# Advanced Monte Carlo integration methods

P. Del Moral (INRIA team ALEA)

INRIA & Bordeaux Mathematical Institute & X CMAP

MCQMC 2012, Sydney, Sunday Tutorial 12-th 2012

**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. Foundations & Trends in Machine Learning (2012). (joint work with Hu & Wu) [+ Refs]
- More references on the website http://www.math.u-bordeaux1.fr/∼delmoral/index.html [+ Links]

# Basic notation

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

- $(\mu, f) \in \mathcal{P}(E) \times \mathcal{B}(E) \quad \longrightarrow \quad \mu(f) = \int \mu(dx)\ f(x)$
- $Q(x_1, dx_2)$ **integral operators** $x_1 \in E_1 \rightsquigarrow x_2 \in E_2$

$$\begin{aligned} Q(f)(x_1) &= \int Q(x_1, dx_2) f(x_2) \\ [\mu Q](dx_2) &= \int \mu(dx_1) Q(x_1, dx_2) \qquad (\Longrightarrow [\mu Q](f) = \mu[Q(f)]\ ) \end{aligned}$$

- **Boltzmann-Gibbs transformation (updating Bayes' type rule)**
  [Positive and bounded potential function $G$]

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

# Basic notation

$E = \{1, \ldots, d\}$ integral operations $\rightsquigarrow$ matrix operations

$$\mu = [\mu(1), \ldots, \mu(d)] \qquad Q = (Q(i,j))_{1 \leq i,j \leq d} \qquad f = \begin{bmatrix} f(1) \\ \vdots \\ f(d) \end{bmatrix}$$

Indeed :

$$\mu Q = [(\mu Q)(1), \ldots, (\mu Q)(d)] \quad \& \quad Q(f) = \begin{bmatrix} Q(f)(1) \\ \vdots \\ Q(f)(d) \end{bmatrix}$$

with

$$(\mu Q)(j) = \sum_i \mu(i) Q(i,j) \quad \& \quad Q(f)(i) = \sum_j Q(i,j) f(j)$$

... and of course the duality formula

$$\mu(f) = \sum_i \mu(i) f(i)$$

# Basic notation

In terms of random variables :

$$\mu = \mathrm{Law}(X) \quad \text{and} \quad Q(x,dy) = \mathbb{P}(Y \in dy \mid X = x)$$

$$\Downarrow$$

$$\begin{array}{rcl} \mu(f) & = & \mathbb{E}(X) \\ (\mu Q)(dy) & = & \mathbb{P}(Y \in dy) \qquad Q(f)(x) = \mathbb{E}(f(Y) \mid X = x) \end{array}$$

Boltzmann-Gibbs transformation (updating Bayes' type rule)

$$\mu(dx) = dp(x) \quad \text{and} \quad G(x) = p(y \mid x)$$

$$\Downarrow$$

$$\Psi_G(\mu)(dx) = \frac{1}{\mu(G)}\ G(x)\ \mu(dx) = \frac{1}{p(y)}\ p(y \mid x)\ dp(x) = dp(x \mid y)$$

# Introduction

**Objective :** **Given a target probab.** $\eta(dx)$ **compute the map**

$$\eta \; : \; f \; \mapsto \; \eta(f) = \int f(x) \; \eta(dx) \quad ??$$

**Monte Carlo methods :**

$$\eta^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{X^i} \;\; \simeq_{N\uparrow} \;\; \eta$$

with two traditional types of algorithms

- $(X^i)_{i \geq 1}$ i.i.d. with common law $\eta$
- $(X^i)_{i \geq 1}$ Markov chain with invariant measure $\eta$

# The Importance sampling trick ($f > 0$)

**Key weight formula**

$$\eta(G) = \int G(x) \, \frac{d\eta}{d\pi}(x) \, \pi(dx) = \pi\left(G \; W\right)$$

with the weight function

$$W = \frac{d\eta}{d\pi}$$

**Sample $X^i$ i.i.d. with common law $\pi$ and set**

$$\eta^N = \frac{1}{N} \sum_{1 \leq i \leq N} W(X^i) \, \delta_{X^i}$$

**Note that**

$$N \, \mathrm{Var}\left(\eta^N(G)\right) = \eta\left(W \; G^2\right) - \eta(G)^2 = 0$$

for the updated twisted measure

$$\pi(dx) = \Psi_G(\eta)(dx) := \frac{1}{\eta(G)} \, G(x) \, \eta(dx) \quad (\Rightarrow W = \eta(G)/G)$$

# The Metropolis-Hasting model


From $x$ propose $x' \sim P(x, dx')$ and accept it with proba

$$a(x, x') := 1 \ \wedge \ \frac{\eta(dx')P(x', dx)}{\eta(dx)P(x, dx')}$$

The Markov transition $M(x, dx')$ is $\eta$-reversible

$$\begin{aligned}
\eta(dx)M(x, dx') &\overset{x \neq x'}{=} \eta(dx)P(x, dx') \times a(x, x') \\
&= [\eta(dx)P(x, dx')] \ \wedge \ [\eta(dx')P(x', dx)] \\
&= \eta(dx')M(x', dx)
\end{aligned}$$

$\Rightarrow$ Fixed point equation

$$\int \eta(dx) \ M(x, dx') = (\eta M)(dx') = \eta(dx')$$

$$\Updownarrow$$

$$\eta M = \eta$$

# The Metropolis-Hasting model

Markov chain samples

$$X_1 \xrightarrow{M} X_2 \xrightarrow{M} X_3 \xrightarrow{M} \ldots \xrightarrow{M} X_{n-1} \xrightarrow{M} X_n \xrightarrow{M} \ldots$$

with the Markov transport equation

$$\underbrace{\mathbb{P}\left(X_n \in dx_n\right)}_{=\eta_n(dx_n)} = \int \underbrace{\mathbb{P}\left(X_{n-1} \in dx_{n-1}\right)}_{=\eta_{n-1}(dx_{n-1})} \underbrace{\mathbb{P}\left(X_n \in dx_n \mid X_{n-1} = x_{n-1}\right)}_{=M(x_{n-1},dx_n)}$$

$$\Updownarrow$$

Linear measure valued equation :

$$\eta_n = \eta_{n-1} M \ \simeq_{n\uparrow\infty} \ \eta = \eta M$$

# Example 1

Boltzmann-Gibbs target measure :

$$\eta(dx) = \eta_n(dx) = \frac{1}{\mathcal{Z}_n}\ e^{-\beta_n V(x)}\ \lambda(dx) \quad \text{with} \quad \beta_n = 1/T_n$$

