Sequential Monte Carlo and Applications in Molecular Dynamics

Qiming Du

Laboratoire de Probabilités, Statistique et Modélisation - UMR 8001
Sorbonne Université

Directeurs de thèse : Arnaud Guyader et Tony Lelièvre

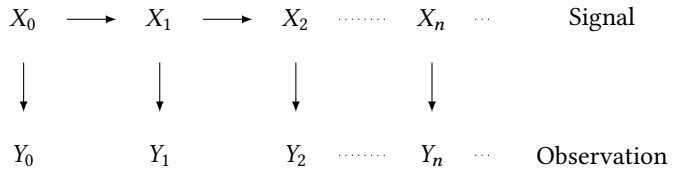
Mercredi 17 juin 2020 Paris

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

LPSM
LABORATOIRE DE PROBABILITÉS STATISTIQUE & MODÉLISATION

Table of Contents

Hidden Markov Models



- $(X_n)_{n\geq 0}$: the underlying Markov chain that is not directly observable;
- $(Y_n)_{n\geq 0}$: the conditionally independent observations, i.e.,

$$\mathbf{P}\left(Y_n \in A \mid X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\right) = \mathbf{P}\left(Y_n \in A \mid X_n = x_n\right)$$

Goal: Conduct inference on the posterior distributions defined as follows:

$$I_n(f) = \mathbf{E}\left[f(X_n) \mid Y_0, \cdots, Y_{n-1}\right]$$

For example, one may consider the setting

$$\begin{cases} X_n = g(X_{n-1}) + W_n \\ Y_n = h(X_n) + V_n \end{cases}$$

Hidden Markov Models: Particle Filters

Bootstrap Filter [Gordon, 1993]:

- Generate an N-particle system;
- Evolve the particle system with *selection* and *mutation* mechanisms.

At each iteration $j \geq 0$:

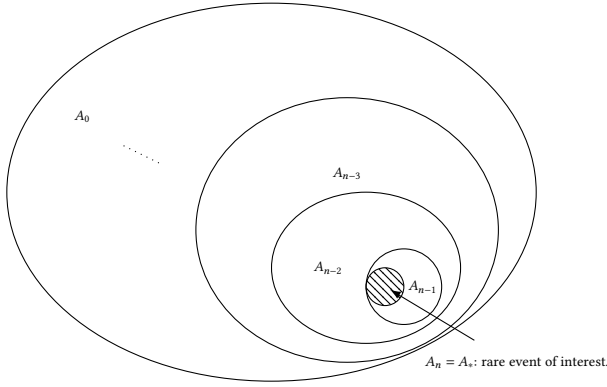
- Selection: Multinomial resampling according to $p\left(Y_j \mid x_j\right)$;
- Mutation: Transition according to the Markov kernel $p\left(x_{j+1} \mid x_j\right)$.

Estimation: use the empirical measure at level n to approximate posterior distribution.

Key idea: Bayes formula + conditional independence yield

$$I_n(f) = \frac{\int f(x_n) \prod_{j=0}^{n-1} p\left(Y_j \mid x_j\right) p(x_0) \prod_{j=0}^{n-1} p\left(x_{j+1} \mid x_j\right) dx_0 dx_1 \ldots dx_n}{\int \underbrace{\prod_{j=0}^{n-1} p(Y_j \mid x_j)}_{\text{Weights (selection)}} \underbrace{p(x_0) \prod_{j=0}^{n-1} p(x_{j+1} \mid x_j) dx_0 dx_1 \ldots dx_n}_{\text{Joint distribution of Markov chain (mutation)}}}$$

Rare-event simulation: Subset Simulation



Goal: Estimate the rare-event probability $p_* := \mathbf{P}(X \in A_*)$.
Key idea: Construct a sequence of decreasing sets $(A_j)_{0 \le j \le n}$, such that

$$E = A_{-1} \supset A_0 \supset A_1 \supset A_2 \supset \cdots \supset A_{n-1} \supset A_n = A_*$$

Bayes formula gives $\mathbf{P}(X \in A_*) = \prod_{j=0}^{n} \mathbf{P}\left(X \in A_j \,\middle|\, X \in A_{j-1}\right)$

Rare-event simulation: Subset Simulation

Subset simulation [Au and Beck, 2001]:

- Construct $\mathbf{P}\left(\cdot \,\middle|\, X \in A_j\right)$-invariant kernels (shaker in A_j);
- Evolve the particle system with selection-mutation mechanism.

At iteration $j \geq 0$:

- Select the surviving particles that succeed in moving to the smaller set A_j;
- Uniformly resample N replicas from the surviving ones;
- Mutate all the N particles according to the shaker in A_j.

Estimation: product of the rates of surviving particles at each step.

$$\underbrace{\mathbf{P}\left(X \in A_*\right)}_{\text{Rare-event probability}} = \prod_{j=0}^{n} \underbrace{\mathbf{P}\left(X \in A_j \,\middle|\, X \in A_{j-1}\right)}_{\text{Non-rare-event probability}}$$

Each term of the "non-rare-event probability" can therefore be estimated separately.

General framework: Feynman-Kac particle models

Mathematical formalization for a wide range of genetic particle methods.
Main reference: the pair of books [Del Moral, 2004, Del Moral, 2013].

Ingredients:

- A sequence of non-negative potential functions $(G_n)_{n\geq 0}$;
- A sequence of Markov kernels $(M_n)_{n\geq 1}$ and an initial distribution η_0.

Key constructions:

- Feynman-Kac operator (semigroup) $Q_n(x, dy) := G_{n-1}(x)M_n(x, dy)$;
- Feynman-Kac measures flows $(\eta_n)_{n\geq 0}$ and $(\gamma_n)_{n\geq 0}$ defined by

$$\gamma_n = \gamma_{n-1}Q_n \quad \text{and} \quad \eta_n = \eta_{n-1}Q_n/\eta_{n-1}(G_{n-1})$$

Relation between η_n and γ_n:

$$\gamma_n = \left\{\prod_{p=0}^{n-1} \eta_p(G_p)\right\} \eta_n \quad \text{and} \quad \eta_n = \gamma_n / \underbrace{\gamma_n(1)}_{\text{mass of } \gamma_n}$$

With proper choices of G_n and M_n:

- In Bootstrap Filter, the posterior distribution $I_n(f)$ is $\eta_n(f)$;
- In Subset Simulation, the rare-event probability p_* is $\gamma_n(1)$.

