Symposium, Univ. Bonn MDV Sept. 2006
Stochastic Algorithms and Markov Processes

Feynman-Kac particle models
Coalescent tree based functional representations

P. DEL MORAL, F. PATRAS, S. RUBENTHALER
Lab. J.A. Dieudonné, Univ. Nice Sophia Antipolis, France

$\hookrightarrow$ *Coalescent tree based functional representations for some Feynman-Kac particle models*
https://hal.ccsd.cnrs.fr/ccsd-00086532
$\hookrightarrow$ (delmoral@math.unice.fr)
$\hookrightarrow$ [preprints+info.] http://math1.unice.fr/ delmoral/

Introduction

- Evolutionary models and Feynman-Kac formulae
- Genetic genealogical models and Feynman-Kac limiting measures
- **Functional representations $\simeq$ precise propagations of chaos expansions**.
  - *Combinatorial differential calculus*
  - *Permutation group analysis of (colored) forests*
    (wreath product of permutation groups, Hilbert series techniques,...)
- (Applications).

Discrete time models $\rightsquigarrow$ *Continuous time version = Moran type genetic models*

($\sim$ joint works with L. Miclo, see also [PhD$\oplus$ articles] M. Rousset)

Evolutionary type models

Simple Genetic Branching Algo.	*Mutation*	*Selection/Branching*
Metropolis-Hastings Algo.	*Proposal*	*Acceptance/Rejection*
Sequential Monte Carlo methods	*Sampling*	*Resampling (SIR)*
Filtering/Smoothing	*Prediction*	*Updating/Correction*
Particle $\in$ Absorbing Medium	*Evolution*	*Killing/Creation/Anhiling*

<u>*Other Botanical Names:*</u> multi-level splitting (Khan-Harris 51), prune enrichment (Rosenbluth 1955), switching algo. (Magill 65), matrix reconfiguration (Hetherington 84), restart (Villen-Altamirano 91), particle filters (Rigal-Salut-DM 92), SIR filters (Gordon-Salmon-Smith 93, Kitagawa 96), go-with-the-winner (Vazirani-Aldous 94), ensemble Kalman-filters (Evensen 1994), quantum Monte Carlo methods (Melik-Nightingale 1999), sequential Monte Carlo Methods (Arnaud Doucet 2001), spawning filters (Fisher-Maybeck 2002), SIR Pilot Exploration Resampling (Liu-Zhang 2002),...

$\Longleftrightarrow$ Particle Interpretations of Feynman-Kac models

Since R. Feynman's phD. on path integrals 1942

Physics $\longleftrightarrow$ Biology $\longleftrightarrow$ Engineering Sciences $\longleftrightarrow$ Probability/Statistics

- **Physics** :
  - $FKS \in$ nonlinear integro-diff. éq. ($\sim$ generalized Boltzmann models).
  - Spectral analysis of Schrödinger operators and large matrices with nonnegative entries. (particle evolutions in disordered/absorbing media)
  - Multiplicative Dirichlet problems with boundary conditions.
  - Microscopic and macroscopic interacting particle interpretations.

- **Biology**:
  - Self-avoiding walks, macromolecular polymerizations.
  - Branching and genetic population models.
  - Coalescent and Genealogical evolutions.

- **Rare events analysis**:
  - Multisplitting and branching particle models (Restart).
  - Importance sampling and twisted probability measures.
  - Genealogical tree based simulation methods.

- **Advanced Signal processing**:
  - Optimal filtering/smoothing/regulation, open loop optimal control.
  - Interacting Kalman-Bucy filters.
  - Stochastic and adaptative grid approximation-models

- **Statistics/Probability**:
  - Restricted Markov chains (w.r.t terminal values, visiting regions,...)
  - Analysis of Boltzmann-Gibbs type distributions (simulation, partition functions,...).
  - Random search evolutionary algorithms, interacting Metropolis/simulated annealing algo.

Simple Genetic evolution/simulation models $\longrightarrow$ only 2 ingredients!!

(Discrete time parameter $n \in \mathbb{N} = \{0, 1, 2, ...\}$, state spaces E_n $(\in \{\mathbb{Z}^d, \mathbb{R}^d, \underbrace{\mathbb{R}^d \times \ldots \times \mathbb{R}^d}_{(n+1)-times}, ...\})$

- *Mutation/exploration/prediction/proposal* :

  $\rightarrow$ Markov transitions $M_n(x_{n-1}, dx_n)$ from E_{n-1} into E_n.

- *Selection/absorption/updating/acceptance* :

  $\rightarrow$ Potential functions G_n from E_n into $[0, 1]$.