The M-H transition $M_n$ s.t. $\eta_n = \eta_n M_n$ :

- Proposition of moves $P(x, dx')$
- Acceptance rate

$$a_n(x, x') = 1 \wedge \left( e^{-\beta_n [V(x') - V(x)]} \times \left[ \frac{\lambda(dx') P(x', dx)}{\lambda(dx) P(x, dx')} \right] \right)$$

If $P$ is $\lambda$-reversible then we have (with $a_+ = \max(a, 0)$)

$$a_n(x, x') = e^{-\beta_n [V(x') - V(x)]_+} \rightsquigarrow \text{stochastic style steepest descent}$$

Some mixing pb. : $\beta_n$ large $\Rightarrow$ high rejection/local minima absorptions

# Example 2

Restriction probability:

$$\eta(dx) = \eta_n(dx) = \frac{1}{\mathcal{Z}_n}\, 1_{A_n}(x)\, \lambda(dx) \quad \text{with} \quad A_n \subset A$$

The M-H transition $M_n$ s.t. $\eta_n = \eta_n M_n$ = "Shaker of the set $A_n$"

- $\lambda$-reversible propositions $P(x, dx')$
- Acceptance iff $x' \in A_n$

*Example of $\lambda$-reversible moves :* $\lambda = \mathcal{N}(0,1) = \mathrm{Law}(W)$

$$x' = a\, x + \sqrt{1-a^2}\; W \quad \forall a \in [0,1]$$

Some mixing pb. : $a$ or $A_n$ too small $\Rightarrow$ high rejection

# Gibbs samplers $\subset$ Metropolis-Hasting model

On product state spaces

$$\eta(dx) := \eta(d(x_1, x_2)) \quad \text{on} \quad E = (E_1 \times E_2)$$

- Desintegration formulae :

$$\eta(d(x_1, x_2)) = \eta_1(dx_1)\ P_2(x_1, dx_2) = \eta_2(dx_2)\ P_1(x_2, dx_1)$$

$$x = (x_1, x_2) \xrightarrow{P_1} x' = (x_1', x_2) \xrightarrow{P_1} x'' = (x_1', x_2')$$

- Unit acceptance rates :

$$a_1(x, x') := 1 \ \wedge \ \frac{[\eta_2(dx_2) P_1(x_2, dx_1')]\ \ P_1(x_2, dx_1)}{[\eta_2(dx_2) P_1(x_2, dx_1)]\ \ P_1(x_2, dx_1')} = \mathbf{1} = a_2(x', x'')$$

Some basic notation

Monte Carlo methods

Measure valued equations

- Sequence of target probabilities
- Probability mass transport models
- Nonlinear Markov transport models
- Boltzmann-Gibbs measures

Feynman-Kac models

Interacting Monte Carlo models

# Measure valued equations

**Measure valued equations**

$=$

**Sequence of (target) probabilities (with $\uparrow$ complexity)**

$$\eta_0 \longrightarrow \eta_1 \longrightarrow \ldots \longrightarrow \eta_{n-1} \longrightarrow \eta_n \longrightarrow \eta_{n+1} \longrightarrow \ldots$$

*Examples :*

- Boltzmann-Gibbs w.r.t. $\beta_n \uparrow$

$$\eta_n(dx) = \frac{1}{\mathcal{Z}_n}\; e^{-\beta_n V(x)}\; \lambda(dx)$$

- Restriction models w.r.t. $A_n \downarrow$

$$\eta_n(dx) = \frac{1}{\mathcal{Z}_n}\; 1_{A_n}(x)\; \lambda(dx)$$

# Probability mass transport models

**Reminder :**

Given a function $G(x) \geq 0$, and a Markov transition $M(x, dy)$

Boltzmann-Gibbs transformation (= Bayes' type updating rule)

$$\Psi_G \ : \ \eta(dx) \mapsto \Psi_G(\eta)(dx) = \frac{1}{\eta(G)} \ G(x) \ \eta(dx)$$

Markov transport equation

$$M \ : \ \eta(dx) \mapsto (\eta M)(dx) = \int \eta(dx') \ M(x', dx)$$

# Key observation

**Boltzmann-Gibbs transform = Nonlinear Markov transport**

$$\Psi_G(\eta) = \eta S_{G,\eta}$$

with the Markov transition

$$S_{G,\eta}(x, dx') = \epsilon G(x)\delta_x(dx') + (1 - \epsilon G(x))\ \Psi_G(\eta)(dx')$$

for any $\epsilon \in [0,1]$ s.t. $\epsilon \|G\| \leq 1$

Proof :

$$S_{G,\eta}(f)(x) = \epsilon G(x) f(x) + (1 - \epsilon G(x))\ \Psi_G(\eta)(f)$$

$$\Downarrow$$

$$\eta(S_{G,\eta}(f)) \quad = \quad \epsilon\eta(Gf) + (1 - \epsilon\eta(G))\ \frac{\eta(Gf)}{\eta(G)} = \Psi_G(\eta)(f)$$

# Boltzmann-Gibbs measures

$$\eta_n(dx) := \frac{1}{\mathcal{Z}_n} \; e^{-\beta_n V(x)} \; \lambda(dx) \quad \text{with} \quad \beta_n \uparrow$$

- For any MCMC transition $M_n$ with target $\eta_n$

$$\eta_n = \eta_n M_n$$

- Updating of the temperature parameter

$$\eta_{n+1} = \Psi_{G_n}(\eta_n) \quad \text{with} \quad G_n = e^{-(\beta_{n+1}-\beta_n)V}$$

Proof : $e^{-\beta_{n+1}V} = e^{-(\beta_{n+1}-\beta_n)V} \times e^{-\beta_n V}$

Consequence :

