Stochastic particle methods in Bayesian statistical learning

P. Del Moral (INRIA team ALEA)

INRIA & Bordeaux Mathematical Institute & X CMAP

INRIA Workshop on Statistical Learning IHP Paris, Dec. 5-6, 2011

Some hyper-refs

- **Feynman-Kac formulae, Genealogical & Interacting Particle Systems with appl., Springer (2004)**
- **Sequential Monte Carlo Samplers JRSS B. (2006). (joint work with Doucet & Jasra)**
- **On the concentration of interacting processes Hal-inria (2011). (joint work with Hu & Wu) [+ Refs]**
- **More references on the website http://www.math.u-bordeaux1.fr/∼delmoral/index.html [+ Links]**

Stochastic particle sampling methods

- Interacting jumps models
- Genetic type interacting particle models
- Particle Feynman-Kac models
- The 4 particle estimates
- Island particle models ($\subset$ Parallel Computing)

Bayesian statistical learning

- Nonlinear filtering models
- Fixed parameter estimation in HMM models
- Particle stochastic gradient models
- Approximate Bayesian Computation
- Interacting Kalman-Filters
- Uncertainty propagations in numerical codes

Concentration inequalities

- Current population models
- Particle free energy
- Genealogical tree models
- Backward particle models

Introduction

Stochastic particle methods
=
Universal adaptive sampling technique

2 types of stochastic interacting particle models:

- Diffusive particle models with mean field drifts
  [McKean-Vlasov style] $\longleftrightarrow$ *Fluid mechanics*

Introduction

Stochastic particle methods
=
Universal adaptive sampling technique

2 types of stochastic interacting particle models:

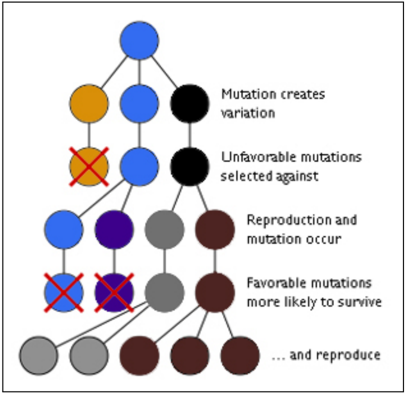
- Diffusive particle models with mean field drifts
  [McKean-Vlasov style] $\longleftrightarrow$ *Fluid mechanics*
- Interacting jump particle models
  [Boltzmann & Feynman-Kac style]
  $\longleftrightarrow$ *Fluid mech.* & *Physics + Biology + Engineering + Statistics*

Lectures $\subset$ Interacting jumps models

- **Interacting jumps = Recycling transitions =**
- **Discrete time models ($\Leftrightarrow$ geometric rejection/jump times)**

Genetic type interacting particle models

- Mutation-Proposals w.r.t. Markov transitions $X_{n-1} \rightsquigarrow X_n \in E_n$.
- Selection-Rejection-Recycling w.r.t. potential/fitness function G_n.



Equivalent particle algorithms

Sequential Monte Carlo	Sampling	Resampling
Particle Filters	Prediction	Updating
Genetic Algorithms	Mutation	Selection
Evolutionary Population	Exploration	Branching-selection
Diffusion Monte Carlo	Free evolutions	Absorption
Quantum Monte Carlo	Walkers motions	Reconfiguration
Sampling Algorithms	Transition proposals	Accept-reject-recycle

Equivalent particle algorithms

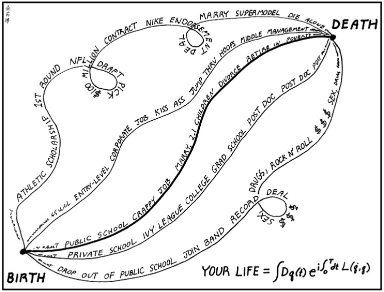
Sequential Monte Carlo	Sampling	Resampling
Particle Filters	Prediction	Updating
Genetic Algorithms	Mutation	Selection
Evolutionary Population	Exploration	Branching-selection
Diffusion Monte Carlo	Free evolutions	Absorption
Quantum Monte Carlo	Walkers motions	Reconfiguration
Sampling Algorithms	Transition proposals	Accept-reject-recycle

More botanical names:
bootstrapping, spawning, cloning, pruning, replenish, multi-level splitting, enrichment, go with the winner, ...

$1950 \leq$ Meta-Heuristic style stochastic algorithms ≤ 1996

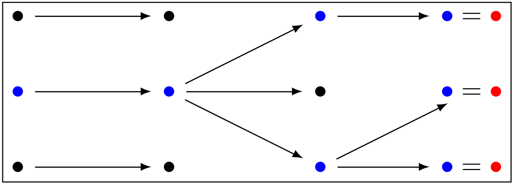
A single stochastic model

Particle interpretation of Feynman-Kac path integrals

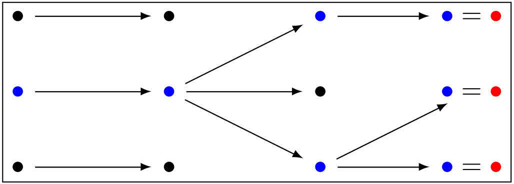


The Path Integral Formulation of Your Life

Genealogical tree evolution (size,time)=$(N, n) = (3, 3)$



Genealogical tree evolution (size,time)=$(N, n) = (3, 3)$

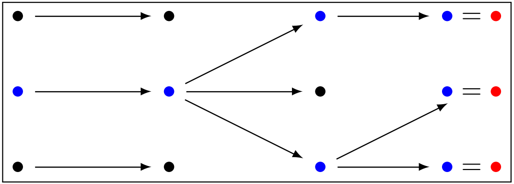


Meta-heuristics $\rightsquigarrow$ "96' Meta-Theorem" :

Ancestral lines $\simeq$ i.i.d. path samples w.r.t. Feynman-Kac measure

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

Genealogical tree evolution (size,time)=$(N, n) = (3, 3)$



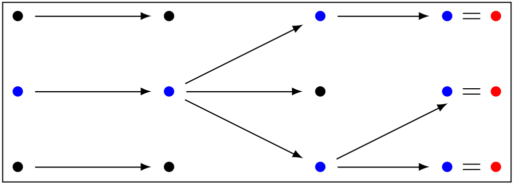
Meta-heuristics $\rightsquigarrow$ "96' Meta-Theorem" :

Ancestral lines $\simeq$ i.i.d. path samples w.r.t. Feynman-Kac measure

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

& Inversely!

