On the convergence and the application of evolutionnary type stochastic particle models

P. Del Moral

INRIA de Bordeaux - Sud Ouest Center

Workshop on Evolutionary Algorithms - Challenges in Theory and Practice

Some references :

- Particle Methods: An introduction with applications. HAL-INRIA RR-699 (2009).
- Feynman-Kac formulae. Genealogical and interacting particle approximations. Springer New York, Series: Probability and Applications (2004).
- A Backward Particle Interpretation of Feynman-Kac Formulae HAL-INRIA RR-7019 (2009).
- Concentration Inequalities for Mean Field Particle Models HAL-INRIA RR-6901(2009).
- On the Convergence and the Applications of the Generalized Simulated Annealing. SIAM Journal on Control and Optimization, Vol. 37, No. 4, 1222-1250, (1999).

Outline

1. A simple genetic particle model
2. Some convergence estimates
3. Some application domains

Summary

A simple genetic particle model $\hookrightarrow$ Only 2 ingredients

- Mutation= A Markov chain X_n on some state space E_n (n=time index).

$$\mathbb{P}\left(X_n \in dx_n \mid X_{n-1} = x_{n-1}\right) = M_n(x_{n-1}, dx_n)$$

Examples:

  - *Simple random walk on* $\mathbb{Z}^d$.
  - *Markov chain on graphs with neighbourhood exploration strategies.*
  - $X_n = F_n(X_{n-1}, W_n)$ *with i.i.d. perturbations/random controls* W_n.
  - *Transition space* $X_n = (X'_{n-1}, X'_n)$.
  - *Path or excursion state space models* $X_n := (X'_0, X'_1, \ldots, X'_n)$.

- Selection=Fitness type potential functions $G_n : x_n \in E_n \to G_n(x_n) \in [0,1]$

Examples:

  - *Indicator functions* $G_n(x_n) = 1_{A_n}(x_n)$.
  - *Boltzmann type exponentials* $G_n(x_n) = e^{-\beta_n V_n(x_n)}$.
  - *Likelihood functions* $G_n(x_n) = p(y_n|x_n) = (2\pi)^{-1/2}\ e^{-\frac{1}{2}(y_n - h_n(x_n))^2}$.
  - *On path space* $G_n(X'_0, X'_1, \ldots, X'_n) = 1_{E_n - \{X'_0, X'_1, \ldots, X'_{n-1}\}}(X'_n)$, ...

Genetic Mutation-Selection type stochastic algorithm

$$\left.\begin{array}{c}\xi_n^1\\ \vdots\\ \xi_n^i\\ \vdots\\ \xi_n^N\end{array}\right]\xrightarrow{\text{Selection w.r.t. } G_n}\left[\begin{array}{ccc}\widehat{\xi}_n^1 & \xrightarrow{\text{Mutation } M_{n+1}} & \xi_{n+1}^1\\ \vdots & & \vdots\\ \widehat{\xi}_n^i & \longrightarrow & \xi_{n+1}^i\\ \vdots & & \vdots\\ \widehat{\xi}_n^N & \longrightarrow & \xi_{n+1}^N\end{array}\right]$$

Acceptance/Rejection-Selection : [$\forall\ \epsilon_n$ that may depend on $(\xi_n^i)_{1\leq i\leq N}$]

$$\begin{array}{ccl}\widehat{\xi}_n^i & = & \xi_n^i \quad \text{with probability } \epsilon_n G_n(\xi_n^i)\\ \widehat{\xi}_n^i & = & \zeta_n^i \quad \text{with probability } 1-\epsilon_n G_n(\xi_n^i)\end{array}$$

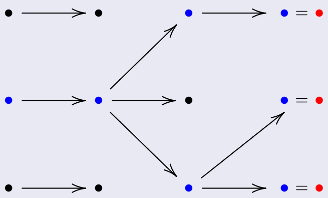
$$\zeta_n^i \text{ r.v. with law } \sum_{j=1}^N \frac{G_n(\xi_n^j)}{\sum_{k=1}^N G_n(\xi_n^k)}\delta_{\xi_n^j}(dx) \propto G_n(x)\times\frac{1}{N}\sum_{j=1}^N \delta_{\xi_n^j}(dx)$$

$\subset$ Mean field interacting particle system ($\subset$ Nonlinear Markov chain)

$$\text{Def. : Transitions } \xi_n^i \rightsquigarrow \xi_{n+1}^i \text{ depending on } \eta_n^N := \frac{1}{N}\sum_{j=1}^N \delta_{\xi_n^j} \simeq_{N\uparrow\infty} \text{Law}(\xi_n^i)$$

Interaction/branch. process $\hookrightarrow$ 3 types of occupation measures

$(N = 3)$



- Current population $\hookrightarrow \eta_n^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_n^i}$ ←*i*-th individual at time *n*

- Genealogical tree $\hookrightarrow \mathbb{Q}_n^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{(\xi_{0,n}^i,\xi_{1,n}^i,\ldots,\xi_{n,n}^i)}$ ←ancestral line of the *i*-th individual