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

$$\Downarrow$$

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

# Restriction models

$$\eta_n(dx) := \frac{1}{\mathcal{Z}_n}\, 1_{A_n}\, \lambda(dx) \quad \text{with} \quad A_n \downarrow$$

- For any MCMC transition $M_n$ with target $\eta_n$

$$\eta_n = \eta_n M_n$$

- Updating of the subset

$$\eta_{n+1} = \Psi_{G_n}(\eta_n) \quad \text{with} \quad G_n = 1_{A_{n+1}}$$

Proof : $\quad 1_{A_{n+1}} = 1_{A_{n+1}} \times 1_{A_n}$

Consequence :

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

$$\Downarrow$$

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

# Product models

$$\eta_n(dx) := \frac{1}{\mathcal{Z}_n} \left\{ \prod_{p=0}^{n} h_p(x) \right\} \lambda(dx) \quad \text{with} \quad h_p \geq 0$$

- For any MCMC transition $M_n$ with target $\eta_n = \eta_n M_n$.
- Updating of the product

$$\eta_{n+1} = \Psi_{G_n}(\eta_n) \quad \text{with} \quad G_n = h_{n+1}$$

Proof : $\left\{ \prod_{p=0}^{n+1} h_p \right\} = h_{n+1} \times \left\{ \prod_{p=0}^{n} h_p \right\}$

Consequence :

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

$$\Downarrow$$

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

# Feynman-Kac models


The solution of any measure valued equation of the form

$$\eta_n = \Psi_{G_{n-1}}(\eta_{n-1})M_n$$

is given by a normalized Feynman-Kac model of the form

$$\eta_n(f) = \gamma_n(f)/\gamma_n(1)$$

with the positive measure $\gamma_n$ defined by

$$\gamma_n(f) = \mathbb{E}\left(f(X_n)\prod_{p=0}^{n-1} G_p(X_p)\right)$$

where $X_n$ stands for the Markov chain with transitions

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

# Path space models

If we take the historical process

$$\mathbf{X}_n = (X_0, \ldots, X_n) \in \mathbf{E}_n = E^{n+1} \quad \text{and} \quad \mathbf{G}_p(\mathbf{X}_n) = G_n(X_n)$$

then we have

$$\gamma_n(f_n) \quad = \quad \mathbb{E}\left(f_n(\mathbf{X}_n) \prod_{p=0}^{n-1} \mathbf{G}_p(\mathbf{X}_p)\right) \Rightarrow \eta_n = \mathbb{Q}_n$$

**with the Feynman-Kac model on path space**

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

Evolution equation :

$$\mathbb{Q}_n = \Psi_{\mathbf{G_n}}(\mathbb{Q}_{n-1})\mathbf{M_n}$$

with the Markov transition $\mathbf{M_n}$ of the historical process $\mathbf{X_n}$


# Application domains extensions

- Confinements :

  RW $X_n \in \mathbb{Z}^d$, $X_0 = 0$ & $G_n := 1_{[-L,L]}$, $L > 0$.

$$\mathbb{Q}_n = \mathrm{Law}\left((X_0, \ldots, X_n) \mid X_p \in [-L, L],\ \forall 0 \leq p < n\right)$$

  and

$$\mathcal{Z}_n = \gamma_n(1) = \mathrm{Proba}\left(X_p \in [-L, L],\ \forall 0 \leq p < n\right)$$

- Self avoiding walks

$$\mathbf{X_n} = (X_0, \ldots, X_n) \quad \& \quad \mathbf{G_n}(\mathbf{X_n}) = 1_{X_n \not\in \{X_0, \ldots, X_{n-1}\}}$$

$$\mathbb{Q}_n = \mathrm{Law}\left((\mathbf{X_0}, \ldots, \mathbf{X_n}) \mid X_p \neq X_q,\ \forall 0 \leq p < q < n\right)$$

  and

$$\mathcal{Z}_n = \gamma_n(1) = \mathrm{Proba}\left(X_p \neq X_q,\ \forall 0 \leq p < q < n\right)$$

- Filtering:

$$M_n(x_{n-1}, dx_n) = p(x_n \mid x_{n-1})\ dx_n \quad \& \quad G_n(x_n) = p(y_n \mid x_n)$$

$$\Downarrow$$

$$\mathbb{Q}_{n+1} = \mathrm{Law}\left((X_0, \ldots, X_{n+1}) \mid Y_0 = y_0, \ldots, Y_n = y_n\right)$$

and

$$\mathcal{Z}_{n+1} = \gamma_{n+1}(1) = p(y_0, \ldots, y_n)$$

- Hidden Markov chain models = product models

$$\underbrace{dp(\theta \mid (y_0, \ldots, y_n))}_{\eta_n(d\theta)} \propto \left\{ \prod_{p=0}^{n} \underbrace{p(y_p \mid \theta, (y_0, \ldots, y_{p-1}))}_{h_p(\theta)} \right\} \underbrace{dp(\theta)}_{\lambda(d\theta)}$$

## $\supset$ *Continuous time models*

$$X_n := X'_{[t_n, t_{n+1}[} \quad \& \quad G_n(X_n) = \exp \int_{t_n}^{t_{n+1}} V_s(X'_s) ds$$

$$\Downarrow$$

$$\prod_{0 \leq p < n} G_p(X_p) = \exp \left\{ \int_{t_0}^{t_n} V_s(X'_s) ds \right\}$$

*or using a simple "Euler's scheme"* $X'_{t_p} = X_p$

$$e^{\int_{t_0}^{t_n} \left[ V_s(X'_s) ds + W_s(X'_s) dB_s \right]} \simeq \prod_{0 \leq p < n} e^{V_{t_p}(X_p)\ \Delta t + W_{t_p}(X_p) \sqrt{\Delta t}\ N_p(0,1)}$$

# A little analysis with 3 keys formulae

- Time marginal measures = Path space measures:

$$[\ \mathbf{X_n} := (X_0, \ldots, X_n)\ \ \&\ \ \mathbf{G_n}(\mathbf{X_n}) = G_n(X_n)] \implies \eta_n = \mathbb{Q}_n$$