Genealogical tree evolution (size,time)=$(N, n) = (3, 3)$



Meta-heuristics $\rightsquigarrow$ "96' Meta-Theorem" :

Ancestral lines $\simeq$ i.i.d. path samples w.r.t. Feynman-Kac measure

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

& Inversely! $\rightsquigarrow$ example

$$\mathbb{Q}_n = \mathrm{Law}((X_0, \ldots, X_n) \mid X_p \in A_p, \; p < n) \Longleftrightarrow G_n = 1_{A_n}$$

Particle estimates

More formally

$\left(\xi^i_{0,n},\xi^i_{1,n},\ldots,\xi^i_{n,n}\right) :=$ i-th ancetral line of the i-th current individual $= \xi^i_n$

$$\Downarrow$$

$$\frac{1}{N}\sum_{1\leq i\leq N}\delta_{\left(\xi^i_{0,n},\xi^i_{1,n},\ldots,\xi^i_{n,n}\right)} \longrightarrow_{N\rightarrow\infty} \mathbb{Q}_n$$

Particle estimates

More formally

$$\left(\xi^i_{0,n}, \xi^i_{1,n}, \ldots, \xi^i_{n,n}\right) := i\text{-th ancetral line of the } i\text{-th current individual} = \xi^i_n$$

$$\Downarrow$$

$$\frac{1}{N} \sum_{1 \le i \le N} \delta_{\left(\xi^i_{0,n}, \xi^i_{1,n}, \ldots, \xi^i_{n,n}\right)} \longrightarrow_{N \to \infty} \mathbb{Q}_n$$

$\oplus$ *Current population models*

$$\eta^N_n := \frac{1}{N} \sum_{1 \le i \le N} \delta_{\xi^i_n} \longrightarrow_{N \to \infty} \eta_n = n\text{-th time marginal of } \mathbb{Q}_n$$

Particle estimates

More formally

$$\left(\xi^i_{0,n},\xi^i_{1,n},\ldots,\xi^i_{n,n}\right) := i\text{-th ancetral line of the } i\text{-th current individual} = \xi^i_n$$

$$\Downarrow$$

$$\frac{1}{N}\sum_{1\leq i\leq N}\delta_{\left(\xi^i_{0,n},\xi^i_{1,n},\ldots,\xi^i_{n,n}\right)} \longrightarrow_{N\to\infty} \mathbb{Q}_n$$

⊕ *Current population models*

$$\eta^N_n := \frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi^i_n} \longrightarrow_{N\to\infty} \eta_n = n\text{-th time marginal of } \mathbb{Q}_n$$

⊕ *Unbiased particle approximation*

$$\mathcal{Z}^N_n = \prod_{0\leq p<n}\eta^N_p(G_p) \longrightarrow_{N\to\infty} \mathcal{Z}_n = \mathbb{E}\left(\prod_{0\leq p<n}G_p(X_p)\right) = \prod_{0\leq p<n}\eta_p(G_p)$$

Particle estimates

More formally

$$\left(\xi^i_{0,n}, \xi^i_{1,n}, \ldots, \xi^i_{n,n}\right) := i\text{-th ancetral line of the } i\text{-th current individual} = \xi^i_n$$

$$\Downarrow$$

$$\frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\left(\xi^i_{0,n}, \xi^i_{1,n}, \ldots, \xi^i_{n,n}\right)} \longrightarrow_{N \to \infty} \mathbb{Q}_n$$

⊕ *Current population models*

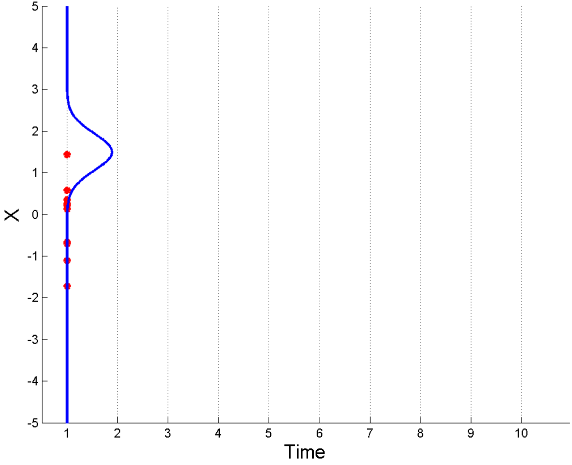
$$\eta^N_n := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi^i_n} \longrightarrow_{N \to \infty} \eta_n = n\text{-th time marginal of } \mathbb{Q}_n$$

⊕ *Unbiased particle approximation*

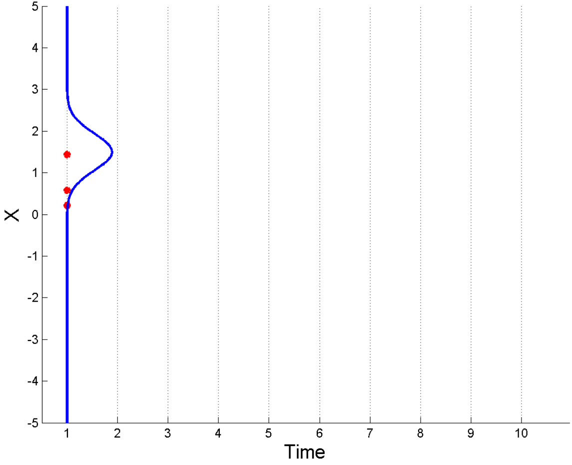
$$\mathcal{Z}^N_n = \prod_{0 \leq p < n} \eta^N_p(G_p) \longrightarrow_{N \to \infty} \mathcal{Z}_n = \mathbb{E}\left(\prod_{0 \leq p < n} G_p(X_p)\right) = \prod_{0 \leq p < n} \eta_p(G_p)$$

Ex.: $G_n = 1_{A_n} \rightsquigarrow \mathcal{Z}^N_n = \prod$ *proportion of success* $\longrightarrow \mathbb{P}(X_p \in A_p, \; p < n)$

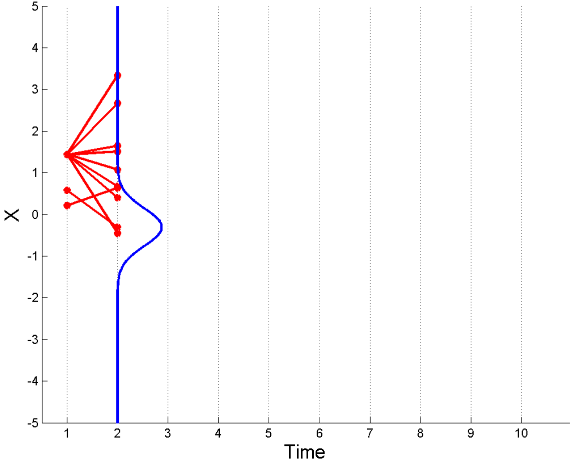
Graphical illustration : $\eta_n^N := \frac{1}{N}\sum_{1\le i\le N}\delta_{\xi_n^i} \simeq \eta_n$



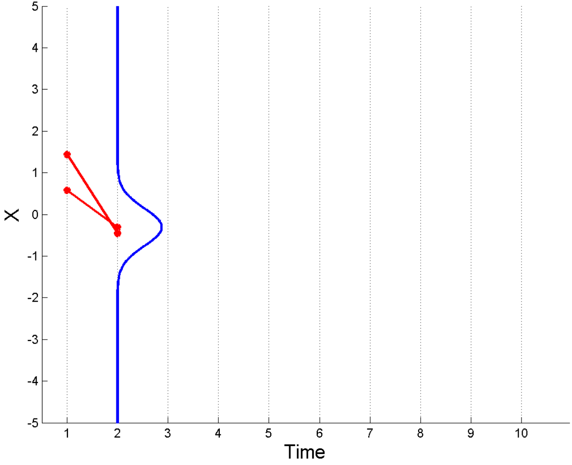
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



