Mean field simulation for Monte Carlo integration

Part III : Some application domains

P. Del Moral

INRIA Bordeaux & Inst. Maths. Bordeaux & CMAP Polytechnique

Lectures, INLN CNRS & Nice Sophia Antipolis Univ. 2012

Some hyper-refs

- **Mean field simulation for Monte Carlo integration**. Chapman & Hall - Maths & Stats [600p.] (May 2013).
- **Feynman-Kac formulae**, Genealogical & Interacting Particle Systems with appl., Springer [573p.] (2004)
- **On the concentration of interacting processes**. Foundations & Trends in Machine Learning [170p.] (2012). (joint work with P. Hu & L.M. Wu)
- **More references on the website on Feynman-Kac models and interacting particle systems**: **http://www.math.u-bordeaux1.fr/∼pdelmora/simulinks.html** [+ Links]

Boltzmann-Gibbs measures
- Interacting Markov chain Monte Carlo
- Subset methodology
- Interacting simulated annealing
- Interacting Island models

Rare event analysis
- Tail probabilities
- Level crossing probabilities
- Excursion level crossing

Sensitivity measures
- Parameter derivatives
- Gradient of Markov semigroup

Estimation with partial observation
- Nonlinear filtering
- Hidden Markov chain problems

Particle absorption models
- QMC & DMC algorithms
- Doob h-processes
- Ground states-Quasi-invariant-Yaglom measures

Boltzmann-Gibbs measures

$$\mu_n(dx) = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \le p < n} h_p(x) \right\} \lambda(dx)$$

- Markov chain Monte Carlo moves

$$\mathbb{P}\left(X_n \in dx \;|X_{n-1}\right) = M_n(X_{n-1}, dx) \quad \text{s.t.} \quad \mu_n M_n = \mu_n$$

- Updating transitions

$$\mu_{n+1} = \Psi_{h_n}(\mu_n)$$

$$\Downarrow$$

$$\mu_{n+1} = \mu_{n+1} M_{n+1} = \Psi_{h_n}(\mu_n) M_{n+1}$$

$$\Downarrow$$

$$\mu_n(f_n) \propto \mathbb{E}\left(f_n(X_n) \prod_{0 \le p < n} h_p(X_p) \right) \rightsquigarrow \text{Interacting MCMC}$$

Subset methodology

$$A_n \text{ decreasing sequence of subset levels}$$

$$\Downarrow$$

$$\mu_n(dx) = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0\leq p<n} 1_{A_{p+1}}(x) \right\} \lambda(dx) = \frac{1}{\lambda(A_n)} \, 1_{A_n}(x) \; \lambda(dx)$$

- Markov chain Monte Carlo moves in each set A_n

$$\mathbb{P}\left(X_n \in dx \;|X_{n-1}\right) = M_n(X_{n-1}, dx) \quad \text{s.t.} \quad \mu_n M_n = \mu_n$$

- Updating transitions $\in A_{n+1}$

$$\mu_{n+1} = \Psi_{1_{A_{n+1}}}(\mu_n)$$

Important observation: [counting pb, volume computation, tail event probability]

$$\lambda(A_n) = \lambda(A_0)\; \mathcal{Z}_n = \lambda(A_0) \prod_{0\leq p<n} \mu_p(1_{A_{p+1}})$$

Interacting simulated annealing

$$h_p(x) = \exp\left(-(\beta_{p+1} - \beta_p)V(x)\right)$$

$$\Downarrow \quad (\beta_0 = 0)$$

$$\mu_n(dx) = \frac{1}{\mathcal{Z}_n}\left\{\prod_{0\leq p<n} h_p(x)\right\}\lambda(dx) = \frac{1}{\lambda\left(e^{-\beta_n V}\right)}\ e^{-\beta_n V(x)}\ \lambda(dx)$$

- Simulated annealing style moves at temperature $1/\beta_n$

$$\mathbb{P}\left(X_n \in dx \ |X_{n-1}\right) = M_n(X_{n-1}, dx) \quad \text{s.t.} \quad \mu_n M_n = \mu_n$$