Interacting Particle System (IPS)

McKean interpretation in Feynman-Kac particle models:

- Construct McKean kernels that connect the probability measures $(\eta_n)_{n\geq 0}$:

$$\eta_{n-1}K_{n,\eta_{n-1}} = \eta_n$$

- Simulate an IPS to approximate the constructed McKean chain, and estimate η_n and γ_n with the associated empirical measures η_n^N, i.e.,

$$\eta_n^N := \frac{1}{N}\sum_{i=1}^{N}\delta_{X_n^i} \quad \text{and} \quad \gamma_n^N := \left\{\prod_{p=0}^{n-1}\eta_p^N(G_p)\right\}\eta_n^N$$

- Mechanism of IPS (replacing η_{p-1} by η_{p-1}^N):
  1. $\mathbf{X_0} = (X_0^1, X_0^2, \ldots, X_0^N) \sim \eta_0^{\otimes N}$;
  2. $\forall p \geq 1, \quad \mathbf{X_p} = (X_p^1, X_p^2, \ldots, X_p^N) \sim \bigotimes_{i=1}^{N} K_{p,\eta_{p-1}^N}(X_{p-1}^i, \cdot)$.

Multinomial resampling scheme:

$$K_{n,\eta}(x, dy) := \frac{\eta Q_n(dy)}{\eta(G_{n-1})}, \quad \text{with } Q_n(x, dy) = G_{n-1}(x)M_n(x, dy).$$

Interacting Particle System (IPS)

Genealogy of IPS:
In practice, the approximated McKean kernel

$$K_{n,\eta_{n-1}^N}(x, dy) = \sum_{k=1}^{N} \frac{G_{n-1}(X_{n-1}^k)}{\sum_{\ell=1}^{N} G_{n-1}(X_{n-1}^\ell)} M_n(X_{n-1}^k, dy)$$

can be decomposed into two steps:

- Selection:

$$\forall 1 \leq i \leq N, \quad A_{n-1}^i \sim \sum_{k=1}^{N} \frac{G_{n-1}(X_{n-1}^k)}{\sum_{\ell=1}^{N} G_{n-1}(X_{n-1}^\ell)} \delta_k$$

- Mutation:

$$\forall 1 \leq i \leq N, \quad X_n^i \sim M_n(X_{n-1}^{A_{n-1}^i}, \cdot)$$

This mechanism applies to each particle at each iteration. We call the indices of the parent $(A_n^i)_{n \geq 0, 1 \leq i \leq N}$ the *genealogy* of the IPS.

Some convergence results ($N \to \infty$)

From now on, we only discuss the behavior of γ_n^N. Suppose that $(G_n)_{n\geq 0}$ are uniformly bounded:

- **Central Limit Theorem:**

$$\sqrt{N}(\gamma_n^N(f) - \gamma_n(f)) \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \sigma_{\gamma_n}^2(f)\right)$$

with

$$\sigma_{\gamma_n}^2(f) := \sum_{p=0}^{n} \left\{\gamma_p(1)\gamma_p(Q_{p,n}(f)^2) - \gamma_n(f)^2\right\}$$

where $Q_{p,n} := Q_{p+1}Q_{p+2}\cdots Q_n$.

- **Consistent variance estimator** [Chan and Lai, 2013, Lee and Whiteley, 2018]:

$$N\gamma_n^N(1)^2 V_n^N(f) \xrightarrow[N\to\infty]{\mathbf{P}} \sigma_{\gamma_n}^2(f)$$

with, if E_n^i denotes the index of the ancestor of the particle X_n^i at step 0,

$$V_n^N(f) := \eta_n^N(f)^2 - \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{E_n^i \neq E_n^j} f(X_n^i) f(X_n^j)$$

Table of Contents

Adaptive SMC via summary statistics [Beskos et al., 2016]

The potential functions $G_{n,z}$ and Markov kernels $M_{n,z}$ are indexed by a parameter $z \in \mathbf{R}^d$.

Optimal Feynman-Kac particle model:
Let $z_n^* = \eta_n(\zeta_n)$ be the optimal parameter for the considered application (e.g. the smallest asymptotic variance). However, z_n^* is unknown.

Idea: Use the information encoded in the history of the samples, through the design of an adaptive algorithm by setting $\hat{z}_n^* = \eta_n^N(\zeta_n)$.

McKean kernel:

$$K_{n,\eta}(x, dy) := \frac{\eta Q_{n,\eta(\zeta_{n-1})}(dy)}{\eta(G_{n-1,\eta(\zeta_{n-1})})},$$

with the adaptive Feynman-Kac operator defined by

$$Q_{n,\eta(\zeta_{n-1})}(x, dy) = G_{n-1,\eta(\zeta_{n-1})}(x) M_{n,\eta(\zeta_{n-1})}(x, dy).$$

Adaptive SMC via summary statistics

Main Assumption:
For any test function f, there exists a function $h_n : E_n \times \mathbf{R}^d \longrightarrow \mathbf{R}^d$ such that

$$\forall (x, z) \in E_n \times \mathbf{R}^d, \quad Q_{n+1,z}(f)(x) - Q_{n+1,z_n^*}(f)(x) = \left\langle h_n(x, z), z - z_n^* \right\rangle$$

The function h_n is assumed to satisfy some regularity properties such as boundedness and smoothness, and a specific "stability" property, that writes

$$\eta_n \left(h_n(\cdot, z_n^*) \right) = 0$$

By denoting $Q_n = Q_{n,z_{n-1}^*}$, we have the following asymptotic result:
CLT [Beskos et al., 2016]: Under the Assumption above (and some other regularity assumptions), we have

$$\sqrt{N} \left(\gamma_n^N(f) - \gamma_n(f) \right) \xrightarrow[N \to \infty]{\mathrm{d}} \mathcal{N} \left(0, \sigma_{\gamma_n}^2(f) \right)$$

The asymptotic variance is *identical* to the one of the optimal reference model.
Question: Is the variance estimator of L&W still consistent in this adaptive setting?

Answer: YES!
However, the theoretical tools in [Lee and Whiteley, 2018] are not easy to use in the adaptive setting to prove the consistency, so a new language is required to conduct the analysis.

Key idea: Separate *adaptiveness* and *genealogy* by Doob's decomposition.

Asymptotic variance [Cérou et al., 2011]:

$$\sigma^2_{\gamma_n}(f) = \sum_{p=0}^{n} \left(\Gamma_n^{(p)}(f^{\otimes 2}) - \Gamma_n^{(\varnothing)}(f^{\otimes 2}) \right).$$

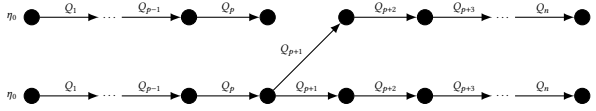


Figure: A representation of the coalescent tree-based measure $\Gamma_n^{(p)} := \gamma_p(1)\gamma_p(Q_{p,n}(f)^2)$.