A Genetic Evolution Model $\Rightarrow$ Markov chain $\xi_n = (\xi_n^1, \ldots, \xi_n^N) \in E_n^N = \underbrace{E_n \times \ldots \times E_n}_{N-times}$

$$\xi_n \in E_n^N \xrightarrow{\text{selection}} \widehat{\xi}_n \in E_n^N \xrightarrow{\text{mutation}} \xi_{n+1} \in E_{n+1}^N$$

- **Selection transition** ($\exists \neq$ types $\rightarrow$ Ex.: accept/reject)

$$\xi_n^i \rightsquigarrow \widehat{\xi}_n^i = \xi_n^i \quad \text{with proba. } G_n(\xi_n^i) \quad \textbf{[Acceptance]}$$

Otherwise we select a better fitted individual in the current configuration

$$\widehat{\xi}_n^i = \xi_n^j \quad \text{with proba. } G_n(\xi_n^j)/\sum_{k=1}^N G_n(\xi_n^k) \quad \textbf{[Rejection + Selection]}$$

- **Mutation transition**

$$\widehat{\xi}_n^i \rightsquigarrow \xi_{n+1}^i \sim M_{n+1}(\widehat{\xi}_n^i, \cdot\,)$$

A Genealogical tree model

Important observation [Historical process]

$$X'_n \in E'_n \quad \text{Markov chain}$$

$$\Downarrow$$

$$X_n = (X'_0, \ldots, X'_n) \in E_n = (E'_0 \times \ldots \times E'_n) \quad \text{Markov chain} \in \text{path spaces}$$

$\rightarrow$ *Markov transitions* $M_n(x_{n-1}, dx_n)$ [elementary extensions]

$$X_{n+1} = ((X'_0, \ldots, X'_n), X'_{n+1}) = (X_n, X'_{n+1})$$

Genetic Evolution Model on Path Spaces=Genealogical tree model

$$X_n = (X'_0, \ldots, X'_n) \quad \text{Markov transitions} \quad M_n \quad \text{and} \quad G_n(X_n) = G'_n(X'_n)$$

$$\downarrow$$

Genetic path-valued particle Model

$$\begin{cases} \xi^i_n &= (\xi^i_{0,n}, \xi^i_{1,n}, \ldots, \xi^i_{n,n}) \\ \widehat{\xi}^i_n &= (\widehat{\xi}^i_{0,n}, \widehat{\xi}^i_{1,n}, \ldots, \widehat{\xi}^i_{n,n}) \in E_n = (E'_0 \times \ldots \times E'_n) \end{cases}$$

- Path acceptance/(rejection+selection).
- Path mutation = path elementary extensions.

Occupation/Empirical measures ($\forall f_n$ test function on E_n)

$$\eta_n^N(f_n) = \frac{1}{N}\sum_{i=1}^N f_n(\xi_n^i) = \frac{1}{N}\sum_{i=1}^N f_n \underbrace{(\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i)}_{i\text{-th ancestral lines}}$$

$$\downarrow$$

Unbias-particle measures & Unnormalized Feynman-Kac measures :

$$\gamma_n^N(f_n) \quad = \quad \eta_n^N(f_n) \ \times \ \prod_{0\leq p<n} \eta_p^N(G_p) \longrightarrow_{N\to\infty} \gamma_n(f_n) = \mathbb{E}(f_n(X_n) \prod_{0\leq p<n} G_p(X_p))$$

Notes:

- $f_n = 1 \Rightarrow \gamma_n^N(1) = \prod_{0\leq p<n} \eta_p^N(G_p) \longrightarrow_{N\to\infty} \gamma_n(1) = \mathbb{E}(\prod_{0\leq p<n} G_p(X_p))$
- *Path-space models*

$$[\ X_n = (X'_0, \ldots, X'_n) \text{ and } G_n(X_n) = G'_n(X'_n)\ ] \Rightarrow \gamma_n(f_n) = \mathbb{E}(f_n(X'_0, \ldots, X'_n) \prod_{0\leq p<n} G'_p(X'_p))$$

$\Longrightarrow$ *Occupation measure & Normalized Feynman-Kac measures:*

$$\eta_n^N(f_n) \quad = \quad \frac{1}{N}\sum_{i=1}^{N} f_n(\xi_n^i) = \gamma_n^N(f_n)/\gamma_n^N(1) \longrightarrow_{N\to\infty} \eta_n(f_n) = \gamma_n(f_n)/\gamma_n(1)$$

Path-space models

$$[\ X_n = (X'_0,\ldots,X'_n) \text{ and } G_n(X_n) = G'_n(X'_n)\ ]$$

$$\Downarrow$$

$$\eta_n(f_n) = \frac{\mathbb{E}(f_n(X'_0,\ldots,X'_n)\prod_{0\leq p<n} G'_p(X'_p))}{\mathbb{E}(\prod_{0\leq p<n} G'_p(X'_p))}$$

Note:

$$\gamma_n(f_n) = \eta_n(f_n) \times \prod_{0\leq p<n} \eta_p(G_p) \quad (\longleftarrow \gamma_n^N(f_n) = \eta_n^N(f_n) \times \prod_{0\leq p<n} \eta_p^N(G_p))$$

Motivating example $\longrightarrow$ filtering/hidden Markov chains/Bayesian Stat.