- Normalizing constants (= Free energy models):

$$\mathcal{Z}_n = \mathbb{E}\left(\prod_{0 \le p < n} G_p(X_p)\right) = \prod_{0 \le p < n} \eta_p(G_p)$$

Proof ($\mathcal{Z}_n = \gamma_n(1)$) :

$$\gamma_{n+1}(1) = \mathbb{E}\left(G_n(X_n) \prod_{0 \le p \le (n-1)} G_p(X_p)\right) = \gamma_n(G_n) = \eta_n(G_n) \times \gamma_n(1)$$

# The last key

- Backward Markov models

$$\mathbb{Q}_n(d(x_0,\ldots,x_n)) \propto \eta_0(dx_0)Q_1(x_0,dx_1)\ldots Q_n(x_{n-1},dx_n)$$

with

$$\begin{aligned} Q_n(x_{n-1},dx_n) &:= G_{n-1}(x_{n-1})M_n(x_{n-1},dx_n) \\ &\overset{hyp}{=} H_n(x_{n-1},x_n)\ \nu_n(dx_n) \\ &\Rightarrow \eta_{n+1}(dx) = \frac{1}{\eta_n(G_n)}\ \eta_n\left(H_{n+1}(\,.\,,x)\right)\ \nu_{n+1}(dx) \end{aligned}$$

If we set

$$\mathbb{M}_{n+1,\eta_n}(x_{n+1},dx_n) = \frac{\eta_n(dx_n)\ H_{n+1}(x_n,x_{n+1})}{\eta_n\left(H_{n+1}(\,.\,,x_{n+1})\right)}$$

then we find the backward equation

$$\eta_{n+1}(dx_{n+1})\ \mathbb{M}_{n+1,\eta_n}(x_{n+1},dx_n) = \frac{1}{\eta_n(G_n)}\ \eta_n(dx_n)\ Q_{n+1}(x_n,dx_{n+1})$$

# The last key (continued)

$$\mathbb{Q}_n(d(x_0,\ldots,x_n)) \propto \eta_0(dx_0) Q_1(x_0,dx_1)\ldots Q_n(x_{n-1},dx_n)$$

$\oplus$

$$\eta_{n+1}(dx_{n+1})\ \mathbb{M}_{n+1,\eta_n}(x_{n+1},dx_n) \propto \eta_n(dx_n)\ Q_{n+1}(x_n,dx_{n+1})$$

$$\Downarrow$$

Backward Markov chain model :

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

**with the dual/backward Markov transitions**

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

# Interacting Monte Carlo models

**Objective : Solve a nonlinear measure valued equation**

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

**Two classes of interacting Monte Carlo models**

- **Interacting/Adaptive MCMC methods**
- **Mean field/Interacting particle systems**

# Interacting/Adaptive MCMC methods

**$\forall n$, run an interacting sequence of MCMC algorithms**

$$\begin{array}{ccccccccc} X_n^1 & \rightarrow & X_n^2 & \rightarrow & \ldots \rightarrow & X_n^j & \rightarrow \ldots\ldots\ldots\ldots & [\text{target } \eta_n\ ] & \\ X_{n+1}^1 & \rightarrow & X_{n+1}^2 & \rightarrow & \ldots \rightarrow & X_{n+1}^j & \rightarrow X_{n+1}^{j+1} \ldots & [\text{target } \eta_{n+1}\ ] \end{array}$$

$$\text{s.t.} \quad X_{n+1}^j \overset{M_\eta^{[n+1]}}{\rightsquigarrow} X_{n+1}^{j+1} \quad \text{depends on} \quad \eta = \eta_n^j := \frac{1}{j} \sum_{1 \leq i \leq j} \delta_{X_n^i} \simeq_{k \uparrow \infty} \eta_n$$

A single "fixed point" compatibility condition :

$$\forall \eta \qquad \Phi_{n+1}(\eta) M_\eta^{[n+1]} = \Phi_{n+1}(\eta)$$

**References**


- **A Functional Central Limit Theorem for a Class of Interacting MCMC Models** EJP (2009). (joint work with Bercu & Doucet)
- **Sequentially Interacting Markov chain Monte Carlo.** AoS (2010). (joint work with Brockwell & Doucet)
- **Interacting MCMC methods for solving nonlinear measure-valued equations.** AAP (2010) (joint work with Doucet)
- **Fluctuations of Interacting MCMC Models**. SPA (2012). (joint work with Bercu & Doucet)

# Mean field interacting particle models

**Key idea :** The solution of **any** measure valued process

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

can be seen as the law $\eta_n = \mathrm{Law}(\overline{X}_n)$ of a Markov chain $\overline{X}_n$ for some transitions

$$\mathbb{P}\left(\overline{X}_n \in dx_n \mid \overline{X}_{n-1} = x_{n-1}\right) = K_{n,\eta_{n-1}}(x_{n-1}, dx_n)$$

- **Notice that $\overline{X}_n$ = Perfect sampler**
- **Example : the Feynman-Kac updating-prediction mapping**

$$\Phi_n(\eta) = \Psi_{G_{n-1}}(\eta) M_n = \eta \underbrace{S_{G_{n-1},\eta} M_n}_{K_{n,\eta}} = \eta K_{n,\eta}$$

# Mean field particle interpretation

We approximate the exact/perfect transitions

$$\overline{X}_n \rightsquigarrow \overline{X}_{n+1} \sim K_{n+1,\eta_n}(\overline{X}_n, dx_{n+1})$$

by running a

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

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

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

$$\xi_n^i \rightsquigarrow \xi_{n+1}^i \sim K_{n+1,\eta_n^N}(\xi_n^i, dx_{n+1})$$

# 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 :

$$
\begin{array}{rcl}
\eta_n^N \simeq_{N\uparrow\infty} \eta_n & \Longrightarrow & K_{n+1,\eta_n^N} \simeq_{N\uparrow\infty} K_{n+1,\eta_n} \\
& \Longrightarrow & \xi_{n+1}^i \text{ almost iid copies of } \overline{X}_{n+1} \\
& \Longrightarrow & \eta_{n+1}^N \simeq_{N\uparrow\infty} \eta_{n+1}
\end{array}
$$