- Updating transitions w.r.t. inverse temperature variations $(\beta_{n+1} - \beta_n)$

$$\mu_{n+1} = \Psi_{e^{-(\beta_{n+1}-\beta_n)V}}(\mu_n)$$

Interacting Island models

$$\xi_{\theta,n} = \text{particle Feynman-Kac model} \sim (M_{\theta,n}, G_{\theta,n}) \quad \text{and} \quad \Theta \sim \nu(d\theta)$$

$$\left.\begin{array}{rcl} x & = & \left(\theta, (\xi_{\theta,n})_{n\in[0,T]}\right) \\ h_n(x) & = & \eta^N_{\theta,n}(G_{\theta,n}) \end{array}\right\} \rightarrow \mu_n(dx) = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0\leq p<n} h_p(x) \right\} \lambda(dx)$$

By the unbiased property *(cf. lecture I)*

$$\mu_n \circ \Theta^{-1} = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0\leq p<n} \eta_{\theta,n}(G_{\theta,n}) \right\} \nu(d\theta)$$

- MCMC shaking moves in (parameter-island)-spaces

$$\mathbb{P}\left(X_n \in dx \;|X_{n-1}\right) = M_n(X_{n-1}, dx) \quad \text{s.t.} \quad \mu_n M_n = \mu_n$$

- Updating w.r.t. the average fitness of the islands $\eta^N_{\theta,n}(G_{\theta,n})$

Rare event analysis

3 type of events:

- red Tail probabilities

$$\mathbb{P}(X \in A) \quad \& \quad \mathrm{Law}(X \mid X \in A) \subset \textbf{Boltzmann-Gibbs models}$$

- Level crossing at a fixed given time

$$\mathbb{P}(V_n(X_n) \geq a) \quad \& \quad \mathrm{Law}((X_0, \ldots, X_n) \mid V_n(X_n) \geq a)$$

- Excursion level crossing

$$\mathbb{P}(X \text{ hits } A_n \text{ before } B) \quad \mathrm{Law}\left((X_n)_{0 \leq n \leq T_{A_n}} \mid X \text{ hits } A_n \text{ before } B \right)$$

Examples:
Networks overload, breakdowns, failures, uncertainty propagation in numerical codes, ruin-default probabilities.

Level crossing probabilities

- Level crossing at a fixed given time

$$\mathbb{P}\left(V_n(X_n) \geq a\right) = \mathbb{E}\left(f_n(X_n)\ e^{V_n(X_n)}\right)$$
$$= \mathbb{E}\left(\mathbf{f_n}(\mathbf{X}_n) \prod_{0 \leq p < n} G_p(\mathbf{X}_p)\right)$$

with

$$\mathbf{X}_n = (X_n, X_{n+1}) \quad \text{and} \quad G_n(\mathbf{X}_n) = e^{V_{n+1}(X_{n+1}) - V_n(X_n)}$$

and the test function

$$f_n(\mathbf{X}_n) = 1_{V_n(X_n) \geq a}\ e^{-V_n(X_n)}$$

Level crossing probabilities

- Level crossing at a fixed given time

$$\mathbb{P}\left(V_n(X_n) \geq a\right) = \mathbb{E}\left(f_n(X_n)\ e^{V_n(X_n)}\right)$$
$$= \mathbb{E}\left(\mathbf{f_n}(\mathbf{X}_n) \prod_{0 \leq p < n} G_p(\mathbf{X}_p)\right)$$

with

$$\mathbf{X}_n = (X_n, X_{n+1}) \quad \text{and} \quad G_n(\mathbf{X}_n) = e^{V_{n+1}(X_{n+1}) - V_n(X_n)}$$

and the test function

$$f_n(\mathbf{X}_n) = 1_{V_n(X_n) \geq a}\ e^{-V_n(X_n)}$$

Level crossing probabilities

- Level crossing at a fixed given time (continued)

$$\mathbf{X}_n = (X_n, X_{n+1}) \quad \text{and} \quad G_n(\mathbf{X}_n) = e^{V_{n+1}(X_{n+1}) - V_n(X_n)}$$

Conditional distributions w.r.t. the critical event

$$\begin{aligned}
&\mathbb{E}\left(\varphi_n(X_0, \ldots, X_n) \mid V_n(X_n) \geq a\right) \\
&= \mathbb{E}\left(F_{n,\varphi_n}(X_0, \ldots, X_n)\ e^{V_n(X_n)}\right) / \mathbb{E}\left(F_{n,1}(X_0, \ldots, X_n)\ e^{V_n(X_n)}\right) \\
&= \mathbb{Q}_n(F_{n,\varphi_n}) / \mathbb{Q}_n(F_{n,1})
\end{aligned}$$

with the function

$$F_{n,\varphi_n}(X_0, \ldots, X_n) = \varphi_n(X_0, \ldots, X_n)\ 1_{V_n(X_n) \geq a}\ e^{-V_n(X_n)}$$