Signal process $\quad X_n$ = Markov chain $\in E_n$

Observation/Sensor eq. $\quad Y_n = H_n(X_n, V_n) \in F_n$ with $\mathbb{P}(H_n(x_n, V_n) \in dy_n) = g_n(x_n, y_n)\ \lambda_n(dy_n)$

Example: $Y_n = h_n(X_n) + V_n \in F_n = \mathbb{R}$, with Gaussian noise $V_n = \mathcal{N}(0,1)$

$$\Downarrow$$

$$\mathbb{P}(h_n(x_n) + V_n \in dy_n) = (2\pi)^{-1/2} e^{-\frac{1}{2}(y_n - h_n(x_n))^2}\ dy_n = \underbrace{\exp\left[h_n(x_n)y_n - h_n^2(x_n)/2\right]}_{g_n(x_n,y_n)}\ \underbrace{\mathcal{N}(0,1)(dy_n)}_{\lambda_n(dy_n)}$$

Prediction/filtering/smoothing $\rightarrow$ **Feynman-Kac representation** $G_n(x_n) = g_n(x_n, y_n)$

$$\eta_n = \text{Law}(X_n \mid Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}) = \text{Law}(X'_0, \ldots, X'_n \mid Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1})$$

Rather complete asymptotic theory $(n,N)\to\infty$ (usual LLN, CLT, LDP,...)

$\hookrightarrow$ *F-K Formulae, Genealogical and IPS*, *Springer (2004)*+**References therein**

Some examples:

- *Weak convergence* [$p\geq 1$ + $\mathcal{F}_n$ not too large + regular mutations]
  (JTP 2000, joint work with M. Ledoux)

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

$$\text{Ex : } E_n=\mathbb{R},\quad \mathcal{F}_n=\{1_{]-\infty,x]}\ ;\ x\in\mathbb{R}\}\Rightarrow \sup_{n\geq 0}\mathbb{E}(\sup_{x\in\mathbb{R}}|\eta_n^N(1_{]-\infty,x]})-\eta_n(1_{]-\infty,x]})|^p)^{1/p}\leq c(p)/\sqrt{N}$$

- *Propagation-of-chaos estimates* [$q\leq N$ finite block size]
  (TVP+SIAM PTA 2006, joint work with A. Doucet)

$$\mathbb{P}_{n,q}^N:=\text{Law}(\xi_n^1,\ldots,\xi_n^q)\simeq\eta_n^{\otimes q}+\frac{1}{N}\,\partial^1\mathbb{P}_{n,q}\quad\text{with}\quad\partial^1\mathbb{P}_{n,q}\quad\text{signed meas. s.t. }\sup_{n\geq 0}\|\partial^1\mathbb{P}_{n,q}\|_{\text{tv}}\leq c\ q^2$$

Problem :

Pb : Find a functional representation at any order?

$$\mathbb{P}^N_{n,q} \simeq \eta_n^{\otimes q} + \frac{1}{N}\ \partial^1 \mathbb{P}_{n,q} + \ldots + \frac{1}{N^k}\ \partial^k \mathbb{P}_{n,q} + \frac{1}{N^{k+1}}\ \partial^k \mathbb{P}^N_{n,q}$$

with a bounded remainder measure $\sup_{N\geq 1} \|\partial^{k+1}\mathbb{P}^N_{n,q}\|_{\text{tv}} < \infty$

Consequences :

- Sharp + strong propagations of chaos estimates at any order.
- Wick product formulae on forests.
- Sharp $\mathbb{L}_p$-mean error bounds.
- Law of large numbers for U-statistics for interacting processes.
- ...

Tensor product measures

$$\rightsquigarrow \quad (\eta_n^N)^{\otimes q} = \frac{1}{N^q} \sum_{a \in [N]^{[q]}} \delta_{\xi_n^a} \quad \text{and} \quad (\eta_n^N)^{\odot q} = \frac{1}{(N)_q} \sum_{a \in \langle q,N \rangle} \delta_{\xi_n^a}$$

with

$$\begin{cases} \xi_n^a & := & (\xi_n^{a(1)}, \ldots, \xi_n^{a(q)}) \\ [N]^{[q]} & := & N^q \text{ mappings } [q] := \{1, \ldots, q\} \rightsquigarrow [N] := \{1, \ldots, N\}; \\ \langle q, N \rangle & := & (N)_q := N!/(N-q)! \text{ one-to-one mappings} \end{cases}$$

$$\textbf{Note :}\quad \mathbb{E}((\eta_n^N)^{\odot q}(F)) = \mathbb{P}_{n,q}^N(F) \quad \text{and} \quad (\eta_n^N)^{\otimes q} = (\eta_n^N)^{\odot q}\left(\frac{1}{N^q}\sum_{b\in[q]^{[q]}} \frac{(N)_{|b|}}{(q)_{|b|}}\, D_b\right)$$

with $|b| = \text{Card}(b([q]))$ and the coalescent-selection transitions

$$D_b(F)(x^1,\ldots,x^q) := F(x^{b(1)},\ldots,x^{b(q)}) = F(x^b)$$

$$\Downarrow$$

$$\delta_{x^a}D_b(F) = D_aD_b(F)(x^a) = D_{ab}(F)(x) \Longleftrightarrow D_aD_b = D_{ab}$$

Proof :

- $\forall c \in [N]^{[q]} \quad \exists (N-|c|)_{q-|c|} \times (q)_{|c|} \neq$ ways to write $c = ab \in \langle q, N\rangle \circ [q]^{[q]}$
- $a \in \langle q, N\rangle \Longrightarrow |b| = |c|$ and $\frac{(N)_{|c|}}{(q)_{|c|}} \times \frac{(N-|c|)_{q-|c|}\times(q)_{|c|}}{(N)_q} = 1$