Term by term estimators:

$$\sigma^2_{\gamma_{n,N}}(f) := \sum_{p=0}^{n} \left(\Gamma^{(p)}_{n,N}(f^{\otimes 2}) - \Gamma^{(\varnothing)}_{n,N}(f^{\otimes 2}) \right) \xrightarrow[N\to\infty]{\mathbf{P}} \sigma^2_{\gamma_n}(f).$$

Key construction: Coalescent tree occupation measures

$b = (b_0, b_1, \ldots, b_n) \in \{0,1\}^{n+1}$: coalescence indicator.
$b_p = 1$ indicates coalescence at step p.
If there exists only *one* coalescence at level p, we denote $b = (p)$ (resp. $(\varnothing)$ if none).

$$\Gamma^b_{n,N}(F) := \gamma^N_n(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{\ell^{[2]}_{0:n} \in ((N)^2)^{\times(n+1)}} \left\{ \prod_{p=0}^{n-1} \lambda^b_p(A^{\ell^{[2]}_{p+1}}_p, \ell^{[2]}_p) \right\} C_{b_n}(F)(X^{\ell^{[2]}_n}_n),$$

where $\lambda^b_p(\tilde{a}^{[2]}_p, \ell^{[2]}_p) \in \{0,1\}$ is an indicator function defined by

$$\lambda^b_p(\tilde{a}^{[2]}_p, \ell^{[2]}_p) := \mathbf{1}_{\{b_p=0\}} \mathbf{1}_{\{\tilde{a}^1_p = \ell^1_p \neq \tilde{a}^2_p = \ell^2_p\}} + \mathbf{1}_{\{b_p=1\}} \mathbf{1}_{\{\tilde{a}^1_p = \ell^1_p = \tilde{a}^2_p \neq \ell^2_p\}}.$$

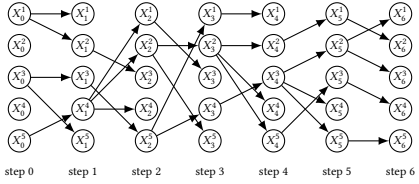
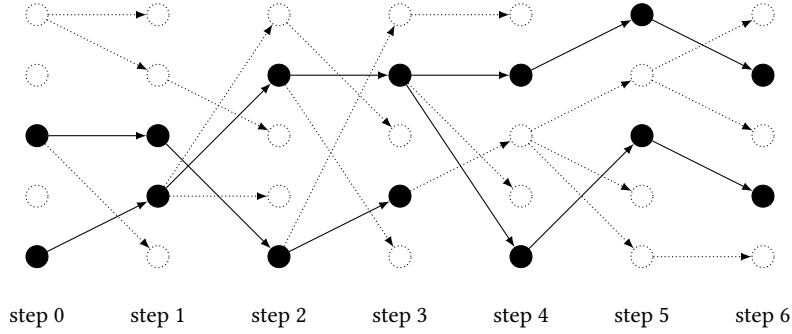


Figure: An IPS (with genealogy) of $n+1 = 7$ levels and $N = 5$ particles at each level.

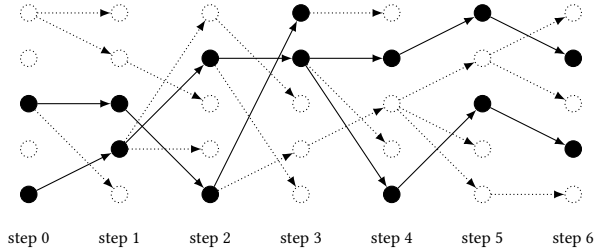
Estimate $\Gamma_6^{(3)}(F)$ by $\Gamma_{6,5}^{(3)}(F)$



It turns out that there are 4 possible choices, taking into account that $F(x, x')$ is not necessarily symmetric in its variables. Namely, the first couple of ancestral lines is:

- $\ell_{0:6}^{[2]} = ((5,3),(4,3),(2,5),(2,4),(2,5),(1,3),(2,4))$,
- $\ell_{0:6}^{[2]} = ((5,3),(4,3),(2,5),(2,4),(5,2),(3,1),(4,2))$.

Estimate $\Gamma_6^{(3)}(F)$ by $\Gamma_{6,5}^{(3)}(F)$



The second couple of ancestral lines is:

- $\ell_{0:6}^{[2]} = ((5,3),(4,3),(2,5),(2,1),(2,5),(1,3),(2,4))$,
- $\ell_{0:6}^{[2]} = ((5,3),(4,3),(2,5),(2,1),(5,2),(3,1),(4,2))$.

Hence, the number of choices of $\ell_{0:6}^{[2]}$ where $\ell_6^{[2]} = (2,4)$ is 2, and the number of choices of $\ell_{0:6}^{[2]}$ where $\ell_6^{[2]} = (4,2)$ is also 2.

$$\Gamma_{6,5}^{(3)}(F) = \gamma_6^5(1)^2 \times 2 \times \left\{ \frac{5^5}{4^7} \left(F(X_6^2, X_6^4) + F(X_6^4, X_6^2) \right) \right\}.$$

Strategy of proof

Step 1: Prove the consistency of coalescent tree occupation measures:

$$\Gamma^b_{n,N}(F) \xrightarrow[N\to\infty]{\mathbf{P}} \Gamma^b_n(F),$$

which proves the consistency of the term by term estimator $\sigma^2_{\gamma_{n,N}}(f)$.
Step 2: Using some combinatorial structure of IPS, prove that the variance estimator of L&W is "close" to our term by term estimator:

$$N\gamma^N_n(1)^2 V^N_n(f) - \sigma^2_{\gamma_{n,N}}(f) \xrightarrow[N\to\infty]{\mathbf{P}} 0.$$

Theorem

Under proper assumptions, the estimator of L&W is consistent in the Adaptive SMC framework:

$$N\gamma^N_n(1)^2 V^N_n(f) \xrightarrow[N\to\infty]{\mathbf{P}} \sigma^2_{\gamma_n}(f).$$

Remarks on our assumptions:

- Our assumptions to have CLT + consistent asymptotic variance estimation are slightly weaker than the ones proposed in [Beskos et al., 2016], which already cover a wide range of real-world applications.
- Our assumptions do not cover the model such as static AMS algorithm proposed in [Cérou and Guyader, 2016]. However, we believe that with a relaxation on the regularity of h_n, we could still prove the consistency of the estimator of L&W.
- The general strategy to separate the adaptiveness and genealogy analysis should also apply to models with *different resampling schemes*.

Different resampling schemes? Which? Why?

Table of Contents

Alternative McKean interpretation

Subset Simulation as Feynman-Kac particle model:

- $G_n(x)$: an indicator function $\mathbf{1}_{A_n}(x)$;
- $M_n(x, dy)$: an η_n-invariant kernel, with $\eta_n := \mathbf{P}(X \in \cdot \mid X \in A_n)$.