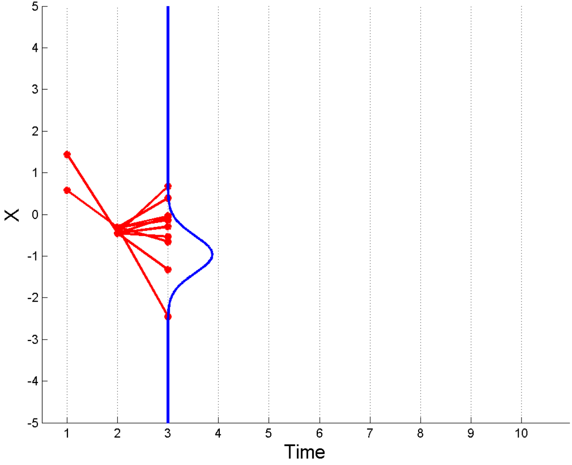
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



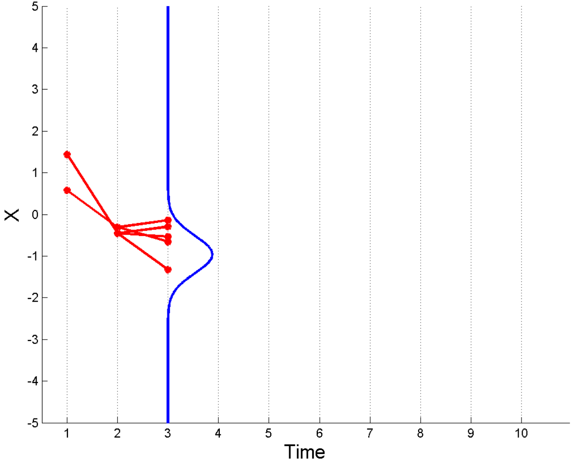
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



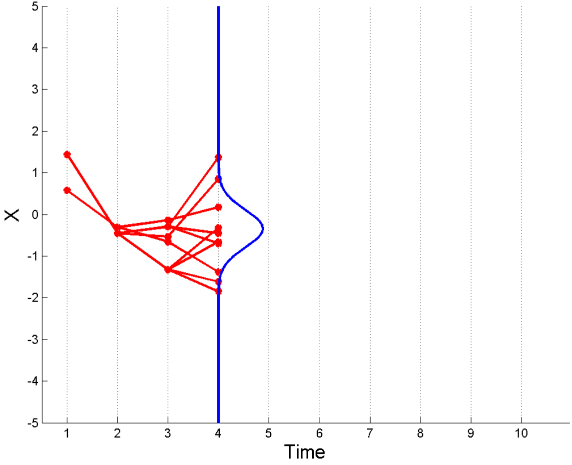
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



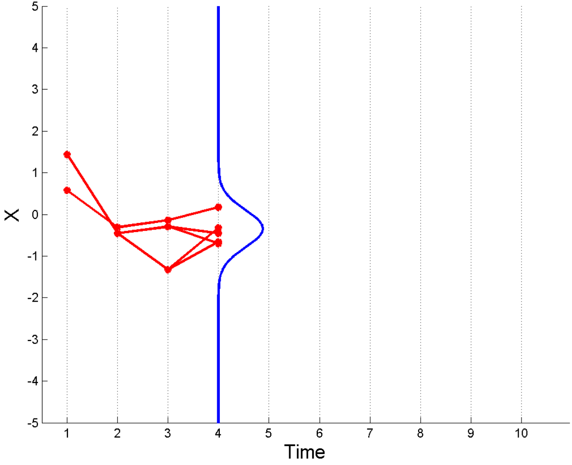
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



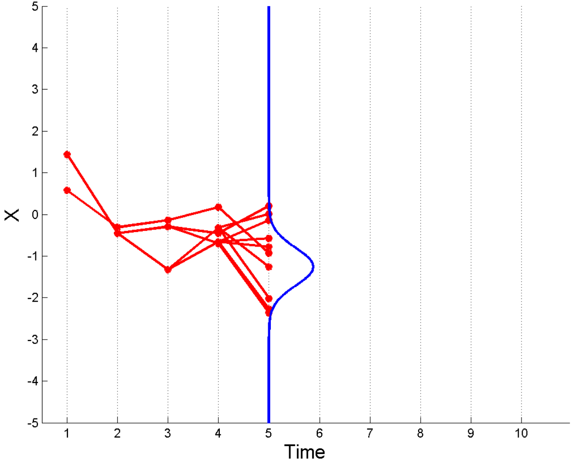
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



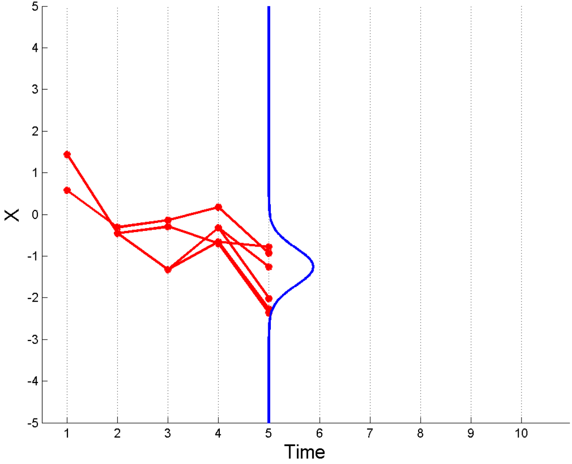
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



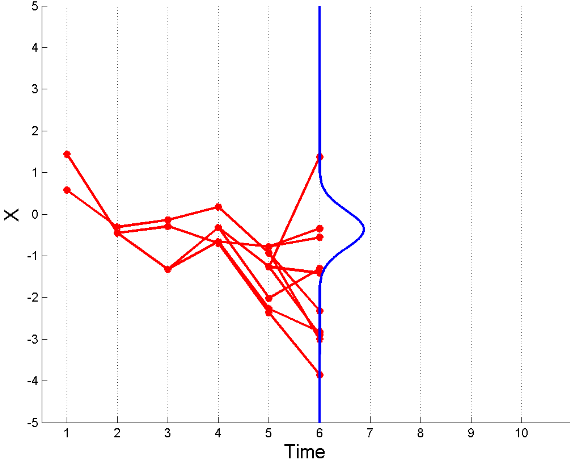
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



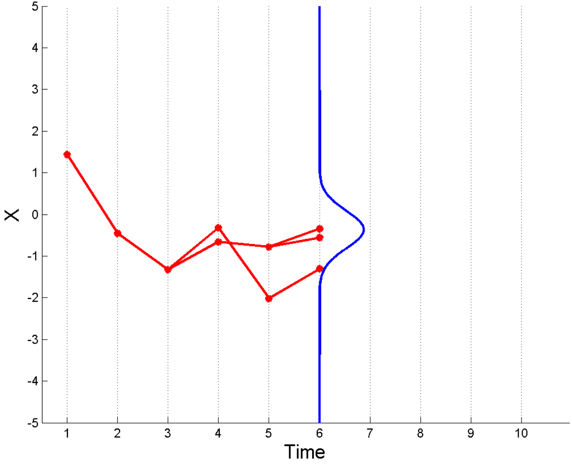
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