Unnormalized (tensor product) measures

$$\gamma_n^N(f) := \gamma_n^N(1) \times \eta_n^N(f) \quad \text{with} \quad \gamma_n^N(1) = \prod_{0 \le p < n} \eta_p^N(G_p) \Longrightarrow \eta_n^N(f) = \gamma_n^N(f)/\gamma_n^N(1)$$

> $\gamma_n^N \sim$ *Martingale end point :*
>
> $$\mathbb{E}(\gamma_n^N(f)) = \gamma_n(f) \quad \text{but} \quad \mathbb{E}(\eta_n^N(f)) = \mathbb{P}_{n,1}^N(f) \neq \eta_n(f) \Rightarrow \text{bias}$$

Proof :

$$\mathbb{E}(\gamma_n^N(f) \mid \xi_{n-1}) = \gamma_{n-1}^N Q_n(f) \quad \text{and} \quad \eta_n^N(f) - \eta_n(f) = \underbrace{\frac{\gamma_n^N(1)}{\gamma_n(1)}}_{\neq 1} \times \gamma_n^N \left(\frac{1}{\gamma_n(1)} (f - \eta_n(f)) \right)$$

with the positive FKS operator $Q_n(x, dy) = G_{n-1}(x) M_n(x, dy)$ $(\to \gamma_n = \gamma_{n-1} Q_n)$

Unnormalized tensor product measures

$$\rightsquigarrow \quad (\gamma_n^N)^{\otimes q} := \gamma_n^N(1)^q \times (\eta_n^N)^{\otimes q} \quad \text{and} \quad (\gamma_n^N)^{\odot q} := \gamma_n^N(1)^q \times (\eta_n^N)^{\odot q}$$

Lemma

$$\begin{aligned}\mathbb{Q}_{n,q}^N(F) &:= \mathbb{E}((\gamma_n^N)^{\otimes q}(F)) \\ &= \frac{1}{N^q} \sum_{a \in [q]^{[q]}} \frac{(N)_{|a|}}{(q)_{|a|}} \, \mathbb{Q}_{n-1,q}^N(Q_n^{\otimes q} D_a F) = \cdots = \frac{1}{N^{q(n+1)}} \sum_{\mathbf{a} \in \mathcal{A}_{n,q}} \frac{(\mathbf{N})_{|\mathbf{a}|}}{(\mathbf{q})_{|\mathbf{a}|}} \, \Delta_{n,q}^{\mathbf{a}}(F)\end{aligned}$$

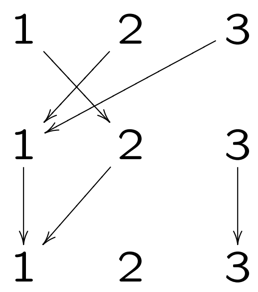
with the measure-valued functional

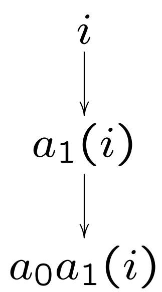
$$\Delta_{n,q} \ : \ \mathbf{a} = (a_0, \ldots, a_n) \in \mathcal{A}_{n,q} \mapsto \Delta_{n,q}^{\mathbf{a}} = \left(\eta_0^{\otimes q} D_{a_0} Q_1^{\otimes q} D_{a_1} \ldots Q_n^{\otimes q} D_{a_n}\right) \in \mathcal{M}(E_n^q)$$

Traditional multi-index notation :

$|\mathbf{a}| = (|a_0|, \ldots, |a_n|)$ and $(\mathbf{N})_{|\mathbf{a}|} = (N)_{|a_0|} \ldots (N)_{|a_n|}$ and $|\mathbf{a}|! = |a_0|! \ldots |a_n|!$ and so on.

Example $q = 3$:





$$\mathbf{a} = (a_0, a_1) \Rightarrow \Delta^{\mathbf{a}}_{n,q}(F) = \int \eta_0(dx^1)\eta_0(dx^2)\eta_0(dx^3)$$

$$Q_1(x^1, dy^1)Q_1(x^1, dy^2)Q_1(x^3, dy^3)$$

$$Q_2(y^1, dz^2)Q_2(y^1, dz^3)Q_2(y^2, dz^1)F(z^1, z^2, z^3)$$

Stirling Formula

$$(N)_p = \sum_{l \le p} s(p,l) N^l \Longrightarrow \forall \mathbf{p} = (p_0, \ldots, p_{n+1}) \quad (\mathbf{N})_\mathbf{p} = \sum_{\mathbf{l} \le \mathbf{p}} s(\mathbf{p}, \mathbf{l}) N^{|\mathbf{l}|}$$

Consequence :

(with $|\mathbf{p}| =: \sum_{0 \le k \le n} p_k$ and $\mathbf{q} = (q)_{0 \le k \le n}$)

$$\mathbb{Q}^N_{n,q}(F) = \sum_{\mathbf{r} < \mathbf{q}} \sum_{\mathbf{q}-\mathbf{r} \le \mathbf{p} \le \mathbf{q}} s(\mathbf{p}, \mathbf{q} - \mathbf{r}) \frac{1}{N^{|\mathbf{r}|}} \frac{1}{(\mathbf{q})_\mathbf{p}} \sum_{\mathbf{a} \in \mathcal{A}_{n,q} : |\mathbf{a}| = \mathbf{p}} \Delta^{\mathbf{a}}_{n,q}(F)$$