Symmetric McKean kernel:

$$K_{n,\eta}^{\mathrm{sym}}(x, dy) := G_{n-1}(x) M_n(x, dy) + (1 - G_{n-1}(x)) \frac{\eta Q_n(dy)}{\eta(G_{n-1})}.$$

Asymmetric McKean kernel:

$$K_{n,\eta}^{\mathrm{asym}}(x, dy) := G_{n-1}(x) \delta_x(dy) + (1 - G_{n-1}(x)) \frac{\eta Q_n(dy)}{\eta(G_{n-1})}.$$

It is readily checked that $\eta_{n-1} K_{n,\eta_{n-1}}^{\mathrm{sym}} = \eta_n$ and $\eta_{n-1} K_{n,\eta_{n-1}}^{\mathrm{asym}} = \eta_n$.

Idea of asymmetric resampling: Since the surviving particles are already "well placed", there is no need to implement another mutation. So, the computational burden can be dramatically reduced, since the mutation kernel M_n is usually the main cost in practice.

Rare-event simulation in Molecular Dynamics

Overdamped Langevin dynamics:

$$\forall t \in \mathbf{N},\ X_{t+1} - X_t = -\nabla V(X_t)h + \sqrt{2\beta^{-1}}(W_{(t+1)h} - W_{th}).$$

with potential function V and inverse temperature β. The process $(W_s)_{s\geq 0}$ denotes a standard Brownian motion in $\mathbf{R}^d$ and $h > 0$ is the associated time step.

Metastability:
Metastable states are typically the wells (open sets) in the potential function V. When trapped in such zone, it is difficult for $(X_t)_{t\in\mathbf{N}}$ to escape.

Goal: Denote by A and B two metastable states. The goal is to estimate the rare-event probability

$$p_* := \mathbf{P}_{x_0}(\tau_B < \tau_A) = \mathbf{P}(\tau_B < \tau_A \mid X_0 = x_0),$$

where the stopping times are defined respectively by

$$\tau_A := \inf\{t \in \mathbf{N} : X_t \in A\} \quad \text{and} \quad \tau_B := \inf\{t \in \mathbf{N} : X_t \in B\}.$$

Rare-event simulation in Molecular Dynamics

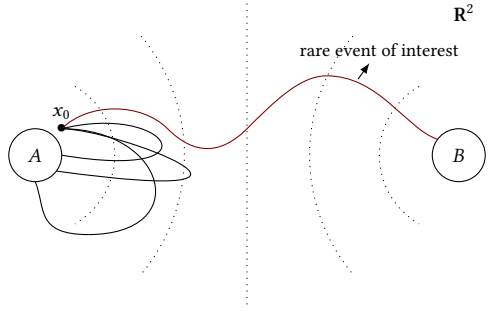


Figure: Schematic picture of a 2-dimensional toy example with the level sets of V (dotted lines).

Difficulties:

- High dimension: finite difference-based numerical PDE methods would not work;
- Metastability: rare-event simulation involved, and naive MC would not work.

Question: Can we implement SMC in this scenario? How?

Generalized Adaptive Multilevel Splitting (gAMS) [Bréhier et al., 2016]

An Adaptive SMC-type algorithm that does not enter into the Adaptive SMC framework discussed above (different resampling scheme; "stability" not verified).

Ingredient:
A *reaction coordinate* $\xi : \mathbf{R}^d \mapsto \mathbf{R}$, that measures the advance of a reaction trajectory towards metastable state B.

Level of trajectory:
For a trajectory $\boldsymbol{x} = (x_t)_{t \geq 0}$, the level associated to the reaction coordinate ξ is the maximum value of ξ on $\boldsymbol{x}$, i.e.,

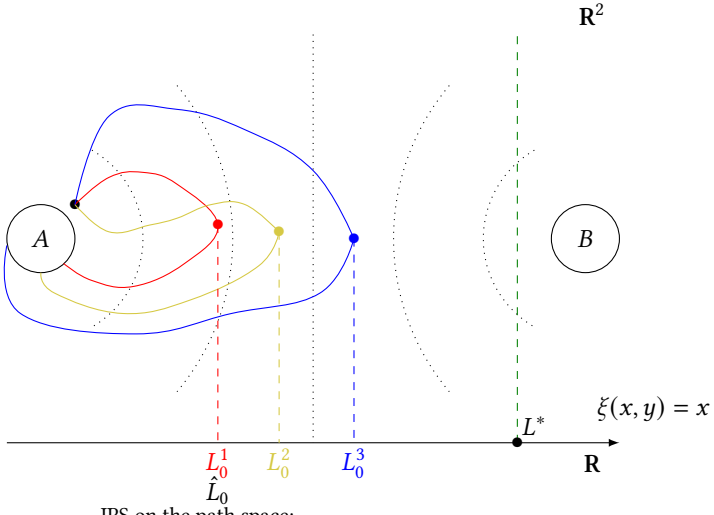
$$\sup_{t < \tau_A \wedge \tau_B} \{\xi(x_t)\}.$$

Idea of gAMS:

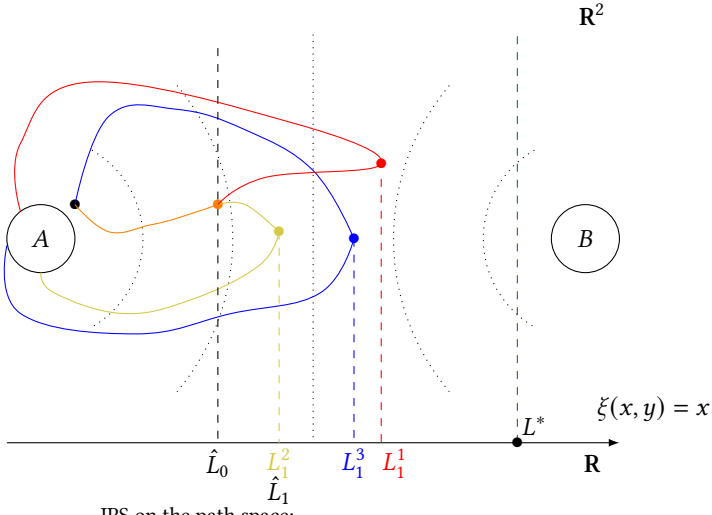
- Generate an N-particle system in the path space;
- Implement an adaptive level generating mechanism;
- Evolve the IPS with asymmetric resampling scheme.

gAMS

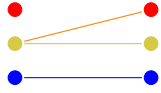
with *continous* reaction coordinate

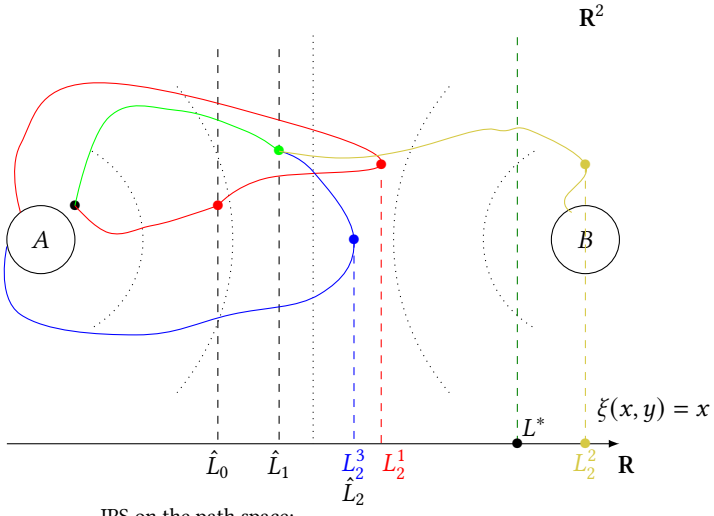


IPS on the path space:



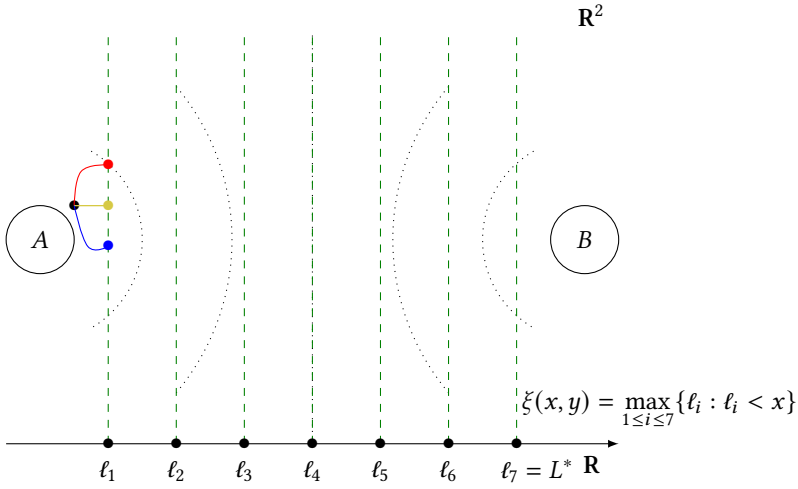
IPS on the path space:



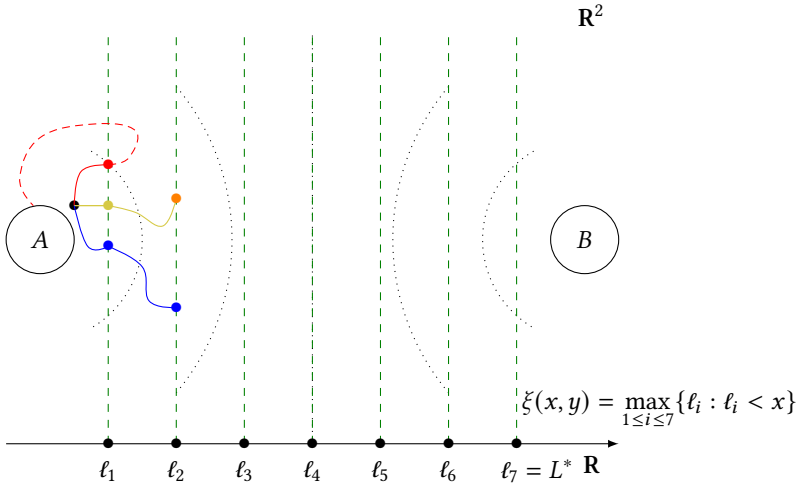

$\mathbf{R}^2$
A
B
$\xi(x,y) = x$
L^*
$\hat{L}_0$
$\hat{L}_1$
L_2^3
$\hat{L}_2$
L_2^1
L_2^2
$\mathbf{R}$
IPS on the path space:
...

gAMS

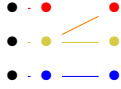
with *discrete* reaction coordinate

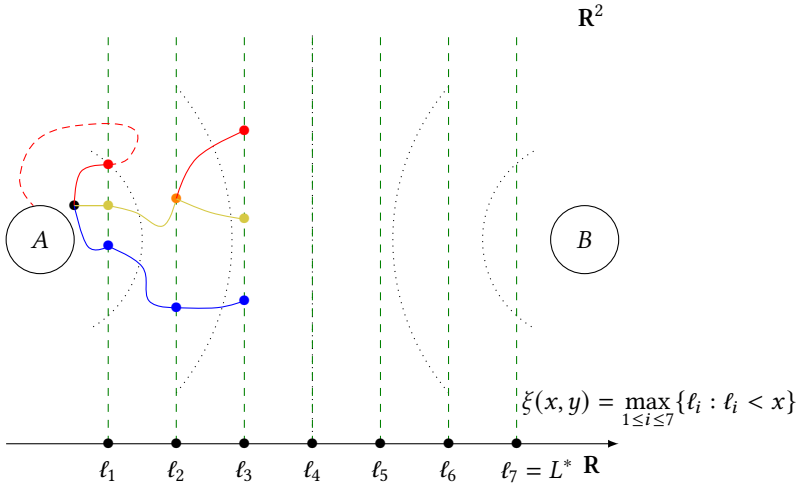


IPS on the state space (Symmetric SMC):

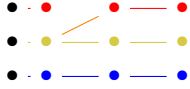


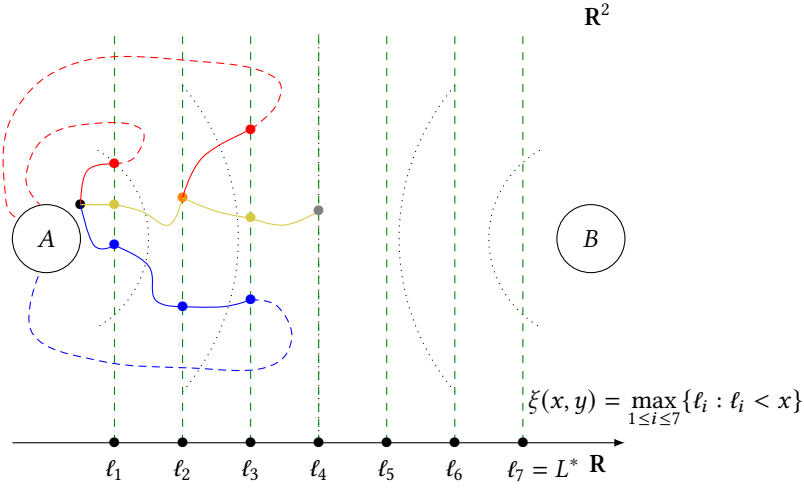
IPS on the state space (Symmetric SMC):