Excursion level crossing

Multilevel decomposition $A_n \downarrow$, with B non critical recurrent subset.

$$\mathbb{P}(X \text{ hits } A_n \text{ before } B) \quad = \quad \mathbb{E}\left(\prod_{0\leq p\leq n} 1_{A_p}(X_{T_p})\right)$$

$$T_n := \inf\{p \geq T_{n-1} \ : \ X_p \in (A_n \cup B)\}$$

Feynman-Kac model in excursion spaces

$$\mathbb{E}\left(\prod_{0\leq p\leq n} 1_{A_p}(X_{T_p})\right) = \mathbb{E}\left(\prod_{0\leq p<n} G_p(\mathbf{X}_p)\right)$$

with

$$\mathbf{X}_n = (X_p)_{p\in[T_n, T_{n+1}]} \quad \& \quad G_n(\mathbf{X}_n) = 1_{A_{n+1}}(X_{T_{n+1}})$$

Excursion level crossing

Multilevel decomposition $A_n \downarrow$, with B non critical recurrent subset.

$$\mathbb{P}(X \text{ hits } A_n \text{ before } B) \quad = \quad \mathbb{E}\left(\prod_{0 \leq p \leq n} 1_{A_p}(X_{T_p})\right)$$

$$\mathbf{X}_n = (X_p)_{p \in [T_n, T_{n+1}]} \quad \& \quad G_n(\mathbf{X}_n) = 1_{A_{n+1}}(X_{T_{n+1}})$$

$$\Downarrow$$

Conditional distributions

$$\mathbb{E}\left(\mathbf{f_n}(X_{[0,T_{n+1}]}) \mid X \text{ hits } A_n \text{ before } B\right) = \mathbb{Q}_n(\mathbf{f}_n)$$

Parameter derivatives

Derivation FK models = FK Integration w.r.t. additive functionals

Computation of Feynman-Kac semigroup derivatives w.r.t. some parameter $\theta \in \mathbb{R}^d$ (here $d = 1$)

$$Q_n^\theta(x_{n-1}, dx_n) = H_n^\theta(x_{n-1}, x_n)\ \nu_n(dx_n)$$

$$\Rightarrow\ \frac{\partial}{\partial\theta}\,\mathbb{E}\left(f_n(X_0^\theta, \ldots, X_n^\theta)\ \prod_{0\leq p<n} G_p^\theta(X_p^\theta)\right) \propto \mathbb{Q}_n^\theta(f_n \times \Gamma_n^\theta)$$

with the additive functional

$$\Gamma_n^\theta(x_0, \ldots, x_n) = \sum_{1\leq p\leq n} \frac{\partial}{\partial\theta} \log H_p^\theta(x_{p-1}, x_p)$$

Parameter derivatives (continued)

Derivation FK models = FK Integration w.r.t. additive functionals

Example $d = 1$ & $(G_n^\theta, M_n^\theta) = (G_n^\theta, M_n)$

$$\frac{\partial}{\partial\theta} \log \mathcal{Z}_n^\theta = \mathbb{Q}_n^\theta(\Gamma_n^\theta) \quad \& \quad \frac{\partial}{\partial\theta} \mathbb{Q}_n^\theta(f_n) = \mathbb{Q}_n^\theta\left(f_n \ \left[\Gamma_n^\theta - \mathbb{Q}_n^\theta(\Gamma_n^\theta)\right]\right)$$

with the additive functional

$$\Gamma_n^\theta(x_0, \ldots, x_n) = \sum_{0 \le p < n} \frac{\partial}{\partial\theta} \log G_p^\theta(x_p)$$

Markov semigroup derivation (here $d = 1$)

$$P_n(f)(x) := \mathbb{E}\left(f(X_n(x))\right)$$

with

$$X_{n+1}(x) = F_n(X_n(x), W_n) \qquad X_0(x) = x$$

$$\Downarrow$$

1st order variational equation

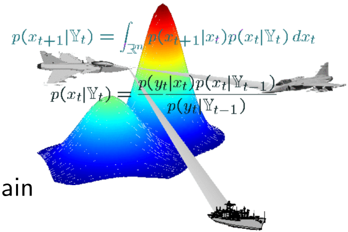
$$\frac{\partial X_n}{\partial x} = \frac{\partial F_{n-1}}{\partial x}(X_{n-1}, W_{n-1})\,\frac{\partial X_{n-1}}{\partial x} = \prod_{0 \leq p < n} G_p(X_p, W_p)$$