Def :

$$\mathcal{A}_{n,q}(\mathbf{r}) := \{\mathbf{a} \in \mathcal{A}_{n,q} : |\mathbf{a}| \geq \mathbf{q} - \mathbf{r}\} \text{ (less than } \mathbf{r} \text{ coalescences)}$$

$$\Downarrow$$

Th:

$$\mathbb{Q}_{n,q}^{N} = \gamma_n^{\otimes q} + \sum_{1 \leq k \leq (q-1)(n+1)} \frac{1}{N^k} \, \partial^k \mathbb{Q}_{n,q}$$

with the measure valued partial derivatives

$$\partial^k \mathbb{Q}_{n,q} = \sum_{\mathbf{r}<\mathbf{q}:|\mathbf{r}|=\mathbf{k}} \; \sum_{\mathbf{a} \in \mathcal{A}_{n,q}(\mathbf{r})} s(|\mathbf{a}|, \mathbf{q}-\mathbf{r}) \, \frac{1}{(\mathbf{q})_{|\mathbf{a}|}} \, \Delta_{n,q}^{\mathbf{a}}$$

Symmetric group left action

$$\mathbf{s} = (s_0, ..., s_{n+1}) \in \mathcal{G}_q^{n+2} \rightsquigarrow \mathbf{s(a)} := (s_0 a_0 s_1^{-1}, s_1 a_1 s_2^{-1}, ..., s_n a_n s_{n+1}^{-1})$$

$$\Downarrow$$

(on sym. test functions) $\quad \Delta_{n,q}^{\mathbf{a}} = \Delta_{n,q}^{\mathbf{s(a)}} \quad \Leftrightarrow$ (leaves re-labelling)

$$\Downarrow$$

Note : $\mathcal{A}_{n,q}/_{\sim} \simeq$ weakly increasing maps sequences $\simeq$ Forests $\mathcal{F}_{n,q}$

$$\Downarrow$$

$$\boxed{\partial^k \mathbb{Q}_{n,q} = \sum_{\mathbf{r}<\mathbf{q}:|\mathbf{r}|=\mathbf{k}} \sum_{\mathbf{f} \in \mathcal{F}_{n,q}(\mathbf{r})} \frac{1}{(\mathbf{q})_{|\mathbf{f}|}} \; s(|\mathbf{f}|, \mathbf{q}-\mathbf{r}) \; \#(\mathbf{f}) \; \Delta_{n,q}^{\mathbf{f}}}$$

with $\#(\mathbf{f})$:= nb of elts in the equivalence class ($\mathcal{A}_{n,q} \simeq$ entangled graphs:=jungles)

Example :

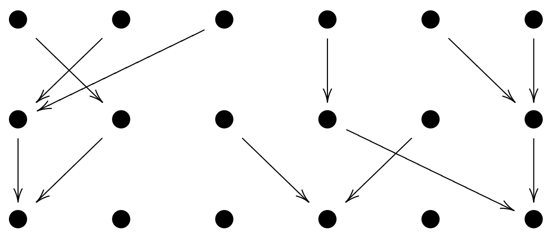


Figure 1: The entangled graph representation of a jungle with the same underlying graph as the planar forest in Fig. 2.

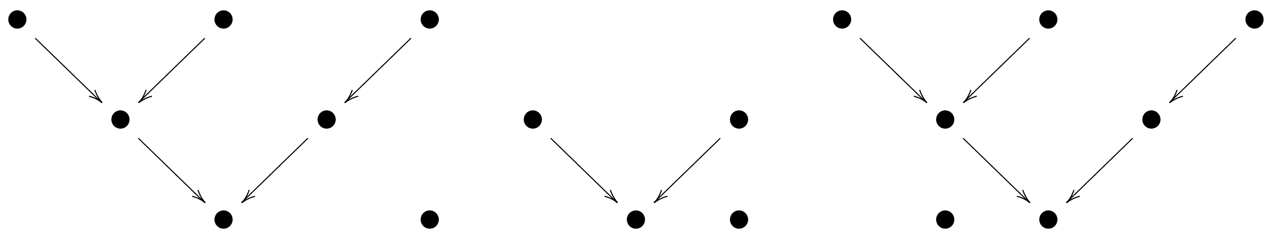


Figure 2: a graphical representation of a planar forest $\mathbf{f} = \mathbf{t}_1\mathbf{t}_3\mathbf{t}_2\mathbf{t}_3\mathbf{t}_3\mathbf{t}_1$ in terms of planar trees (corresponding forest $\mathbf{t}_1^2\mathbf{t}_2\mathbf{t}_3^3$=normal form).