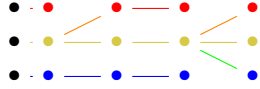


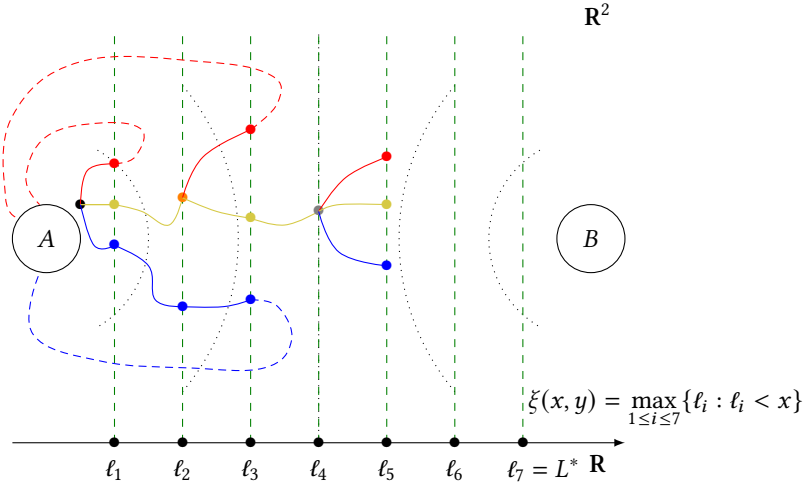
IPS on the state space (Symmetric SMC):



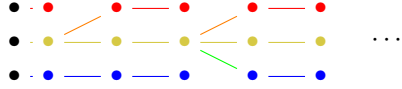


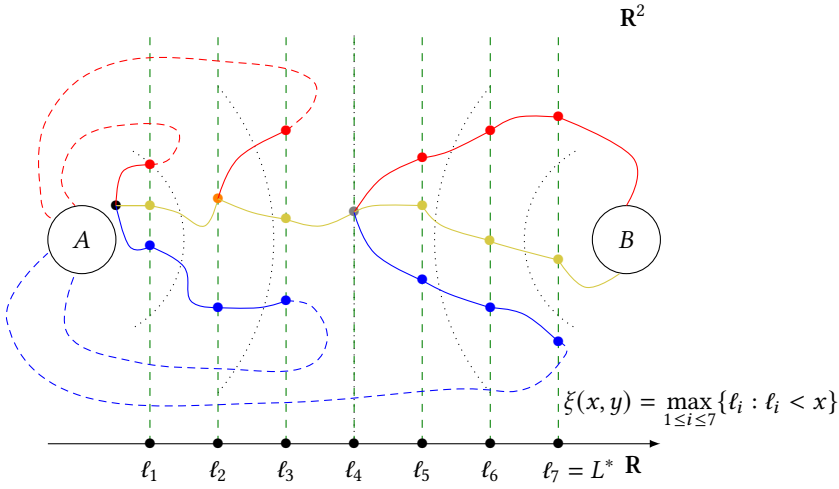
IPS on the state space (Symmetric SMC):



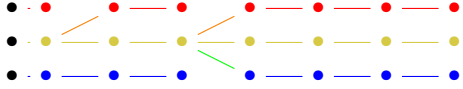


IPS on the state space (Symmetric SMC):





IPS on the state space (Symmetric SMC):



Summary of gAMS

- When the reaction coordinate is continuous, gAMS is an Adaptive SMC with asymmetric resampling scheme;
- When the reaction coordinate is discrete, gAMS enters into Symmetric SMC framework with fixed levels;
- In both cases mentioned above, the rare-event probability can be expressed as $\gamma_n(f)$.

Main results

Theorem (CLT)

Assume that $(G_n)_{n\geq 0}$ *is* $[0,1]$*-valued. We have*

$$\sqrt{N}\left(\gamma_n^N(f) - \gamma_n(f)\right) \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \widetilde{\sigma}^2_{\gamma_n}(f)\right),$$

with $\widetilde{\sigma}^2_{\gamma_n}(f) := \sum_{p=0}^{n}\left(\Gamma_n^{\dagger,(p)}(f^{\otimes 2}) - \Gamma_n^{\dagger,(\varnothing)}(f^{\otimes 2})\right) + \sum_{p=0}^{n-1}\widetilde{\Gamma}_n^{\dagger,(p)}(f^{\otimes 2})$.

Theorem (Variance estimation)

- *We propose an unbiased nonasymptotic variance estimator for* $\mathrm{Var}[\gamma_n^N(f)]$ *in Symmetric SMC, of time complexity* $\mathcal{O}(nN^2)$*;*
- *We also propose a biased consistent asymptotic variance estimator for* $\widetilde{\sigma}^2_{\gamma_n}(f)$ *in Asymmetric SMC, of time complexity* $\mathcal{O}(nN)$.

The analysis is done by developing a language called *generalized coalescent tree-based expansion*, which is notationally very heavy. It is then not covered in this presentation.

Table of Contents

Committor function

The committor function is defined by $\xi^*(x) = \mathbf{P}_x(\tau_B < \tau_A)$.

Why do we estimate the committor function?

- Optimal choice of reaction coordinate for gAMS in terms of asymptotic variance (see, e.g., [Cérou et al., 2019]);
- Ingredient for other methods (e.g. Importance Sampling [Lelièvre and Stoltz, 2016]).

PDE point of view:
Denote by A, B two metastable states. The committor function ξ^* is the solution of the following elliptic PDE:

$$-\nabla V \cdot \nabla u + \beta^{-1}\Delta u = 0 \quad \text{on } \mathbf{R}^d \backslash (A \cup B),$$

with the boundary conditions

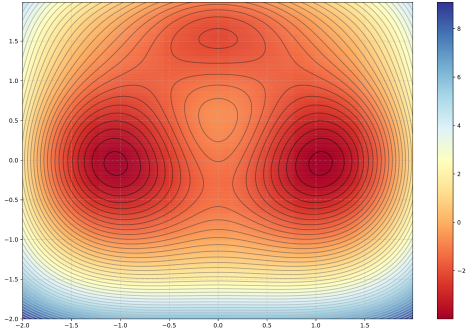
$$\begin{cases} u = 0 & \text{on } \partial A; \\ u = 1 & \text{on } \partial B. \end{cases}$$

This allows to calculate the committor function with finite difference method in low dimension, which provides a reference for our toy example.

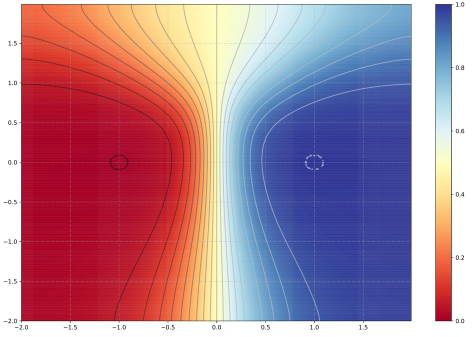
A 2-dimensional toy example

Three-hole potential:

$$V(x,y) := 3\exp\left(-x^2-(y-1/3)^2\right) - 3\exp\left(-x^2-(y-5/3)^2\right) - 5\exp\left(-(x-1)^2-y^2\right) - 5\exp\left(-(x+1)^2-y^2\right) + x^4/5 + (y-1/3)^2/5.$$



(a) Energy landscape of three-hole potential.