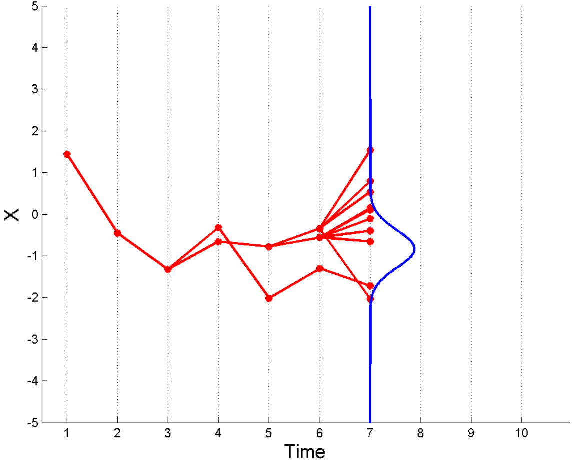
Graphical illustration : $\eta_n^N := \frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi_n^i} \simeq \eta_n$



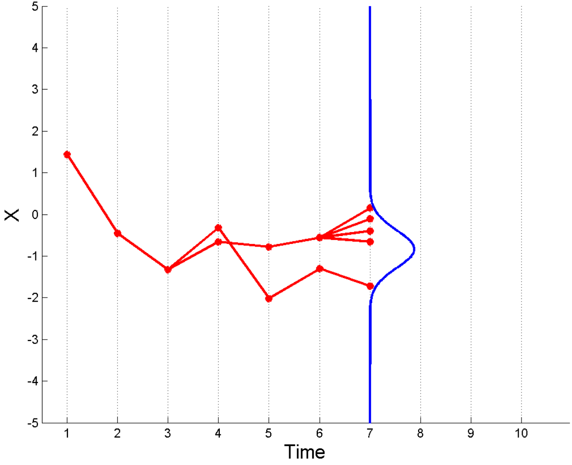
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



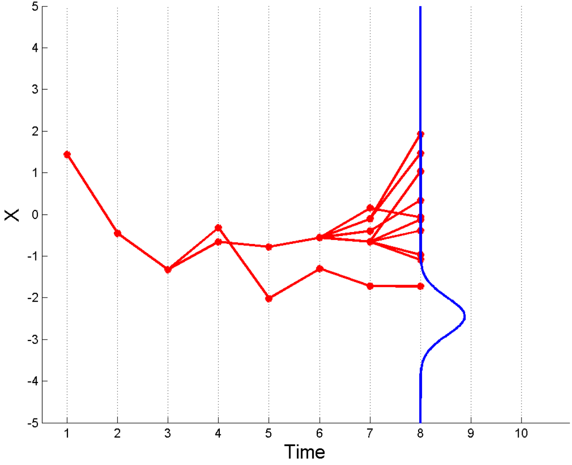
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



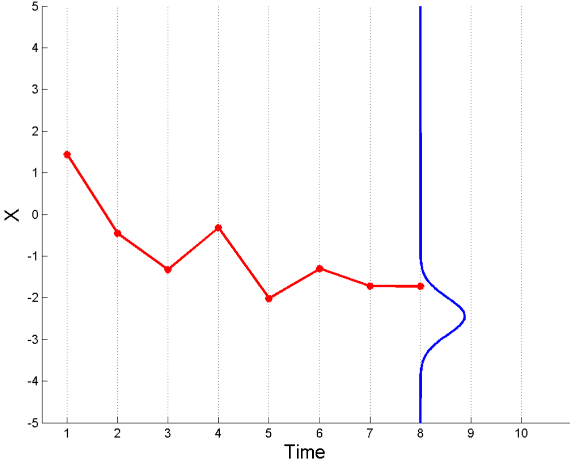
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



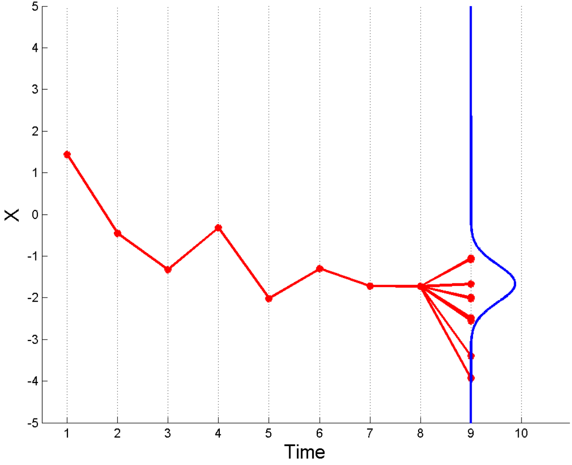
Graphical illustration : $\eta_n^N := \frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi_n^i} \simeq \eta_n$



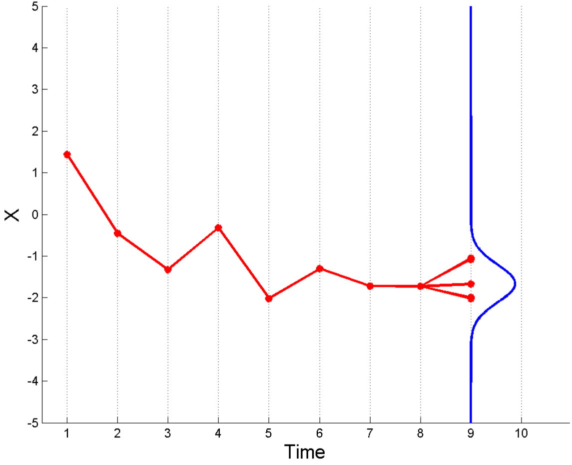
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



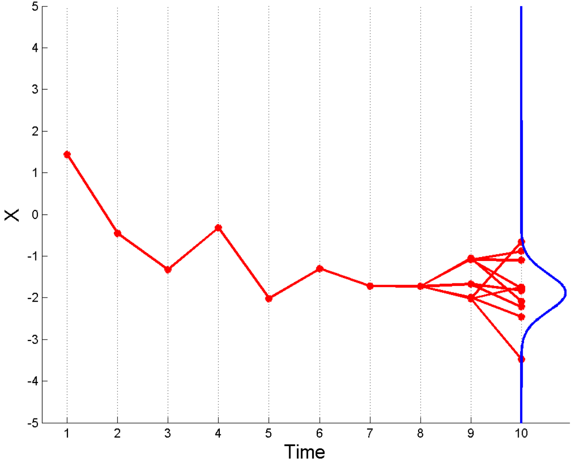
Graphical illustration : $\eta_n^N := \frac{1}{N}\sum_{1\leq i\leq N} \delta_{\xi_n^i} \simeq \eta_n$