Definitions :

$$\begin{array}{rcl} B(\mathbf{t}) & = & \text{the forest deduced from cutting the root of tree } \mathbf{t} \\ B^{-1}(\mathbf{f}) & = & \text{the tree deduced from the forest } \mathbf{f} \text{ by adding a root.} \end{array}$$

Symmetry multisets :

$$\begin{array}{rcl} \mathbf{t} = B^{-1}(\mathbf{t}_1^{m_1} \dots \mathbf{t}_k^{m_k}) \Rightarrow \mathbf{S}(\mathbf{t}) & := & (m_1, \dots, m_k) \\ \mathbf{S}(\mathbf{t}_1^{m_1} \dots \mathbf{t}_k^{m_k}) & := & \left(\underbrace{\mathbf{S}(\mathbf{t}_1), \dots, \mathbf{S}(\mathbf{t}_1)}_{m_1-\text{terms}}, \dots, \underbrace{\mathbf{S}(\mathbf{t}_k), \dots, \mathbf{S}(\mathbf{t}_k)}_{m_k-\text{terms}} \right) \end{array}$$

$\Downarrow$ (class formula + recursive multiplication principles)

Th. [closed formula]:

$$\forall \mathbf{f} \in \mathcal{F}_{q,n} \qquad \#(\mathbf{f}) = (q!)^{n+2} / \prod_{i=-1}^{n} \mathbf{S}(B^i(\mathbf{f}))!$$

$\oplus$ (Hilbert series tech. $\rightsquigarrow$ # forests with a given type)

Definitions :

$$\begin{array}{rcl} \mathcal{B}_0^{sym}(E_n^q) & = & F \text{ on } E_n^q \text{ such that } \int F(x_1,\ldots,x_{q-1},x_q)\ \gamma_n(dx_q) = 0. \\ \mathbf{t}_k & = & \text{the tree with a single coal. at level } k \text{ (its two leaves at level } (n+1)) \\ \mathbf{u}_k & = & \text{the trivial tree of height } k. \end{array}$$

$$\Downarrow$$

Cor.: $\forall q$ even $\leq N,\ F \in \mathcal{B}_0^{sym}(E_n^q)$

$$\forall k < q/2 \quad \partial^k \mathbb{Q}_{n,q}(F) = 0, \quad \partial^{q/2}\mathbb{Q}_{n,q}(F) = \sum_{\mathbf{r}<\mathbf{q},|\mathbf{r}|=\frac{q}{2}} \frac{q!}{2^{q/2}\ \mathbf{r}!}\ \Delta_{n,q}^{\mathbf{f_r}} F$$

with

$$\mathbf{r} = (r_k)_{0\leq k\leq n} < \mathbf{q} = (q)_{0\leq k\leq n} \rightsquigarrow \mathbf{f_r} := \mathbf{t}_0^{r_0}\mathbf{u}_0^{r_0}\ldots\mathbf{t}_n^{r_n}\mathbf{u}_n^{r_n}$$

($\forall q$ odd $\leq N$, the partial derivatives are the null measure on $\mathcal{B}_0^{sym}(E_n^q)$, up to any order $k \leq \lfloor q/2 \rfloor$)$\oplus$($\exists$ Gaussian field interpretation)

Extension $\mathbb{Q}^N_{n,q} \rightsquigarrow \mathbb{P}^N_{n,q}$ **:**

Same type of results + a remainder unif. bounded measure

$\longrightarrow$ *$\sim$ techniques $\oplus$ 3 main ingredients*

- $\mathbb{E}((\gamma_n^N)^{\otimes q}(F)) \rightsquigarrow \mathbb{E}([(\gamma_0^N)^{\otimes q_0} \otimes \ldots \otimes (\gamma_n^N)^{\otimes q_n}](F))$
- Forests $\rightsquigarrow$ colored forests
- $\gamma_n^N \rightsquigarrow \eta_n^N \Longrightarrow$ renormalisation techniques.

Applications :

- Particle physics (absorbing medium, ground states)
- Biology (polymers, macromolecules)
- Statistics (particle simulation, restricted Markov, target distributions)
- Rare event analysis (importance sampling, multilevel branching)
- Signal processing, filtering

Particle physics: Markov $X_n \in$ Absorbing medium $G(x) = e^{-V(x)} \in [0,1]$

$$X_n^c \in E^c = E \cup \{c\} \xrightarrow{absorption} \widehat{X}_n^c \xrightarrow{exploration} X_{n+1}^c$$

Absorption/killing: $\longrightarrow \widehat{X}_n^c = X_n^c$, with proba $G(X_n^c)$; otherwise the particle is killed and $\widehat{X}_n^c = c$.

$$\Downarrow$$

$$\begin{aligned} A &= \{x \,:\, G(x) = 0\} \longrightarrow \text{Hard obstacles} \\ T &= \inf\{n \geq 0 \,;\, \widehat{X}_n^c = c\} \longrightarrow \text{Absorption time} \quad X_{T+n}^c = \widehat{X}_{T+n}^c = c \end{aligned}$$

$\Longrightarrow$ **Feynman-Kac models (G, X_n) :** $\gamma_n = \text{Law}(X_n^c \,;\, T \geq n)$ and $\gamma_n(1) = \text{Proba}(T \geq n)$

$$\Downarrow$$

$$\eta_n = \text{Law}(X_n^c \mid T \geq n) = \text{Law}((X_0'^c, \ldots, X_n'^c) \mid T \geq n)$$