(b) Committor function approximated by finite difference method with $\beta = 1.67$.

Iterative update strategy with Mondrian Forests

Mondrian Forests (MF, [Lakshminarayanan et al., 2014]) are a nonparametric regressor that allows online learning. More concretely, MF is a variant of Random Forests whose splits do not depend on the response variables Y_i.

Supervised learning: Given a dataset $\{(X_i, Y_i) : 1 \leq i \leq K\}$, the goal is to estimate $\mathbf{E}[Y \mid X = x]$. In our situation, the training data is the couple

$$(x_0\text{: Starting point of gAMS algorithm}, \hat{\xi}^*(x_0)\text{: Estimated probability}),$$

and the conditional expectation is the committor function $\xi^*(x) = \mathbf{P}(\tau_B < \tau_A \mid X_0 = x)$.

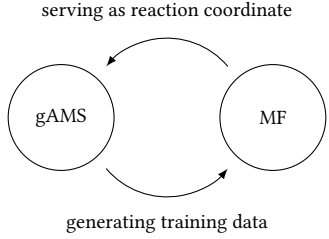


Figure: The illustration of iterative updating strategy.

Why Mondrian Forests?

Characteristics of our problem:

- Estimation should be extremely fast;
- We can control the quality of our training data (with variance estimation);
- We can control the distribution of X_i (active learning).

Advantages of MF:

- Nearly no tuning needed;
- Online learning is available, and the performance does not depend on the order of arrival of data;
- Fast in terms of estimation (average of binary decision trees);
- Prediction is a weighted average of the training data (the image is of finite values, hence theoretical consistency and variance estimation are available).

Other regressors? If the goal is only to estimate the committor function (rather than to improve gAMS at the same time), then almost all the SOTA machine/statistical learning methods can be considered. This is however beyond the scope of this thesis.

Mondrian Forests (lifetime $\lambda = \infty$)

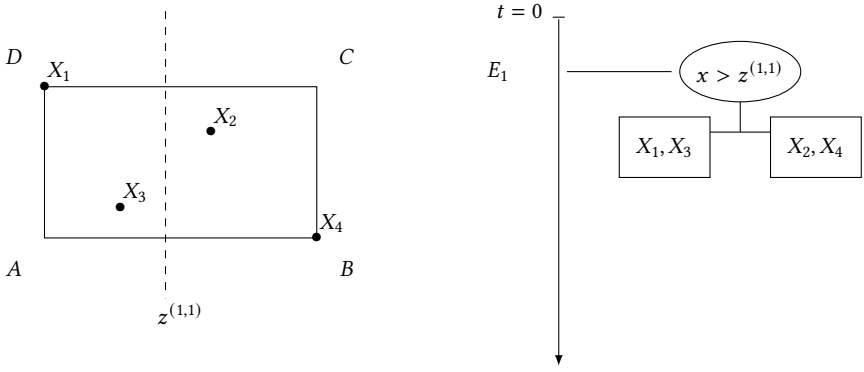


Figure: The first cut of a Mondrian tree along with lifetime.

Mondrian Forests (lifetime $\lambda = \infty$)

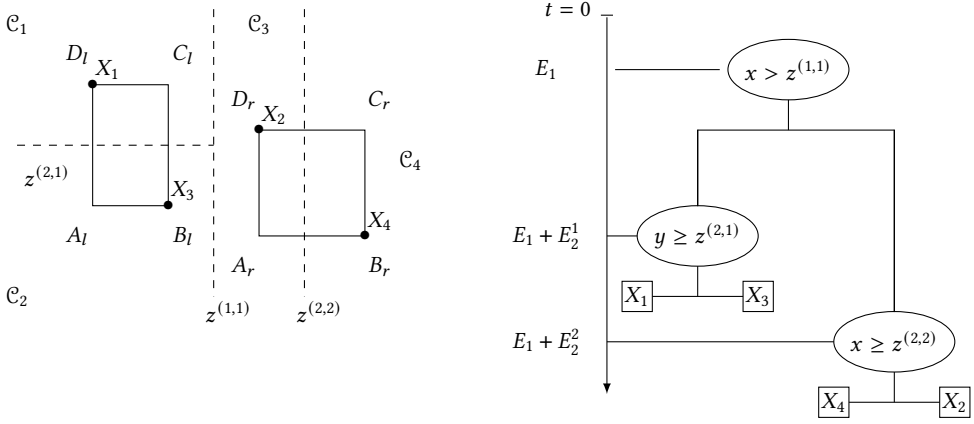
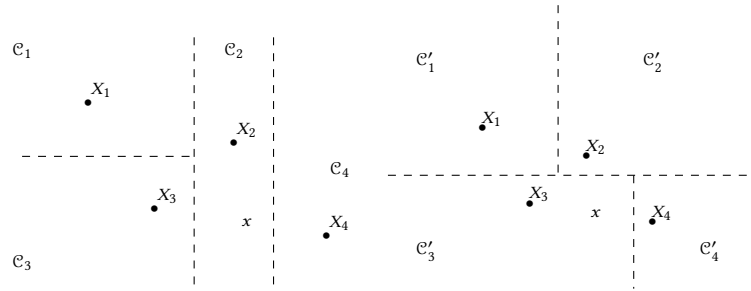


Figure: The second and third cuts of a Mondrian tree along with lifetime.

Prediction of a MF that contains two Mondrian trees



(a) First Mondrian tree.

(b) Second Mondrian tree.

Final prediction at the point x is the average of the prediction made by the Mondrian trees in the forest, i.e.,

$$\frac{1}{2}Y_2 + \frac{1}{2}Y_3$$

Online learning: Memoryless property of exponential random vriables.

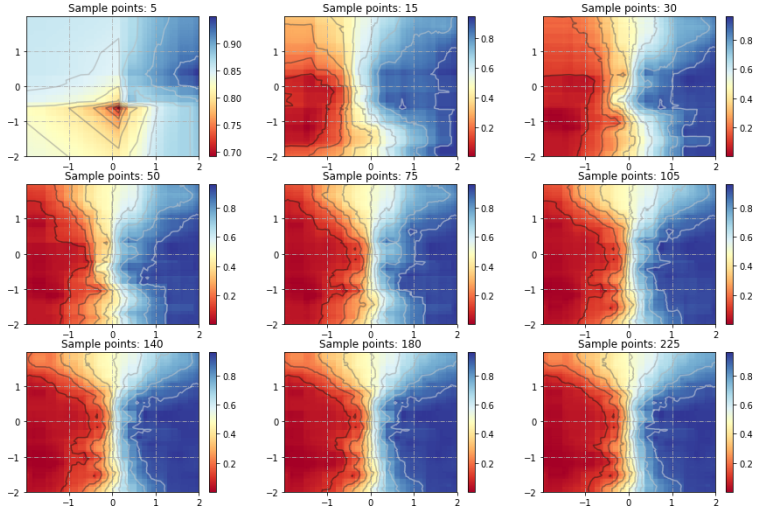


Figure: Learning with the training data generated by gAMS algorithm. The reaction coordinate is set to be the distance to the state A. All the sample points are uniformly drawn in the domain.

