mathcal{B}(E)$ as above s.t. :

$$|[\eta_1 - \eta_2] K_\mu(f)| \leq \int |[\eta_1 - \eta_2](g)| \Gamma_\mu^{(1)}(f, dg) \quad \text{[Contraction prop.]}$$

$$|\mu [K_{\eta_1} - K_{\eta_2}](f)| \leq \int |[\eta_1 - \eta_2](g)| \Gamma_{\eta_1}^{(2)}(f, dg) \quad \text{[Regularity prop.]}$$

with

$$\lambda^{(1)} + \lambda^{(2)} < 1 \quad \text{and} \quad \lambda^{(i)} := \sup_{\mu \in \mathcal{P}(E), f \in \mathcal{F}} \Gamma_\mu^{(i)}(f, 1), \quad i = 1, 2$$

Th 2 : [weaker version of DM & Miclo [2003]]

(H') :

$$\exists! \mu = \mu K_\mu$$

and the three types of decay rates : $\Lambda := \lambda_2 / (1 - \lambda_1)$

$$\mathbb{E} \left([S_n(f) - \mu]^2 \right)^{\frac{1}{2}} \leq \begin{cases} \frac{1}{\sqrt{n}} & \text{if } \Lambda \in [0, \frac{1}{2}[\\ \sqrt{\frac{(\log n)}{n}} & \text{if } \Lambda = \frac{1}{2} \\ \frac{1}{n^{(1-\Lambda)}} & \text{if } \Lambda \in]\frac{1}{2}, 1] \end{cases}$$

Extensions/Proof ideas :

- Under very weak conditions (\oplus Resolvent estimates \oplus martingales \leq) :

$$\sup_{n \geq 1} \sqrt{n} \times \mathbb{E} \left([S_n(f) - S_n^\omega(f)]^2 \right)^{\frac{1}{2}} < \infty \quad \text{with} \quad S_n^\omega := \frac{1}{n} \sum_{0 \leq p < n} \omega(S_p)$$

- \oplus Weaker conditions \rightsquigarrow Fixed point mapping $\omega(\eta) = \omega(\eta)K_\eta$ analysis.
- \mathbb{L}_p -mean error estimates with the same decay rates.

- 1 Introduction
- 2 Self Interacting sequences
- 3 Self interacting Markov chains
- 4 Interacting MCMC models
 - Nonlinear distribution flows
 - A functional Central Limit Theorem

i-MCMC sampling of nonlinear m.v.p.

- **Problem** : sample series of random variables w.r.t. η_n with

$$\eta_n = \Phi_n(\eta_{n-1})$$

~ Cooling schemes, temp. variations, constraints sequences, subset restrictions, observation data, conditional events, Feynman-Kac flows:

$$\eta_n(dx) \propto e^{-\beta_n V(x)} \lambda(dx), \mathbf{1}_{A_n}(x) \lambda(dx), \mathbb{P}(X_n \in dx \mid Y_0, \dots, Y_n), \dots$$

- **Natural interacting sampling idea** :

Use η_{n-1} or its empirical approx. to sample w.r.t. η_n

\rightsquigarrow **Interacting MCMC models** :

$$\left\{ \begin{array}{l} \text{Use the occupation measures} \\ \text{of an MCMC with target } \eta_{n-1} \end{array} \right\} \rightsquigarrow \text{MCMC target } \eta_n$$

The key natural idea

Find a series of MCMC models $X^{(n)} := (X_k^{(n)})_{k \geq 0}$ s.t.

$$\frac{1}{k+1} \sum_{0 \leq l \leq k} \delta_{X_l^{(n)}} := \eta_k^{(n)} \simeq_{k \uparrow \infty} \eta_n$$

↓

Use $\eta_k^{(n)} \simeq \eta_n$ to define $X^{(n+1)}$ with target η_{n+1}

⇕

Find a collection of Markov transitions $M_{n+1, \eta}$ s.t. given the flow $(\eta_l^{(n)})_{l \leq k}$

$$\mathbb{P} \left(X_{k+1}^{(n+1)} \in dy \mid X_k^{(n+1)} = x \right) = M_{n+1, \eta_k^{(n)}}(x, dy)$$