Biology: Macromolecules and Directed Polymers

- *Self avoiding walks* $X'_n \in \mathbb{Z}^d$

$$X_n = (X'_0, \ldots, X'_n) \quad \text{and} \quad G_n(X_n) = 1_{\not\in \{X'_0, \ldots, X'_{n-1}\}}(X'_n)$$

$$\gamma_n(1) = \text{Proba}(\forall 0 \leq p \neq q \leq n,\ \ X'_p \neq X'_q) \quad \text{and} \quad \eta_n = \text{Law}(X'_0, \ldots, X'_n \mid \forall 0 \leq p \neq q \leq n,\ \ X'_p \neq X'_q)$$

- *Edwards' model*

$$X_n = (X'_0, \ldots, X'_n) \quad \text{and} \quad G_n(X_n) = \exp\{-\beta \sum_{0 \leq p < n} 1_{X'_p}(X'_n)\}$$

Statistics: Sequential MCMC and Feynman-Kac-Metropolis models

Metropolis potential [π target measure]+[(K, L) pair Markov transitions]

$$G(y_1, y_2) = \frac{\pi(dy_2)L(y_2, dy_1)}{\pi(dy_1)K(y_1, dy_2)}$$

Ex. π Gibbs measure:

$$\pi(dy) \propto e^{-V(y)} \ \lambda(dy) \Rightarrow G(y_1, y_2) = e^{(V(y_1)-V(y_2))} \ \frac{\lambda(dy_2)L(y_2, dy_1)}{\lambda(dy_1)K(y_1, dy_2)}$$

<u>*Note:*</u> $(K = L \ \ \lambda-\text{reversible})$ or $\left(\lambda K = \lambda \text{ and } L(y_2, dy_1) = \lambda(dy_1)\frac{dK(y_1, \centerdot)}{d\lambda}(y_2)\right)$

$$\Downarrow$$

$$G(y_1, y_2) = \exp\left(V(y_1) - V(y_2)\right)$$

Notation $\mathbb{E}_{\nu}^{M}(\,.\,)$=Expectation w.r.t. Markov [transition M, initial condition ν]

Theorem: (**Time reversal formula**), [*A. Doucet, P.DM; (Séminaire Probab. 2003)*]

$$\mathbb{E}_{\pi}^{L}(f_n(Y_n, Y_{n-1}\ldots, Y_0)|Y_n = y) = \frac{\mathbb{E}_y^K(f_n(Y_0, Y_1, \ldots, Y_n)\ \{\prod_{0\leq p<n} G(Y_p, Y_{p+1})\})}{\mathbb{E}_y^K(\{\prod_{0\leq p<n} G(Y_p, Y_{p+1})\})}$$

In addition :

$\oplus$ *FK-Metropolis n-marginal:* $\lim_{n\to\infty}\eta_n = \pi$ (cv. decays $\perp \pi$)

$\oplus$ *Nonhomogeneous models:* (π_n, L_n, K_n)

$\pi_n(dy) \propto e^{-\beta_n V(y)}\ \lambda(dy)$, cooling schedule $\beta_n \uparrow \infty$, mutation s.t. $\pi_n = \pi_n K_n$, and Law$(X_0) = \pi_0$

$$\Downarrow$$

$$G_n(y_1, y_2) = \exp\left[-(\beta_{n+1} - \beta_n)V(y_1)\right] \Longrightarrow \eta_n = \pi_n$$

Rare events analysis

- **Importance sampling and Twisted Feynman-Kac measures**

$$\mathbb{P}(V_n(X_n) \geq a) \quad = \quad \mathbb{E}(1_{V_n(X_n)\geq a}\ e^{-\beta_n V_n(X_n)}\ e^{+\beta_n V_n(X_n)})$$

$$\Downarrow$$

Importance potentials/measures:

$$G_n(X_n, X_{n+1}) = e^{\beta_n(V_{n+1}(X_{n+1})-V_n(X_n))} \Longrightarrow \mathbb{P}(V_n(X_n) \geq a) = \gamma_n(1_{V_n\geq a}e^{-\beta_n V_n})$$

In addition:

$$\mathbb{E}(f_n(X_n) \mid V_n(X_n) \geq a) = \eta_n(f_n\ 1_{V_n\geq a}e^{-\beta_n V_n})/\eta_n(1_{V_n\geq a}e^{-\beta_n V_n})$$

$\oplus$ *Path-space models $\Rightarrow$ weighted genealogies*

$$X_n = (X'_0, \ldots, X'_n) \text{ and } V_n(X_n) = V'_n(X'^n)$$

$$\Downarrow$$

$$\mathbb{E}(f_n(X'_0, \ldots, X'_n) \mid V'_n(X'_n) \geq a) = \eta_n(f_n\ 1_{V_n\geq a}e^{-\beta_n V_n})/\eta_n(1_{V_n\geq a}e^{-\beta_n V_n})$$

- **Multi-splitting Feynman-Kac models** ($\neq$ importance sampling)
  $(E = A \cup A^c)$, Y_n Markov, $Y_0 \in A_0 (\subset A)$ $\rightsquigarrow$ $A^c = (B \cup C)$, $C =$ absorbing set/hard obstacle

  Multi-level decomposition $B = B_m \subset \ldots \subset B_1 \subset B_0$ $\quad (A_0 = B_1 - B_0,\ B_0 \cap C = \emptyset)$

$$\Downarrow$$

$$\mathbb{P}(Y_n \text{ hits } B \text{ before } C) = \mathbb{E}(\prod_{1 \leq p \leq m} G_p(X_p))$$