# Updating-prediction models $\rightsquigarrow$ genetic particle model :

$$
\left.\begin{array}{c}
\xi_n^1 \\ \vdots \\ \xi_n^i \\ \vdots \\ \xi_n^N
\end{array}\right]
\xrightarrow{S_{G_n,\eta_n^N}}
\left[\begin{array}{ccc}
\widehat{\xi}_n^1 & \xrightarrow{M_{n+1}} & \xi_{n+1}^1 \\
\vdots & & \vdots \\
\widehat{\xi}_n^i & \longrightarrow & \xi_{n+1}^i \\
\vdots & & \vdots \\
\widehat{\xi}_n^N & \longrightarrow & \xi_{n+1}^N
\end{array}\right]
$$

Accept/Reject/Recycling/Selection transition :

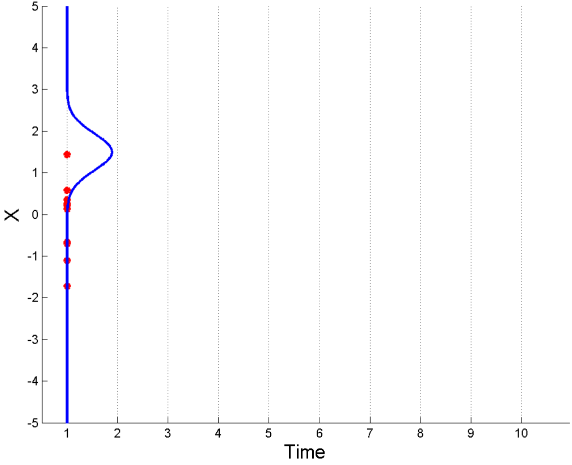
$$S_{G_n,\eta_n^N}(\xi_n^i, dx)$$

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

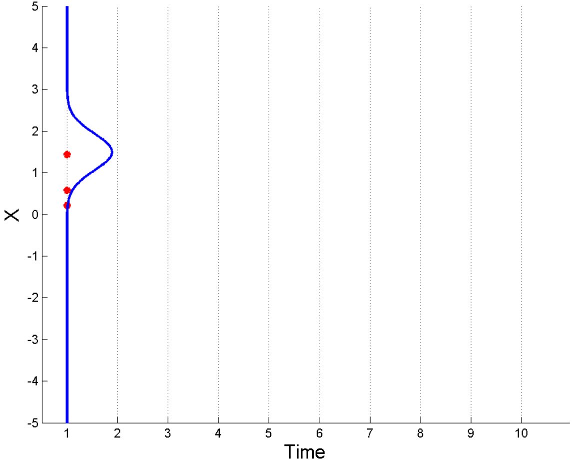
**Ex.** **:** $G_n = 1_A,\ \epsilon_n = 1 \rightsquigarrow G_n(\xi_n^i) = 1_A(\xi_n^i)$

$\hookrightarrow$ FK-Mean field particle models = *sequential Monte Carlo, population Monte Carlo, genetic algorithms, particle filters, pruning, spawning, reconfiguration, quantum Monte carlo, go with the winner...*

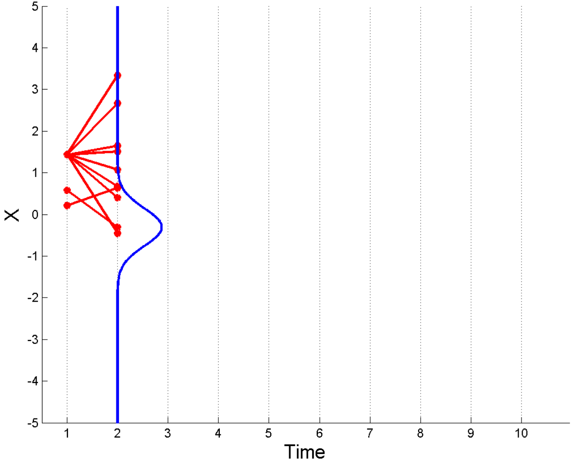
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



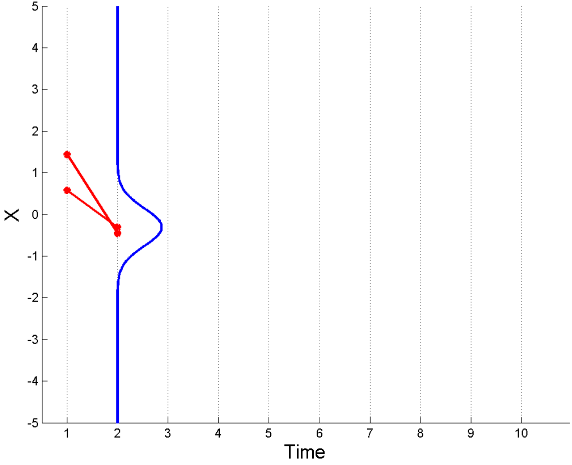
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



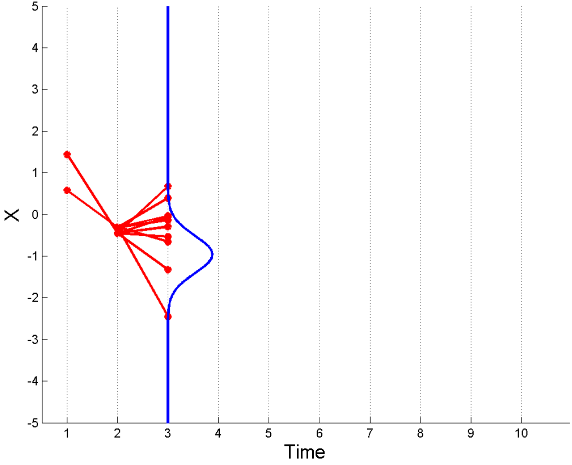
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



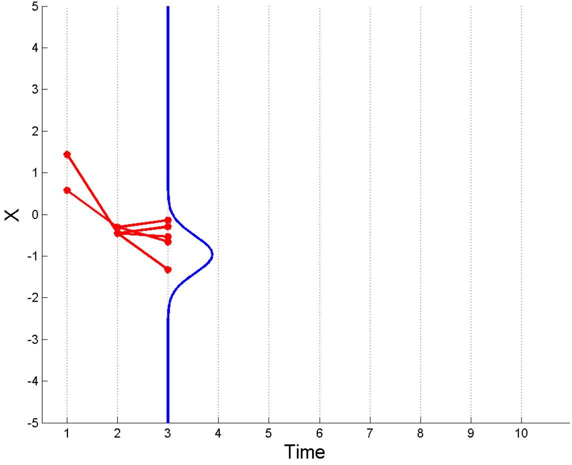
# Graphical illustration : $\eta_n^N := \frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi_n^i} \simeq \eta_n$