- **Ancestral tree** $\hookrightarrow \mathbb{K}_n^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{(\xi_0^i,\xi_1^i,\ldots,\xi_n^i)}$ ←*i*-th ancestral line

- $\oplus$ Particle partition functions $\hookrightarrow \mathcal{Z}_n^N := \mathcal{Z}_{n-1}^N \times \underbrace{\left(\frac{1}{N}\sum_{i=1}^{N} G_n(\xi_n^i)\right)}_{\text{proportion of success}}$

Equivalent Stochastic Algorithms :

- Genetic and evolutionary type algorithms.
- 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

Limiting Feynman-Kac or Boltzmann-Gibbs measures

No choice! We have $\mathbb{Q}_n^N \simeq_{N\uparrow\infty} \mathbb{Q}_n$ and $\mathcal{Z}_n^N \simeq_{N\uparrow\infty} \mathcal{Z}_n$ and inversely!

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

Examples:

- Indicator functions $G_n(x_n) = 1_{A_n}(x_n) \rightsquigarrow$ Confinement models

$$\mathbb{Q}_n = \text{Law}((X_0, \ldots, X_n) \mid \forall 0 \leq p < n \;\; X_p \in A_p) \quad \mathcal{Z}_n = \mathbb{P}(\forall 0 \leq p < n \;\; X_p \in A_p)$$

- Likelihood functions $G_n(x_n) = p(y_n|x_n) \rightsquigarrow$ Filtering models

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

- $X_n = (X'_p)_{0\leq p\leq n}$, $G_n(X_n) = 1_{E_n - \{X'_0, X'_1, \ldots, X'_{n-1}\}}(X'_n) \rightsquigarrow$ Self-avoiding chains

$$\mathbb{Q}_n = \text{Law}((X'_0, X'_1, \ldots, X'_n) \mid \forall 0 \leq p < q < n \;\; X'_p = X'_q)$$

Summary

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

- Empirical processes (example $E_n = \mathbb{R}^d$ & $f_n \in \mathcal{F}_n = \{1_{[0,x]}\ ,\ x \in \mathbb{R}^d\}$)

$$\eta_n^N(f_n) := \frac{1}{N}\sum_{i=1}^N f_n(\xi_n^i) \simeq_{N\uparrow\infty} \int f_n(x_n)\ \eta_n(dx_n) := \eta_n(f_n)$$

$$\sup_{n\geq 0}\sup_{N\geq 1}\sqrt{N}\ \mathbb{E}\left(\sup_{f_n\in\mathcal{F}_n}|\eta_n^N(f_n)-\eta_n(f_n)|^p\right)^{1/p} \leq a(p)\ b$$

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

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

  $+$ *Guionnet* $\sup_{n\geq 0}$ `(IHP 01)` & *Ledoux* $\sup_{\mathcal{F}_n}$ `(JTP 00)` & *Rio* `hal-09`

- Same results for $\mathbb{Q}_n^N$ and $\mathbb{Q}_n$ BUT not uniform w.r.t. the time n and the constant b is now replaced by $b(n) = b \times n$.

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

- Propagations of chaos exansions *(+Patras,Rubenthaler* (AAP 09-10) *)* :

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

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

- *Normalizing Cts* (+Cerou & Guyader Hal-INRIA 08 & IPH 2010) *:*

$$\mathbb{E}(\mathcal{Z}^N_n) = \mathcal{Z}_n \quad \text{and} \quad \mathbb{E}\left(\left[1-\mathcal{Z}^N_n/\mathcal{Z}_n\right]^2\right) \leq\ c\ n/N$$

Note : Traditional Monte Carlo methods

$$\mathbb{E}\left(\left[1-\mathcal{Z}^N_n/\mathcal{Z}_n\right]^2\right) = \frac{1}{N}\,\frac{1-\mathcal{Z}_n}{\mathcal{Z}_n} \rightsquigarrow \mathrm{Pb\ if\ } \mathcal{Z}_n \simeq 10^{-n}$$

Summary

1. A simple genetic particle model

2. Some convergence estimates

3. Some application domains
   - Confinement, optimization, combinatorial pb, rare events.
   - Filtering $\leftrightarrows$ Regulation
   - Parameter estimation HMM
   - Additive functionals

Confinement, optimization, combinatorial pb, rare events.