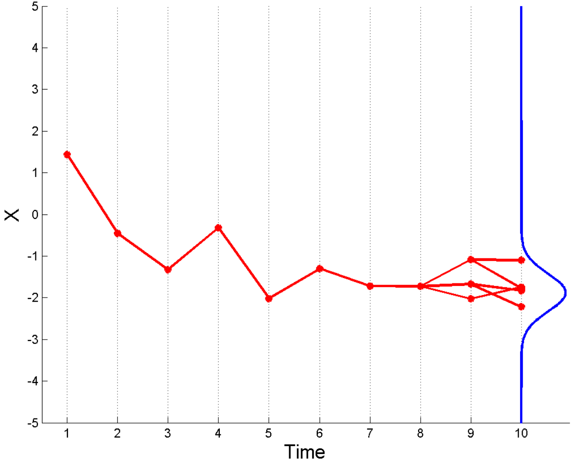
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \le i \le N} \delta_{\xi_n^i} \simeq \eta_n$



Complete ancestral tree when $G_{n-1}(x)M_n(x,dy) = H_n(x,y)\ \lambda(dy)$

Backward Markov chain model

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

with the random particle matrices:

$$\mathbb{M}_{n+1,\eta_n^N}(x_{n+1},dx_n) \propto \eta_n^N(dx_n)\ H_{n+1}(x_n,x_{n+1})$$

Complete ancestral tree when $G_{n-1}(x)M_n(x,dy) = H_n(x,y)\ \lambda(dy)$

Backward Markov chain model

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

with the random particle matrices:

$$\mathbb{M}_{n+1,\eta_n^N}(x_{n+1},dx_n) \propto \eta_n^N(dx_n)\ H_{n+1}(x_n,x_{n+1})$$

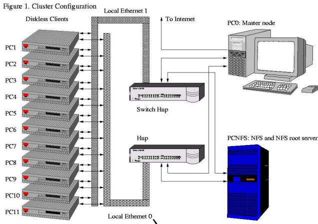
Example: Normalized additive functionals

$$\mathbf{f_n}(x_0,\ldots,x_n) = \frac{1}{n+1}\sum_{0\leq p\leq n} f_p(x_p)$$

$$\Downarrow$$

$$\mathbb{Q}_n^N(\mathbf{f_n}) := \frac{1}{n+1}\sum_{0\leq p\leq n} \eta_n^N \underbrace{\mathbb{M}_{n,\eta_{n-1}^N}\ldots\mathbb{M}_{p+1,\eta_p^N}(f_p)}_{\text{matrix operations}}$$

Island models ($\subset$ Parallel Computing)



Reminder : the unbiased property

$$\mathbb{E}\left(\mathbf{f_n}(\mathbf{X_n}) \prod_{0\leq p<n} \mathbf{G_p}(\mathbf{X_p})\right) = \mathbb{E}\left(\eta_n^N(\mathbf{f_n}) \prod_{0\leq p<n} \eta_p^N(\mathbf{G_p})\right)$$

$$= \mathbb{E}\left(\mathbf{F_n}(\mathcal{X}_n) \prod_{0\leq p<n} \mathcal{G}_p(\mathcal{X}_p)\right)$$

with the Island evolution Markov chain model

$$\mathcal{X}_n := \eta_n^N \qquad \text{and} \qquad \mathcal{G}_n(\mathcal{X}_n) = \eta_n^N(\mathbf{G_n}) = \mathcal{X}_n(\mathbf{G_n})$$

$$\Downarrow$$

particle model with $(\mathcal{X}_n, \mathcal{G}_n(\mathcal{X}_n))$ = Interacting Island particle model

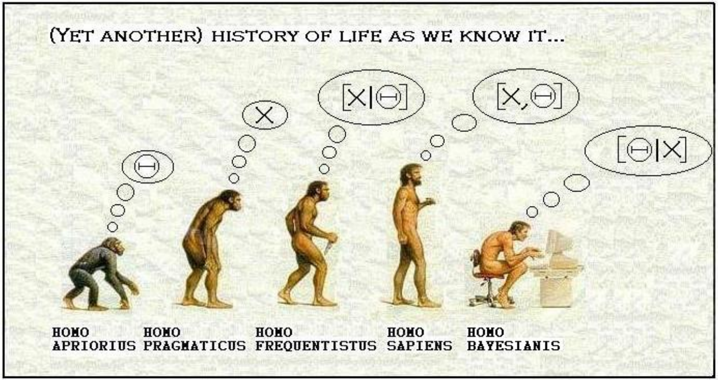
Stochastic particle sampling methods

Bayesian statistical learning

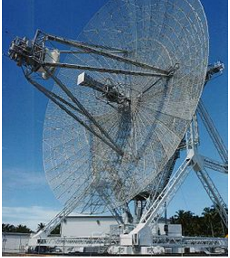
- Nonlinear filtering models
- Fixed parameter estimation in HMM models
- Particle stochastic gradient models
- Approximate Bayesian Computation
- Interacting Kalman-Filters
- Uncertainty propagations in numerical codes

Concentration inequalities

Bayesian statistical learning



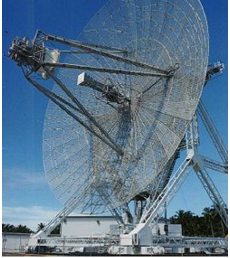
Signal processing & filtering models



$$\text{Law}\left(\textbf{Markov process } X \mid \textbf{Noisy \& Partial observations } Y\right)$$

- **Signal** X **:** target evolution (missile, plane, robot, vehicle, image contours), forecasting models, assets volatility, speech signals, ...
- **Observation** Y **:** Radar/Sonar/Gps sensors, financial assets prices, image processing, audio receivers, statistical data measurements, ...

Signal processing & filtering models



$$\text{Law}\left(\textbf{Markov process } X \;\middle|\; \textbf{Noisy \& Partial observations } Y\right)$$

- **Signal** X **:** target evolution (missile, plane, robot, vehicle, image contours), forecasting models, assets volatility, speech signals, ...
- **Observation** Y **:** Radar/Sonar/Gps sensors, financial assets prices, image processing, audio receivers, statistical data measurements, ...

$\subset$ **Multiple objects tracking models (highly more complex pb)**

- On the Stability and the Approximation of Branching Distribution Flows, with Applications to Nonlinear Multiple Target Filtering. Francois Caron, Pierre Del Moral, Michele Pace, and B.-N. Vo (HAL-INRIA RR-7376) [50p]. Stoch. Analysis and Applications Volume 29, Issue 6, 2011.
- Comparison of implementations of Gaussian mixture PHD filters. M. Pace, P. Del Moral, Fr. Caron 13th International Conference on Information. FUSION, EICC, Edinburgh, UK, 26-29 July (2010)

$$\text{Law}\left(X = \sum_{1\leq i\leq N_t^X} \delta_{X_t^i} \;\middle|\; Y = \sum_{1\leq i\leq N_t^Y} \delta_{Y_t^i}\right)$$

Filtering (prediction $\oplus$ smoothing)

$$p((x_0,\ldots,x_n) \mid (y_0,\ldots,y_n)) \quad \& \quad p(y_0,\ldots,y_n) \ ?$$

Bayes' rule

$$p((x_0,\ldots,x_n) \mid (y_0,\ldots,y_n)) \propto \underbrace{p((y_0,\ldots,y_n) \mid (x_0,\ldots,x_n))}_{\prod_{0\leq k\leq n} p(y_k|x_k) \leftarrow \text{likelihood functions } G_k} \times p(x_0,\ldots,x_n)$$