FK-representation formula

$$\nabla P_n(f)(x) := \mathbb{E}_x\left(\nabla f(X_n) \prod_{0 \leq p < n} G_p(X_p, W_p)\right)$$

$d > 1$ Noncommutative models $\rightsquigarrow$ sequential proj. on the unit sphere

Nonlinear filtering and smoothing



Signal-Observation model: $(X, Y) =$ Markov chain

$$\mathbb{P}\left((X_n, Y_n) \in d(x,y) | (X_{n-1}, Y_{n-1})\right) := M_n(X_{n-1}, dx)\; g_n(x,y)\; \nu_n(dy)$$

- Conditional distributions: fixed obs. $Y = y$, $G_n(x_n) \propto g_n(x_n, y_n)$

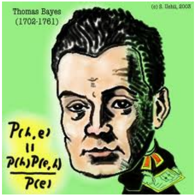
$$\mathbb{Q}_n = \mathrm{Law}((X_0, \ldots, X_n) \mid \forall 0 \leq p < n \;\; Y_p = y_p)$$

- Normalizing constants:

$$\mathcal{Z}_{n+1} \propto p_n(y_0, \ldots, y_n)$$

Hidden Markov chain problems

Pb : $\theta \mapsto (X^\theta, Y^\theta) \rightsquigarrow \mathrm{Arg\text{-}max}_\theta\, p_n(y_0, \ldots, y_n \mid \theta)$



Bayesian formulation:

$$dp(\theta \mid (y_0, \ldots, y_n)) = \frac{1}{\mathcal{Z}_{n+1}}\ \mathcal{Z}_{n+1}(\theta)\ dp(\theta)$$

with

$$\begin{aligned}
\mathcal{Z}_{n+1}(\theta) &= p_n(y_0, \ldots, y_n \mid \theta) \\
&= \mathbb{E}\left(\prod_{0 \le p \le n} g_{p,\theta}(X_p^\theta, y_p)\right) \\
&= \prod_{0 \le p \le n} \eta_{\theta,p}(g_{p,\theta}) = \prod_{0 \le p \le n} h_p(\theta) \qquad \subset \textbf{Boltzmann-Gibbs models}
\end{aligned}$$

NB.: Conditionally linear-Gaussian $\eta_{\theta,p}(g_{p,\theta})$ or Island-models $\eta^N_{\theta,p}(g_{p,\theta})$

Hidden Markov chains (continued)

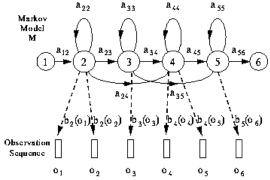


Fig. 1.3 The Markov Generation Model

$\text{Pb} : \theta \mapsto (X^\theta, Y^\theta) \rightsquigarrow \text{Arg-max}_\theta\, p_n(y_0, \ldots, y_n \mid \theta)$

Frequentists formulation: EM style algo. $\oplus$ Stochastic gradient

$$\theta_n = \theta_{n-1} + \tau_n \, \nabla \log \mathcal{Z}_n^{\theta_{n-1}}$$

Note:

Derivation FK models = FK Integration w.r.t. additive functionals

$$\frac{\partial}{\partial\theta} \log \mathcal{Z}_n^\theta = \mathbb{Q}_n^\theta(\Gamma_n^\theta)$$

Optimal stopping & Snell envelop

Optimal stopping time problems

$$\sup_{T \leq n} \mathbb{E}\left(f_T(X_T) \prod_{0 \leq p < T} G_p(X_p) \right)$$

Solution:

$$T^\star = \inf \{0 \leq p \leq n \ : \ U_p(X_p) \leq f_p(X_p)\}$$

with the Snell envelop backward equation $(U_n = f_n)$:

$$\begin{aligned} U_p(x) &= f_p(x) \vee \int Q_{p+1}(x, dy)\ U_{p+1}(y) \\ &= f_p(x) \vee \int \eta_{p+1}(dy)\ \frac{dQ_{p+1}(x, .)}{d\eta_{p+1}}(y)\ U_{p+1}(y) \\ &\simeq f_p(x) \vee \eta_p^N(G_p)\ \int \eta_{p+1}^N(dy)\ \frac{H_{p+1}(x, y)}{\eta_p^N(H_{p+1}(., y))}\ U_{p+1}(y) \end{aligned}$$