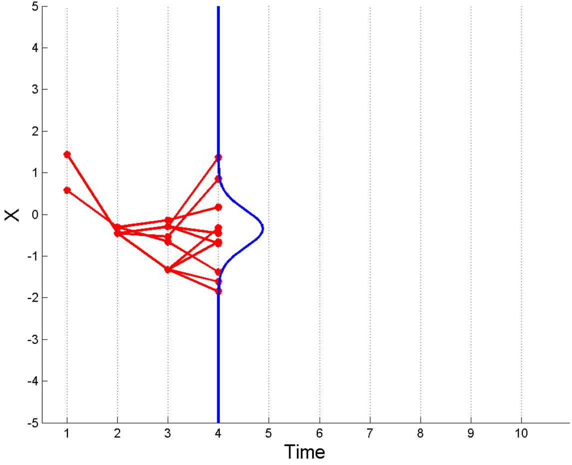
# Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



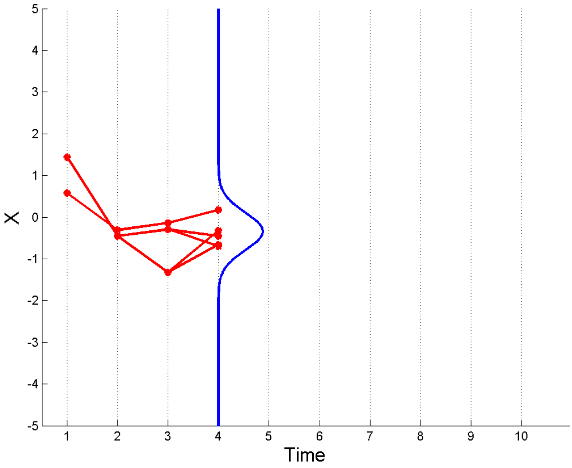
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



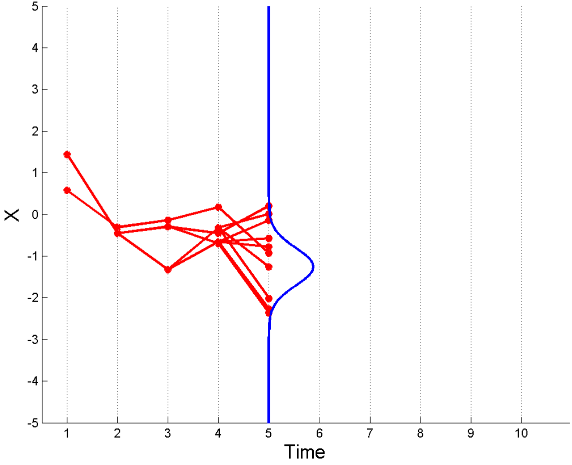
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



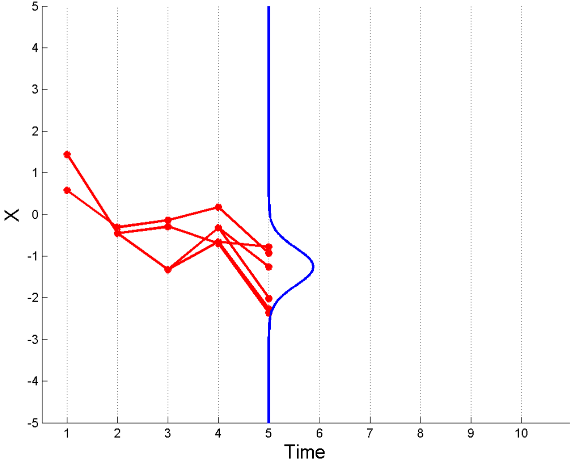
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



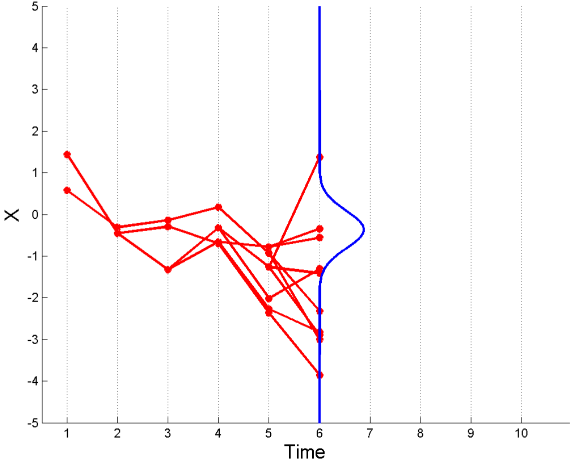
# Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



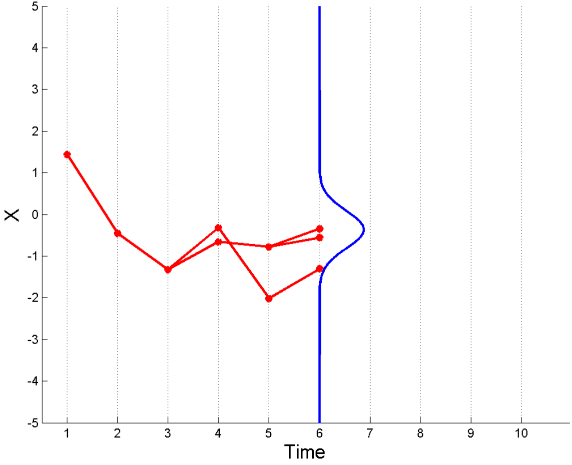
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