$((n-1)$ -th chain)

$$\begin{array}{c} X_0^{(n-1)} \\ \downarrow \\ X_1^{(n-1)} \\ \downarrow \\ \vdots \\ \downarrow \\ X_k^{(n-1)} \\ \downarrow \\ \vdots \end{array}$$

$$\xrightarrow{\eta_k^{(n-1)} \simeq \eta_{n-1}}$$

$(n$ -th chain)

$$\begin{array}{c} X_0^{(n)} \\ \downarrow \\ \vdots \\ \downarrow \\ \vdots \\ \downarrow \\ X_k^{(n)} \\ \downarrow \\ M_{n,\eta_k^{(n-1)}} \simeq M_{n,\eta_{n-1}} \\ \downarrow \\ X_{k+1}^{(n)} \end{array}$$

Key condition : $M_{n,\eta}(x, dy)$ with invariant measure $\Phi_n(\eta)$

$$\Rightarrow \Phi_n(\eta_k^{(n-1)}) \simeq \Phi_n(\eta_{n-1}) = \eta_n$$

Advantages

- Using η_n the sampling η_{n+1} is often easier.
- Improve the proposition step in any Metropolis type model with target η_{n+1} (\rightsquigarrow enters the stability prop. of the flow η_n)
- Increases the precision at every time step.
But CLT variance often \geq CLT variance mean field models.
- Easy to combine with mean field stochastic algorithms :
 \rightsquigarrow a natural technique to refine and improve the convergence of a given sampled mean field particle algorithm.

Interacting Markov chain Monte Carlo models

- Find M_0 and a collection of transitions $M_{n,\mu}$ s.t.

$$\eta_0 = \eta_0 M_0 \quad \text{and} \quad \Phi_n(\mu) = \Phi_n(\mu) M_{n,\mu}$$

- $(X_k^{(0)})_{k \geq 0}$ Markov chain $\sim M_0$.
- Given $X^{(n)}$, we let $X_k^{(n+1)}$ with Markov transitions $M_{n+1,\eta_k^{(n)}}$

Rationale :

$$\eta_k^{(n)} \simeq \eta_n \quad \Rightarrow \quad \left(M_{n+1,\eta_k^{(n)}} \simeq M_{n+1,\eta_n} \right) \implies \eta_k^{(n+1)} \simeq \Phi_n(\eta_n) = \eta_{n+1}$$

Example : $M_{n,\mu}(x, \cdot) = \Phi_n(\mu) \rightsquigarrow X_k^{(n+1)}$ random var. $\sim \Phi_{n+1} \left(\eta_k^{(n)} \right)$

Functional Central Limit Theorem :

$$\sqrt{k} \left[\eta_k^{(n)} - \eta_n \right] \xrightarrow{k \uparrow \infty} \sum_{q=0}^n \alpha_n(q) V_q D_{q,n}$$

- The parameters $\alpha_n(n-q) = \sqrt{(2q)!}/q!$ (**Mean field models** $\alpha_n(q) = 1$)
- \perp Centered Gauss. fields $(V_q)_{q \geq 0} \sim$ Fluctuations of the q -th MCMC sampler around the limiting measure η_q

$$\mathbb{E} (V_q(f)^2)$$

$$= \eta_q [(f - \eta_q(f))^2] + 2 \sum_{m \geq 1} \eta_q [(f - \eta_q(f)) M_{q, \eta_{q-1}}^m (f - \eta_q(f))]$$

- **First order decomposition of the (nonlinear) sg** $\rightarrow \Phi_{p,n}(\eta_p) = \eta_n$

$$\Phi_{p,n}(\eta) - \Phi_{p,n}(\mu) \simeq (\eta - \mu) D_{p,n} + (\eta - \mu)^{\otimes 2} \dots$$

\rightsquigarrow **Extensions** : Non asymptotic \mathbb{L}_p -estimates \oplus exponential concentration
 \subset Bercu, DM, Doucet (2008).