Filtering (prediction $\oplus$ smoothing)

$$p((x_0,\ldots,x_n) \mid (y_0,\ldots,y_n)) \quad \& \quad p(y_0,\ldots,y_n) \ ?$$

Bayes' rule

$$p((x_0,\ldots,x_n) \mid (y_0,\ldots,y_n)) \propto \underbrace{p((y_0,\ldots,y_n) \mid (x_0,\ldots,x_n))}_{\prod_{0\leq k\leq n} p(y_k|x_k) \leftarrow \text{likelihood functions } G_k} \times p(x_0,\ldots,x_n)$$

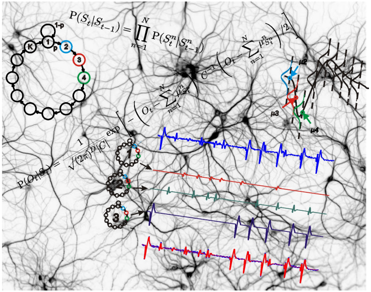
$$\Downarrow$$

Feynman-Kac models : $G_n(x_n) := p(y_n|x_n)$ & $\mathbb{P}_n := \mathrm{Law}\,(X_0,\ldots,X_n)$

$$\mathrm{Law}\,((X_0,\ldots,X_n) \mid Y_p = y_p,\ p < n) = \frac{1}{\mathcal{Z}_n}\left\{\prod_{0\leq p<n} G_p(X_p)\right\}\ \mathbb{P}_n$$

Note : **Not unique stochastic model!**

Hidden Markov chains problems



$$\Theta \rightsquigarrow \text{Signal} \quad X^{\Theta} \rightsquigarrow \text{observations} \quad Y^{\Theta}$$

$$\textbf{Law}\left(\textbf{fixed parameter } \Theta \;\middle|\; \textbf{Noisy \& Partial observations } Y^{\Theta}\right)$$

- **Parameter** Θ **:** kinetic model unknown parameters, statistical parameters (signal/sensors), hypothesis testing, ..
- **Signal** X^{Θ} **:** Single or multiple targets evolution, forecasting models, financial assets volatility, speech signals, video images, ...
- **Observation** Y^{Θ} **:** Radar/Sonar/Gps sensors, financial assets prices, image processing, statistical data measurements, ...

Posterior density

$$p(\ \theta \mid (y_0, \ldots, y_n)) \propto \underbrace{p((y_0, \ldots, y_n) \mid \theta)}_{\prod_{0 \leq k \leq n} p(y_k | \theta, (y_0, \ldots, y_{k-1})) \leftarrow \text{likelihood functions}} \times \ p(\theta)$$

Posterior density

$$p(\ \theta \mid (y_0,\ldots,y_n)) \propto \underbrace{p((y_0,\ldots,y_n) \mid \theta)}_{\prod_{0\leq k\leq n} p(y_k|\theta,(y_0,\ldots,y_{k-1}))\leftarrow \text{likelihood functions}} \times\ p(\theta)$$

$$\Downarrow$$

Multiplicative formulation

$$\mathrm{Law}\,(\Theta \mid (y_0,\ldots,y_n)) \propto \left\{\prod_{0\leq p\leq n} h_p(\theta)\right\}\ \lambda(d\theta)$$

with

$$h_n(\theta) := p(y_n|\theta,(y_0,\ldots,y_{n-1})) \quad \& \quad \lambda := \mathrm{Law}\,(\Theta)$$

First key observation

$$p((y_0,\ldots,y_n)|\theta) = \prod_{0\leq p\leq n} h_p(\theta) = \mathcal{Z}_n(\theta)$$

with the normalizing constant $\mathcal{Z}_n(\theta)$ of the conditional distribution

$$p((x_0,\ldots,x_n)|(y_0,\ldots,y_n),\theta)$$

$$= \frac{1}{p((y_0,\ldots,y_n)|\theta)}\; p((y_0,\ldots,y_n)|(x_0,\ldots,x_n),\theta)\; p((x_0,\ldots,x_n)|\theta)$$

First key observation

$$p((y_0,\ldots,y_n)|\theta) = \prod_{0\leq p\leq n} h_p(\theta) = \mathcal{Z}_n(\theta)$$

with the normalizing constant $\mathcal{Z}_n(\theta)$ of the conditional distribution

$$p((x_0,\ldots,x_n)|(y_0,\ldots,y_n),\theta)$$

$$= \frac{1}{p((y_0,\ldots,y_n)|\theta)}\ p((y_0,\ldots,y_n)|(x_0,\ldots,x_n),\theta)\ p((x_0,\ldots,x_n)|\theta)$$

Second key observation

$h_n(\theta)$ and $\mathcal{Z}_n(\theta)$ easy to compute for linear/gaussian models

Third key observation : Any target measure of the form

$$\eta_n(d\theta) = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0\leq p\leq n} h_p(\theta) \right\} \times \lambda(d\theta)$$

is the n-th time marginal of the Feynman-Kac measure

$$\mathbb{Q}_n := \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0\leq p<n} G_p(\Theta_p) \right\} \mathbb{P}_n$$

with

$$G_n = h_{n+1} \quad \text{and} \quad \mathbb{P}_n := \text{Law}\left(\Theta_0, \ldots, \Theta_n\right)$$

where

$$\Theta_{p-1} \rightsquigarrow \Theta_p \text{ as an MCMC move with target measure } \eta_p$$

Particle auxiliary variables $\theta \rightsquigarrow \xi^\theta \sim P(\theta, d\xi)$

$$\overline{\eta}_n(d\overline{\theta}) \propto \left\{ \prod_{0\leq p\leq n} \overline{h}_p(\overline{\theta}) \right\} \quad \underbrace{\overline{\lambda}(d\overline{\theta})}_{=\lambda(d\theta)\times P(\theta,d\xi)}$$

with $\overline{\theta} = (\theta, \xi)$ and

$$\overline{h}_n(\overline{\theta}) := \frac{1}{N}\sum_{i=1}^{N} p(y_n \mid \xi_n^{\theta,i}) \simeq_{N\uparrow\infty} p(y_n|\theta,(y_0,\ldots,y_{n-1})) = h_p(\theta)$$

Particle auxiliary variables $\theta \rightsquigarrow \xi^\theta \sim P(\theta, d\xi)$

$$\overline{\eta}_n(d\overline{\theta}) \propto \left\{ \prod_{0 \leq p \leq n} \overline{h}_p(\overline{\theta}) \right\} \underbrace{\overline{\lambda}(d\overline{\theta})}_{=\lambda(d\theta)\times P(\theta, d\xi)}$$

with $\overline{\theta} = (\theta, \xi)$ and

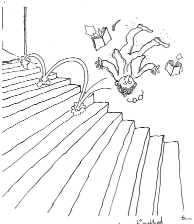
$$\overline{h}_n(\overline{\theta}) := \frac{1}{N} \sum_{i=1}^{N} p(y_n \mid \xi_n^{\theta,i}) \simeq_{N\uparrow\infty} p(y_n|\theta, (y_0, \ldots, y_{n-1})) = h_p(\theta)$$