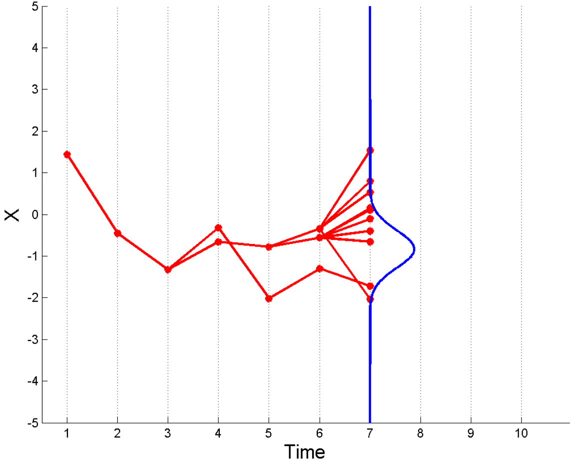
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



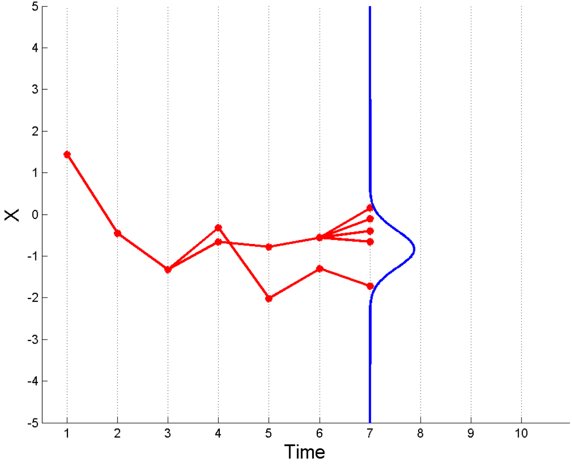
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



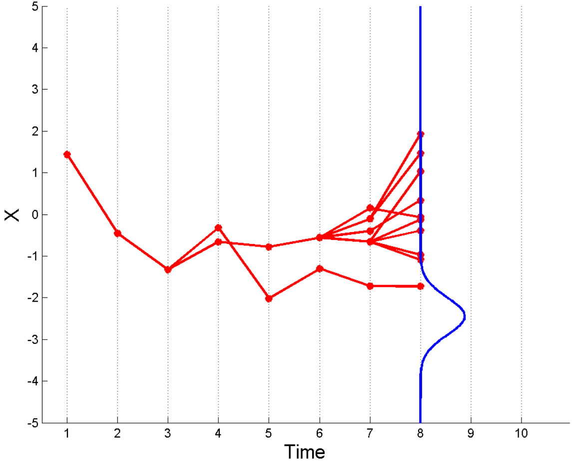
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



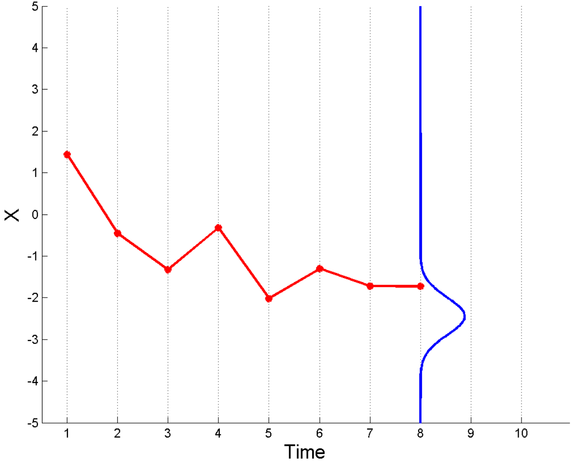
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



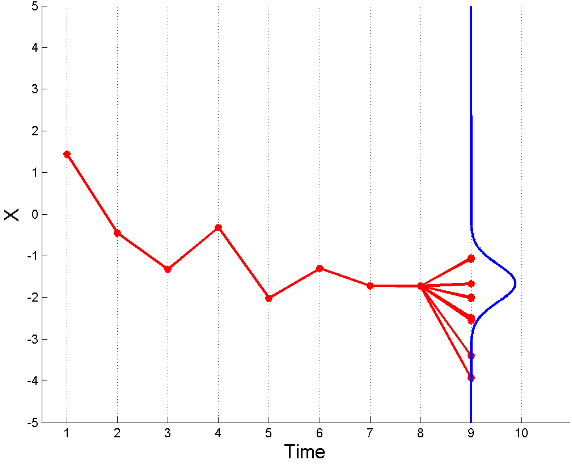
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \le i \le N} \delta_{\xi_n^i} \simeq \eta_n$



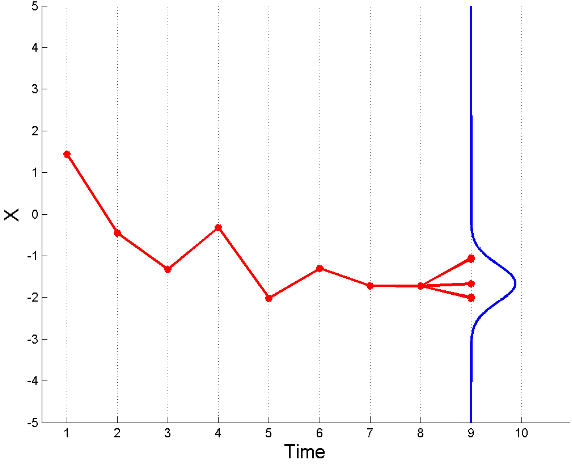
## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



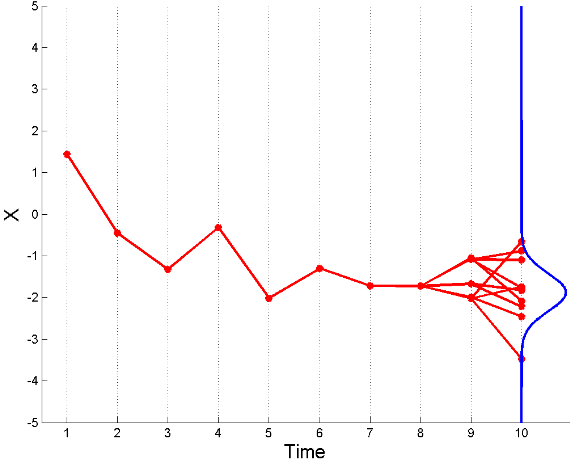
# Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \le i \le N} \delta_{\xi_n^i} \simeq \eta_n$