1. $\eta_n = \mathrm{Loi}((X_0, \ldots, X_n) \mid \forall 0 \leq p \leq n \quad X_p \in A_p)$
2. $\eta_n(dx) \propto e^{-\beta_n V(x)} \lambda(dx)$ with $\beta_n \uparrow$
3. $\eta_n(dx) \propto 1_{A_n}(x) \lambda(dx)$ with $A_n \downarrow$
4. $\eta_n = \mathrm{Loi}_\pi((X_0, \ldots, X_n) \mid X_n = x_n)$.
5. $\eta_n = \mathrm{Loi}(\ X \text{ hits } B_n \mid X \text{ hits } B_n \text{ before } A)$

Stochastic particle algorithms

1. X_n-local moves $\oplus$ individual selections $\in A_n$ i.e. $\sim G_n = 1_{A_n}$
2. η_n-MCMC local moves (i.e. η_n-shakers) $\oplus$ selections $G_n = e^{-(\beta_{n+1} - \beta_n)V}$
3. η_n-MCMC local moves (i.e. η_n-shakers) $\oplus$ selections w.r.t. $G_n = 1_{A_{n+1}}$
4. X_n-local moves $\oplus$ Selection w.r.t. $G(x_1, x_2) = \frac{\pi(dx_2) M(x_2, dx_1)}{\pi(dx_1) M(x_1, dx_2)}$
5. X_n-local moves $\oplus$ Selection-branching on upper/lower levels B_n.

Filtering (max likelihood path) $\leftrightarrows$ Regulation

Signal-Observation $(X, Y) \rightsquigarrow$ reference path $Y = y$

$$X_n = F_n(X_{n-1}, W_n) \quad \text{and} \quad Y_n = H_n(X_n) + V_n$$

If W_n, V_n centered Gaussian r.v.

$$\begin{aligned}\mathbb{Q}_n &:= \mathrm{Law}((W_0, \ldots, W_n) \mid (Y_0, \ldots, Y_n) = (y_0, \ldots, y_n)) \\ &\propto \exp\left(-\frac{1}{2}\sum_{p=0}^{n} \|W_p\|^2 - \frac{1}{2}\sum_{p=0}^{n} \|y_n - H_n(X_n)\|^2\right)\end{aligned}$$

Genealogical tree associated with $(W_0, \ldots, W_n)$ = Control Tree Chart (open loop)

$$\mathbb{Q}_n^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{(W^i_{0,n}, W^i_{1,n}, \ldots, W^i_{n,n})}$$

Parameter estimation HMM

- Boltzmann-Gibbs measures

$$\eta_n(d\theta) \propto G_n(\theta)\ p(d\theta) \quad \text{with} \quad G_n(\theta) = G_{n-1}(\theta) \times g_n(\theta)$$

$\Longrightarrow$ η_n-MCMC mutations (shakers) $\oplus$ Selections w.r.t. g_n

- Example: Θ r.v. with law $p(d\theta)$ & Linear-Gaussian HMM model

$$X_n = A_n(\Theta)X_{n-1} + B_n(\Theta)W_n \quad \text{and} \quad Y_n = C_n(\Theta)X_n + D_n(\Theta)V_n$$

Given the data $Y = y$, if we set $g_n(\theta) = p(y_n|\theta, y_0, \ldots, y_{n-1})$, then we have

$$G_n(\theta) = p(y_0, \ldots, y_n|\theta) \quad \text{and} \quad \eta_n = \mathrm{Law}(\Theta|y_0, \ldots, y_n)$$

- Nonlinear HMM : $G_n(\theta) = p(y_0, \ldots, y_n|\theta) = \mathcal{Z}_n(\theta)$ normalizing Cts of

$$\mathbb{Q}_n^\theta := \mathrm{Law}((X_0, \ldots, X_n) \mid \theta, (y_0, \ldots, y_n))$$

$\rightsquigarrow$ 3 Solutions:

As above with $\mathcal{Z}_n^N(\theta)$ or Particle MCMC models (Andrieu, Doucet, Holenstein, JRSS B (2010)) or Gradient algo. w.r.t. additive functionals.

Additive functionals
(with Doucet & Singh `Hal-INRIA 09`*, to appear in M2AN (2010))*

- Hypothesis :

$$\begin{aligned}\mathbb{P}\left(X_n \in dx_n | X_{n-1} = x_{n-1}\right) &= H_n(x_{n-1}, x_n)\ \lambda_n(dx_n) \\ &\stackrel{ex.}{=} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}(x_n - F_n(x_{n-1}))^2}\ dx_n\end{aligned}$$

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

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 :

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

Some non asymptotic estimates

Test functions: Additive functionals $F_n(x_0, \ldots, x_n) := \frac{1}{n+1} \sum_{0 \le p \le n} f_p(x_p)$

- Bias estimate+uniform $\mathbb{L}_p$-bounds + variance

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

- Uniform exponential concentration

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

- Illustration $(G_n, \mathbb{Q}_n, \mathcal{Z}_n) \rightsquigarrow (G_n^\theta, \mathbb{Q}_n^\theta, \mathcal{Z}_n^\theta)$ with some varying θ

$$\frac{1}{n+1} \frac{\partial}{\partial \theta} \log \mathcal{Z}_{n+1}^\theta = \mathbb{Q}_{n+1}^\theta(F_n^\theta) \quad \text{with} \quad f_p^\theta = \frac{\partial \log G_p^\theta}{\partial \theta}$$