But by the unbiased property the θ-marginal of $\overline{\eta}_n$ coincides with

$$\mathrm{Law}\left(\Theta \mid (y_0, \ldots, y_n)\right) \propto \left\{ \prod_{0 \leq p \leq n} h_p(\theta) \right\} \lambda(d\theta)$$

Feynman-Kac formulation :
Markov chain $\overline{\Theta}_k = \left(\Theta_k, \xi^{(k)}\right)$ MCMC with target $\overline{\eta}_n$ and $G_n = \overline{h}_{n+1}$

Particle gradient models ($\theta \in \mathbb{R}^d$)



Just after learning the "Steepest Descent" method in optimization class...

$$\mathcal{Z}_n(\theta) = p((y_0, \ldots, y_{n-1}) \mid \theta) = \mathbb{E}\left(\prod_{0 \le q < n} p(y_q \mid \theta,\ X_q^\theta) \right)$$

$$\Rightarrow \underbrace{\nabla \log \mathcal{Z}_n(\theta)}_{\text{derivative}} \quad = \quad \underbrace{\mathbb{Q}_n^{(\theta)}(\Lambda_n)}_{\text{path-integral}}$$

with the Feynman-Kac measure $\mathbb{Q}_n^{(\theta)}$ on path space associated with

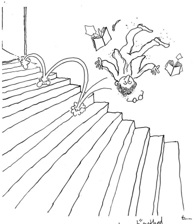
$$(X_n^\theta, G_n^\theta(x_n)) = (X_n^\theta, p(y_q \mid \theta, x_n))$$

and with the additive functional

$$\Lambda_n(x_0, \ldots, x_n) = \sum_{0 \le p < n} \nabla \log \left(p(x_{q+1} | \theta, x_q) p(y_q \mid \theta, x_q) \right)$$

Particle gradient models ($\theta \in \mathbb{R}^d$)

$$\mathcal{Z}_n(\theta) = p((y_0, \ldots, y_{n-1}) \mid \theta) = \mathbb{E}\left(\prod_{0 \leq q < n} p(y_q \mid \theta,\ X_q^\theta)\right)$$



Just after learning the "Steepest Descent" method in optimization class...

$$\Rightarrow \underbrace{\nabla \log \mathcal{Z}_n(\theta)}_{\text{derivative}} = \underbrace{\mathbb{Q}_n^{(\theta)}(\Lambda_n)}_{\text{path-integral}}$$

with the Feynman-Kac measure $\mathbb{Q}_n^{(\theta)}$ on path space associated with

$$(X_n^\theta, G_n^\theta(x_n)) = (X_n^\theta, p(y_q \mid \theta, x_n))$$

and with the additive functional

$$\Lambda_n(x_0, \ldots, x_n) = \sum_{0 \leq p < n} \nabla \log \left(p(x_{q+1} | \theta, x_q) p(y_q \mid \theta, x_q)\right)$$

$\rightsquigarrow$ **Particle gradient algorithm**

$$\Theta_n = \Theta_{n-1} + \tau_n\ \nabla \mathbb{Q}_n^{(\theta)}(\Lambda_n) \simeq \Theta_{n-1} + \tau_n\ \nabla \mathbb{Q}_n^{(\theta), N}(\Lambda_n)$$

Approximate Bayesian Computation

When $p(y_n|x_n)$ is untractable or impossible to compute in reasonable time

$$\left\{ \begin{array}{lcl} X_n & = & F_n(X_{n-1}, W_n) \\ Y_n & = & H_n(X_n, V_n) \end{array} \right. \xrightarrow{\mathcal{X}_n = (X_n, Y_n)} \left\{ \begin{array}{lcl} \mathcal{X}_n & = & \mathcal{F}_n(\mathcal{X}_{n-1}, \mathcal{W}_n) \\ Y_n^\epsilon & = & Y_n + \epsilon \; V_n^\epsilon \end{array} \right.$$

$$\Downarrow$$

$$\textbf{Law}\,(X \mid Y^\epsilon = y) \simeq_{\epsilon \downarrow 0} \textbf{Law}\,(X \mid Y = y)$$

Approximate Bayesian Computation

When $p(y_n|x_n)$ is untractable or impossible to compute in reasonable time

$$\left\{\begin{array}{rcl} X_n & = & F_n(X_{n-1}, W_n) \\ Y_n & = & H_n(X_n, V_n) \end{array}\right. \xrightarrow{\mathcal{X}_n=(X_n,Y_n)} \left\{\begin{array}{rcl} \mathcal{X}_n & = & \mathcal{F}_n(\mathcal{X}_{n-1}, \mathcal{W}_n) \\ Y_n^\epsilon & = & Y_n + \epsilon\, V_n^\epsilon \end{array}\right.$$

$$\Downarrow$$

$$\textbf{Law}\,(X \mid Y^\epsilon = y\,) \simeq_{\epsilon\downarrow 0} \textbf{Law}\,(X \mid Y = y\,)$$

$$\Downarrow$$

Feynman-Kac model with the Markov chain and the potentials :

$$\mathcal{X}_n = (X_n, Y_n) \quad \text{and} \quad G_n(\mathcal{X}_n) = p(Y_n^\epsilon | Y_n)$$

Interacting Kalman-Filters

$X_n = (X_n^1, X_n^2)$ with X_n^1 Markov and $(X_n^2, Y_n)|X^1$ linear-gaussian model

$$\left\{ \begin{array}{rcl} X_n^2 & = & A_n(X_n^1)\ X_{n-1}^2 + B_n(X_n^1)\ W_n \\ Y_n & = & C_n(X_n^1)\ X_n^2 + D_n(X_n^1)\ V_n \end{array} \right.$$

$$\Downarrow$$

$$\textbf{Law}\left(X_n^2 \ \middle|\ X^1,\ Y_p = y_p,\ p < n\right) = \eta_{X^1,n} = \text{Kalman gaussian predictor}$$

Interacting Kalman-Filters

$X_n = (X_n^1, X_n^2)$ with X_n^1 Markov and $(X_n^2, Y_n)|X^1$ linear-gaussian model

$$\begin{cases} X_n^2 &= A_n(X_n^1)\ X_{n-1}^2 + B_n(X_n^1)\ W_n \\ Y_n &= C_n(X_n^1)\ X_n^2 + D_n(X_n^1)\ V_n \end{cases}$$

$$\Downarrow$$

$$\textbf{Law}\left(X_n^2 \mid X^1,\ Y_p = y_p,\ p < n\right) = \eta_{X^1,n} = \text{Kalman gaussian predictor}$$

$$\Downarrow$$

Law$((X^1, X^2) \mid Y)$= Feynman-Kac model with

$$\mathcal{X}_n = (X_n^1, \eta_{X^1,n}) \quad \& \quad G_n(\mathcal{X}_n) = \int p(Y_n \mid (x_n^1, x_n^2))\ \eta_{X^1,n}(dx_n^2)$$