Inter-level excursions : $T_n = \inf \{p \geq T_{n-1} \ : \ Y_p \in B_n \cup C\}$

$$X_n \quad = \quad (Y_p \ ; T_{n-1} \leq p \leq T_n) \in \text{Excursion space} \qquad G_n(X_n) = 1_{B_n}(Y_{T_n})$$

$$\Downarrow$$

FK interpretation

$$\mathbb{E}(f(Y_0, \ldots, Y_{T_m}) \ 1_{B_m}(X_{T_m})) = \mathbb{E}(f(X_0, \ldots, X_m) \prod_{1 \leq p \leq m} G_p(X_p))$$

Advanced signal processing $\longrightarrow$ filtering/hidden Markov chains/Bayesian methodology

Signal process $\quad X_n$ = Markov chain $\in E_n$

Observation/Sensor eq. $\quad Y_n = H_n(X_n, V_n) \in F_n$ with $\mathbb{P}(H_n(x_n, V_n) \in dy_n) = g_n(x_n, y_n)\ \lambda_n(dy_n)$

Example: $Y_n = h_n(X_n) + V_n \in F_n = \mathbb{R}$, with Gaussian noise $V_n = \mathcal{N}(0,1)$

$$\Downarrow$$

$$\mathbb{P}(h_n(x_n) + V_n \in dy_n) = (2\pi)^{-1/2} e^{-\frac{1}{2}(y_n - h_n(x_n))^2}\ dy_n = \underbrace{\exp\left[h_n(x_n)y_n - h_n^2(x_n)/2\right]}_{g_n(x_n,y_n)}\ \underbrace{\mathcal{N}(0,1)(dy_n)}_{\lambda_n(dy_n)}$$

Prediction/filtering/smoothing $\rightarrow$ **Feynman-Kac representation** $G_n(x_n) = g_n(x_n, y_n)$

$$\eta_n = \text{Law}(X_n \mid Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}) = \text{Law}(X'_0, \ldots, X'_n \mid Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1})$$

Partially linear/Gaussian models

$$X_n^1 = \text{Markov} \quad \in E_n \quad + \quad \begin{cases} X_n^2 &= A_n(X_n^1)\ X_{n-1}^2 + a_n(X_n^1) + B_n(X_n^1)\ W_n \in \mathbb{R}^d \\ Y_n &= C_n(X_n^1)\ X_n^2 + c_n(X_n^1) + D_n(X_n^1)\ V_n \in \mathbb{R}^{d'} \end{cases}$$

Given a realization $X^1 = x \to$ *Kalman-Bucy optimal one step predictor*

$$\widehat{X}^{2\ -}_{x,n+1} = \mathbb{E}(X^2_{n+1} \mid Y_0, \ldots, Y_n,\ X^1 = x) \quad \text{and} \quad P^-_{x,n+1} = \mathbb{E}([X^2_{n+1} - \widehat{X}^{2\ -}_{x,n+1}][X^2_{n+1} - \widehat{X}^{2\ -}_{x,n+1}]')$$

$$\Downarrow$$

Quenched Kalman-Bucy recursion: $(\widehat{X}^{2\ -}_{x,n+1}, P^-_{x,n+1}) = \mathcal{B}_{n+1}[(x_n, x_{n+1}), (\widehat{X}^{2\ -}_{x,n}, P^-_{x,n})]$

Feynman-Kac representation: $\eta_n \sim (\mathbf{X}_n, \mathbf{G}_n)$ s.t.

$$\mathbf{X}_n = (X_n^1, (\widehat{X}^{2\,-}_{X^1,n+1}, P^-_{X^1,n+1})) \text{ Markov chain} \in \mathbf{E}_n = (E_n \times \mathbb{R}^d \times \mathbb{R}^{d\times d})$$

$$\mathbf{G}_n(x,m,P) = \frac{d\mathcal{N}(C_n(x)\ m + c_n(x), C_n(x)\ P\ C_n(x)' + D_n(x)R_n^v D_n(x)')}{d\mathcal{N}(0, D_n(x)R_n^v D_n(x)')}(y_n)$$

$$\Downarrow \quad [\textit{virtual sensor :} \quad Y_n = \{C_n(X_n^1)\ \widehat{X}^{2\,-}_{X^1,n} + c_n(X_n^1)\} + \widehat{V}_{X^1,n}\ ]$$

$$F_n(x,m,P) = f_n(x) \quad \Longrightarrow \quad \eta_n(F_n) = \mathbb{E}(f_n(X_n^1) \mid Y_0,\ldots,Y_{n-1})$$
$$F_n(x,m,P) = \mathcal{N}(m,P)(f_n) \quad \Longrightarrow \quad \eta_n(F_n) = \mathbb{E}(f_n(X_n^2) \mid Y_0,\ldots,Y_{n-1})$$

Note: $\rightsquigarrow$ Interacting Kalman-Bucy filters and for path-space models we have

$$X_n^1 = (X_0^{1\,\prime},\ldots,X_n^{1\,\prime}) \rightsquigarrow \text{Law}((X_0^{1\,\prime},\ldots,X_n^{1\,\prime}) \mid Y_0,\ldots,Y_{n-1})$$