Boltzmann-Gibbs measures

Rare event analysis

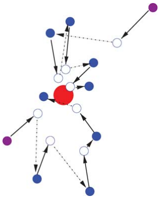
Sensitivity measures

Estimation with partial observation

Particle absorption models

- QMC & DMC algorithms
- Doob h-processes
- Ground states-Quasi-invariant-Yaglom measures

Absorption models $\rightsquigarrow$ QMC+DMC algo.



- Sub Markov operators

$$Q_n(x,dy) = G_{n-1}(x)\ M_n(x,dy) \rightsquigarrow E_n^c = E_n \cup \{c\}$$

- Absorbed Markov chain model

$$X_n^c \in E_n^c \xrightarrow{\text{absorption } \sim (1-G_n)} \widehat{X}_n^c \xrightarrow{\text{exploration } \sim M_{n+1}} X_{n+1}^c$$

$$\Downarrow$$

$$\mathbb{Q}_n = \mathrm{Loi}((X_0^c, \ldots, X_n^c) \mid T^{absorption} \geq n)$$

et

$$\mathcal{Z}_n = \mathrm{Proba}\left(T^{absorption} \geq n\right)$$

Doob h-processes $(G_n, M_n) = (G, M)$

- Reversibility condition : $\mu(dx)M(x, dy) = \mu(dy)M(y, dx)$

$$\mathrm{Proba}\left(T^{absorption} \geq n\right) \simeq \lambda^n$$

with $\lambda =$ top of the spectrum of

$$Q(x, dy) = G(x)\ M(x, dy)$$

- $Q(h) = \lambda h \rightsquigarrow$ Doob h-process X^h

$$M^h(x, dy) = \frac{1}{\lambda} h^{-1}(x) Q(x, dy) h(y) = \frac{Q(x, dy)h(y)}{Q(h)(x)} = \frac{M(x, dy)h(y)}{M(h)(x)}$$

Doob h-processes : FK formulation

$$\mathbb{Q}_n(d(x_0,\ldots,x_n)) \propto \mathrm{Proba}((X_0^h,\ldots,X_n^h) \in d(x_0,\ldots,x_n))\ h^{-1}(x_n)$$

$$\Downarrow$$

Correspondance formulae with the invariant measure $\mu_h = \mu_h M^h$

$$\eta_n = \Psi_{1/h}(\eta_n^h) \quad \& \quad \eta_n^h = \Psi_h(\eta_n)$$

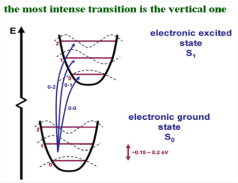
and

$$\eta_\infty = \Psi_{M(h)}(\mu) \quad \& \quad \eta_\infty^h := \Psi_h(\eta_\infty) = \Psi_{hM(h)}(\mu)$$

Pysical interpretation $G = e^{-V}$:

$$\widehat{\eta}_\infty = \Psi_G(\eta_\infty) = \Psi_h(\mu) \propto h\ d\mu$$

$\Rightarrow h =$ground state energy



Mathematical biology: $\eta_\infty =$ Quasi-inv. or Yaglom measure

Doob h-processes : Particle approximations

- Updated limiting population distribution:

$$\Psi_G(\eta_\infty) \simeq_N \Psi_G(\eta_\infty) = \Psi_h(\mu) \propto h\ d\mu$$

- Particle invariant measure approx.

$$\mu_h(f) \simeq_n \mathbb{Q}_n(\overline{F}_n) \simeq_N \mathbb{Q}_n^N(\overline{F}_n) \quad \text{with} \quad \overline{F}_n(\mathbf{x_n}) = \frac{1}{n+1} \sum_{0 \le p \le n} f(x_p)$$

- For $G = G^\theta$ related to $\theta \in \mathbb{R} \rightsquigarrow f := \frac{\partial}{\partial\theta} \log G^\theta$

$$\frac{\partial}{\partial\theta} \log \lambda^\theta \simeq_n \frac{1}{n+1} \frac{\partial}{\partial\theta} \log \mathcal{Z}_{n+1}^\theta \simeq_N \mathbb{Q}_n^N(\overline{F}_n)$$