- **Uncertainty propagations in numerical codes**

$$\textbf{Law}\left(\textbf{Inputs } \mathcal{I} \;\middle|\; \textbf{Outputs } \mathcal{O} = C(\mathcal{I}) \in \textbf{Reference or Critical event}\right)$$

$$\Updownarrow$$

$$\left.\begin{array}{l} \mu = \mathrm{Law}(\mathcal{I}) \\ A = \{\mathcal{I} \;:\; C(\mathcal{I}) \in B\} \end{array}\right\} \longrightarrow \mathbb{P}\left(\mathcal{I} \in A\right) = \mu(A) \;\&\; \mathrm{Law}\left(\mathcal{I} \mid \mathcal{I} \in A\right) = \mu_A$$

- **Uncertainty propagations in numerical codes**

Law $\left(\textbf{Inputs } \mathcal{I} \;\middle|\; \textbf{Outputs } \mathcal{O} = C(\mathcal{I}) \;\in \textbf{ Reference or Critical event}\right)$

$$\Updownarrow$$

$$\left.\begin{array}{l} \mu = \mathrm{Law}(\mathcal{I}) \\ A = \{\mathcal{I} \;:\; C(\mathcal{I}) \in B\} \end{array}\right\} \longrightarrow \mathbb{P}\left(\mathcal{I} \in A\right) = \mu(A) \;\&\; \mathrm{Law}\left(\mathcal{I} \mid \mathcal{I} \in A\right) = \mu_A$$

Multi-level decomposition

$$h_n = 1_{A_n} \text{ with } A_n \downarrow \;\Longrightarrow \mu_{A_n}(dx) \propto \left\{\prod_{0 \leq p \leq n} h_p(x)\right\} \mu(dx)$$

- **Uncertainty propagations in numerical codes**

Law $\left(\textbf{Inputs } \mathcal{I} \;\middle|\; \textbf{Outputs } \mathcal{O} = C(\mathcal{I}) \;\in \textbf{Reference or Critical event}\;\right)$

$$\Updownarrow$$

$$\left.\begin{array}{l}\mu = \mathrm{Law}(\mathcal{I}) \\ A = \{\mathcal{I} \;:\; C(\mathcal{I}) \in B\}\end{array}\right\} \longrightarrow \mathbb{P}\left(\mathcal{I} \in A\right) = \mu(A) \;\&\; \mathrm{Law}\left(\mathcal{I} \mid \mathcal{I} \in A\right) = \mu_A$$

Multi-level decomposition

$$h_n = 1_{A_n} \text{ with } A_n \downarrow \implies \mu_{A_n}(dx) \propto \left\{\prod_{0 \le p \le n} h_p(x)\right\} \mu(dx)$$

$\Rightarrow$ **Feynman-Kac representation**

$$(X_{n-1} \rightsquigarrow X_n) = \text{MCMC moves with target } \mu_{A_n} \quad \& \quad G_n = 1_{A_{n+1}}$$

Stochastic particle sampling methods

Bayesian statistical learning

Concentration inequalities

- Current population models
- Particle free energy
- Genealogical tree models
- Backward particle models

Current population models

Constants (c_1, c_2) related to (bias,variance), c universal constant
Test funct. $\|f_n\| \leq 1$

- $\forall\ (x \geq 0, n \geq 0, N \geq 1)$, the probability of the event

$$\left[\eta_n^N - \eta_n\right](f) \leq \frac{c_1}{N}\ \left(1 + x + \sqrt{x}\right) + \frac{c_2}{\sqrt{N}}\ \sqrt{x}$$

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

- $y = (y_i)_{1 \leq i \leq d} \rightsquigarrow (-\infty, y] = \prod_{i=1}^{d}(-\infty, y_i]$ cells in $E_n = \mathbb{R}^d$.

$$F_n(y) = \eta_n\left(1_{(-\infty,y]}\right) \quad \text{and} \quad F_n^N(y) = \eta_n^N\left(1_{(-\infty,y]}\right)$$

$\forall\ (x \geq 0, n \geq 0, N \geq 1)$, the probability of the following event

$$\sqrt{N}\ \left\|F_n^N - F_n\right\| \leq c\ \sqrt{d\ (x+1)}$$

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

Particle free energy models

Constants (c_1, c_2) related to (bias,variance), c universal constant.

- $\forall\ (x \geq 0, n \geq 0, N \geq 1, \epsilon \in \{+1, -1\})$, the probability of the event

$$\frac{\epsilon}{n} \log \frac{\mathcal{Z}_n^N}{\mathcal{Z}_n} \leq \frac{c_1}{N}\ \left(1 + x + \sqrt{x}\right) + \frac{c_2}{\sqrt{N}}\ \sqrt{x}$$

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

Note : $(0 \leq \epsilon \leq 1 \Rightarrow (1 - e^{-\epsilon}) \vee (e^{\epsilon} - 1) \leq 2\epsilon)$

$$e^{-\epsilon} \leq \frac{z^N}{z} \leq e^{\epsilon} \Rightarrow \left|\frac{z^N}{z} - 1\right| \leq 2\epsilon$$

Genealogical tree models $:= \eta_n^N$ (in path space)

Constants (c_1, c_2) related to (bias,variance), c universal constant
$\mathbf{f_n}$ test function $\|\mathbf{f_n}\| \leq 1$.

- $\forall \ (x \geq 0, n \geq 0, N \geq 1)$, the probability of the event

$$\left[\eta_n^N - \mathbb{Q}_n\right](f) \leq c_1 \ \frac{n+1}{N} \ \left(1 + x + \sqrt{x}\right) + c_2 \ \sqrt{\frac{(n+1)}{N}} \ \sqrt{x}$$

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

- $\mathcal{F}_n =$ indicator fct. $\mathbf{f_n}$ of cells in $\mathbf{E_n} = \left(\mathbb{R}^{d_0} \times \ldots, \times \mathbb{R}^{d_n}\right)$
  $\forall \ (x \geq 0, n \geq 0, N \geq 1)$, the probability of the following event

$$\sup_{\mathbf{f_n} \in \mathcal{F}_n} \left|\eta_n^N(\mathbf{f_n}) - \mathbb{Q}_n(\mathbf{f_n})\right| \leq c \ (n+1) \ \sqrt{\frac{\sum_{0 \leq p \leq n} d_p}{N} \ (x+1)}$$

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

Backward particle models

Constants (c_1, c_2) related to (bias,variance), c universal constant.
$\mathbf{f_n}$ normalized additive functional with $\|f_p\| \leq 1$.

- $\forall\ (x \geq 0, n \geq 0, N \geq 1)$, the probability of the event

$$\left[\mathbb{Q}_n^N - \mathbb{Q}_n\right](\overline{\mathbf{f}}_n) \leq c_1 \ \frac{1}{N}\ (1 + (x + \sqrt{x})) + c_2\ \sqrt{\frac{x}{N(n+1)}}$$

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

- $\mathbf{f}_{a,n}$ normalized additive functional w.r.t. $f_p = 1_{(-\infty,a]}$, $a \in \mathbb{R}^d = E_n$.

  $\forall\ (x \geq 0, n \geq 0, N \geq 1)$, the probability of the following event

$$\sup_{a \in \mathbb{R}^d} \left|\mathbb{Q}_n^N(\mathbf{f_{a,n}}) - \mathbb{Q}_n(\mathbf{f_{a,n}})\right| \leq c\ \sqrt{\frac{d}{N}(x+1)}$$

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