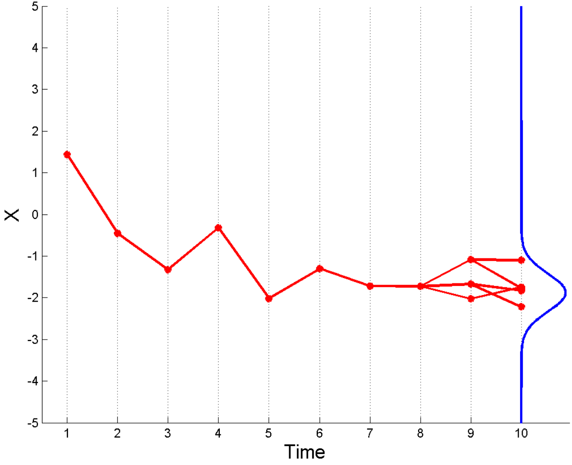
Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



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

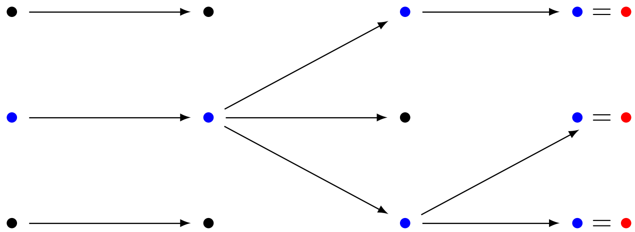


## Graphical illustration : $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq \eta_n$



# The 4 particle estimates

## Genealogical tree evolution $(N, n) = (3, 3)$

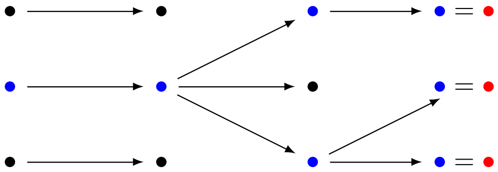


- Individuals in the current population

  $=$ *Almost* i.i.d. samples w.r.t. FK marginal meas. $\eta_n$

$$\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \longrightarrow_{N \to \infty} \eta_n = \textbf{solution of a nonlinear m.v.p.}$$

# Two more particle estimates



- Ancestral lines $=$ *Almost* i.i.d. sampled paths w.r.t. $\mathbb{Q}_n$.

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

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

- Unbiased particle free energy functions

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

# ... and the last particle measure

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

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

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

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

$$\Downarrow$$

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

# Island models

we can also write

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

with "the product style measure" on the sequence $\mathcal{X} = (\mathcal{X}_n)_n$

$$\Gamma_n(d\mathcal{X}) = \left\{\prod_{0\leq p<n} h_p(\mathcal{X})\right\} \mathbb{P}(d\mathcal{X}) \quad \text{with} \quad h_p(\mathcal{X}) = \mathcal{G}_p(\mathcal{X}_p)$$

$$\Downarrow$$

MCMC algorithms or their SMC version

# Island models (continued)

$$\pi_n(d\theta) = \frac{1}{\mathcal{Z}_n}\ \mathbb{E}\left(\prod_{p=0}^{n} G_{\theta,p}(X_{\theta,p})\right)\ \lambda(d\theta) = \frac{1}{\mathcal{Z}_n}\ \left\{\prod_{p=0}^{n} h_p(\theta)\right\}\ \lambda(d\theta)$$

with

$$h_p(\theta) = \eta_{\theta,p}(G_{\theta,p})$$

Examples :

$$\begin{aligned} G_{\theta,p}(x_p) = p(y_p \mid x_p, \theta) \quad &\Rightarrow \quad \pi_n(d\theta) = dp(\theta \mid (y_0, \ldots, y_n)) \\ G_{\theta,p}(x_p) = 1_{A_p}(x_p) \quad &\Rightarrow \quad \pi_n(d\theta) = dp(\theta \mid X_0 \in A_0, \ldots, A_n \in A_n) \end{aligned}$$

Unbiased property $\Rightarrow$ $\pi$ is the $\theta$-marginal of the product measure

$$\overline{\pi}_n(d\overline{\theta}) = \frac{1}{\mathcal{Z}_n}\ \left\{\prod_{p=0}^{n} \overline{h}_p(\overline{\theta})\right\}\ \overline{\lambda}(d\overline{\theta})$$

with

$$\overline{\theta} = (\theta, \xi)\ \sim\ \lambda(d\theta)P(\theta, d\xi) \quad \& \quad \overline{h}_p(\overline{\theta}) = \eta^N_{\theta,p}(G_{\theta,p})$$

# Island models (continued)

Metropolis-Hastings model on $(\overline{\theta} = (\theta, \xi))$ with target

$$\overline{\pi}_n(d\overline{\theta}) = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{p=0}^{n} \overline{h}_p(\overline{\theta}) \right\} \underbrace{\overline{\lambda}(d\overline{\theta})}_{=\lambda(d\theta)P(\theta,d\xi)}$$

Proposition transition

$$\overline{\theta} = (\theta, \xi) \longrightarrow \overline{\theta}' = (\theta', \xi') \sim \overline{Q}(\overline{\theta}, d\overline{\theta}') = Q(\theta, d\theta')\ P(\theta', d\xi')$$

Acceptance-Rejection rate

$$\begin{aligned} a(\overline{\theta}, \overline{\theta}') &= 1 \wedge \frac{\overline{\pi}_n(d\overline{\theta}')\overline{Q}(\overline{\theta}', d\overline{\theta})}{\overline{\pi}_n(d\overline{\theta})\overline{Q}(\overline{\theta}, d\overline{\theta}')} \\ &= \frac{\left\{\prod_{p=0}^{n} \overline{h}_p(\overline{\theta}')\right\}}{\left\{\prod_{p=0}^{n} \overline{h}_p(\overline{\theta})\right\}} \times \frac{\lambda(d\theta')Q(\theta', d\theta)}{\lambda(d\theta)Q(\theta, d\theta')} \end{aligned}$$

$\rightsquigarrow$ Particle MCMC (Andrieu-Doucet-Holenstein 2010).