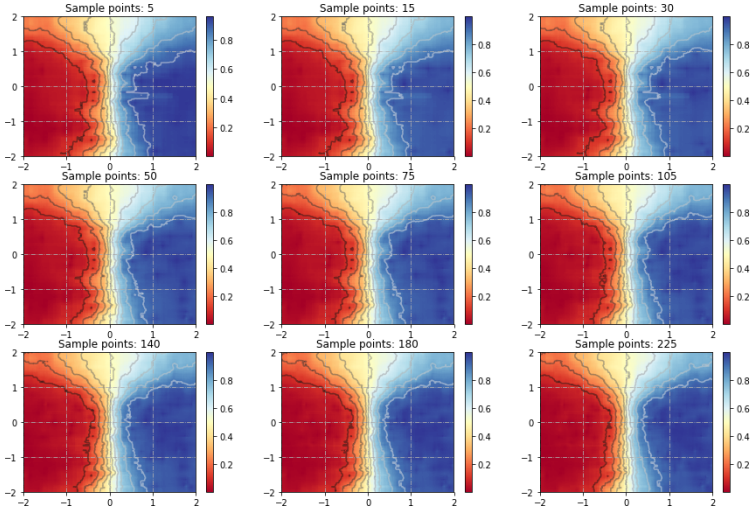


Figure: Learning with the data re-generated by gAMS algorithm. The reaction coordinate is the trained MF model. All the sample points are uniformly drawn in the domain.

Possible improvements

Tempering:

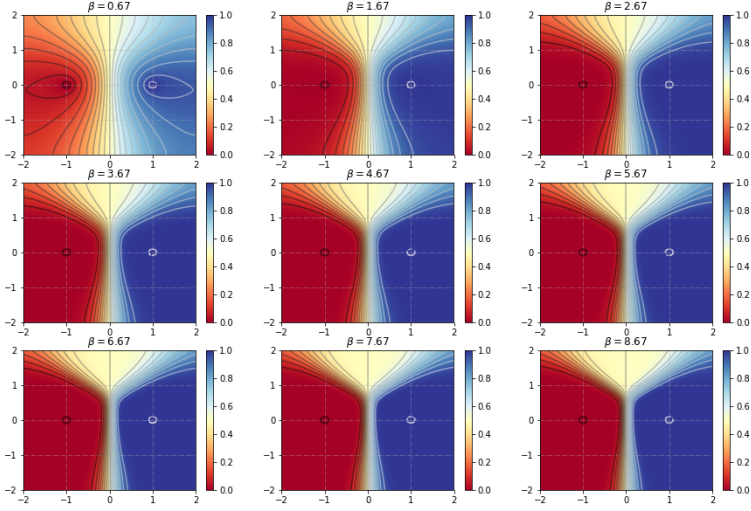


Figure: Illustrations of committor functions for different temperatures.

Possible improvements

Tempering (potential barrier):

- In high temperature setting, committor function can be estimated efficiently with naive Monte Carlo + MF;
- Use the MF model estimated with β_{low} (high temperature) to implement gAMS algorithm in the setting of β_{high} (low temperature);
- No pre-defined reaction coordinate needed.

Active learning (still open problems):

- In high dimensional setting, it becomes more difficult to use uniformly sampled X_i to estimate the committor function (curse of dimension);
- We only need to estimate part of the committor function, namely the "most visited zone" for the reaction trajectories, in order to improve the gAMS algorithm for a certain problem.

Conclusion

Contributions and Perspectives

Results:

- Provide uncertainty control for a wide range of Adaptive SMC models;
- Provide uncertainty control for gAMS in a specific setting;
- Combine the gAMS algorithm with modern statistical/machine learning.

Theoretical tools:

- Provide new strategy to deal with Adaptive Feynman-Kac particle models;
- Develop the language of generalized coalescent tree-based measures;
- Enrich the general theory of Feynman-Kac particle models.

Future research:

- Continuous-time Asymmetric SMC;
- Coalescent tree-based expansion in general IPS context (e.g. PMCMC);
- Develop an efficient, automated strategy, to estimate the committor function in a high dimensional and low temperature setting.

Thank you for your attention!

References

- [Au and Beck, 2001] Au, S. and Beck, J. (2001).
  Estimation of small failure probabilities in high dimensions by subset simulation.
  Probabilistic Engineering Mechanics, 16(4):263–277.

- [Beskos et al., 2016] Beskos, A., Jasra, A., Kantas, N., and Thiery, A. (2016).
  On the convergence of adaptive sequential Monte Carlo methods.
  Ann. Appl. Probab., 26(2):1111–1146.

- [Bréhier et al., 2016] Bréhier, C.-E., Gazeau, M., Goudenège, L., Lelièvre, T., and Rousset, M. (2016).
  Unbiasedness of some generalized adaptive multilevel splitting algorithms.
  Ann. Appl. Probab., 26(6):3559–3601.

- [Cérou et al., 2011] Cérou, F., Del Moral, P., and Guyader, A. (2011).
  A nonasymptotic theorem for unnormalized Feynman-Kac particle models.
  Ann. Inst. Henri Poincaré Probab. Stat., 47(3):629–649.

- [Cérou and Guyader, 2016] Cérou, F. and Guyader, A. (2016).
  Fluctuation analysis of adaptive multilevel splitting.
  Ann. Appl. Probab., 26(6):3319–3380.

- [Chan and Lai, 2013] Chan, H. P. and Lai, T. L. (2013).
  A general theory of particle filters in hidden Markov models and some applications.
  Ann. Statist., 41(6):2877–2904.

- [Cérou et al., 2019] Cérou, F., Guyader, A., and Rousset, M. (2019).
  Adaptive multilevel splitting: Historical perspective and recent results.
  Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(4):043108.

References (cont.)

- [Del Moral, 2004] Del Moral, P. (2004).
  Feynman-Kac formulae: Genealogical and interacting particle systems with applications.
  Probability and its Applications. Springer-Verlag, New York.

- [Del Moral, 2013] Del Moral, P. (2013).
  Mean field simulation for Monte Carlo integration, volume 126 of *Monographs on Statistics and Applied Probability.*
  CRC Press, Boca Raton, FL.

- [Gordon, 1993] Gordon, N. (1993).
  Novel approach to nonlinear/non-gaussian bayesian state estimation.
  IEE Proceedings F (Radar and Signal Processing), 140:107–113(6).

- [Lakshminarayanan et al., 2014] Lakshminarayanan, B., Roy, D. M., and Teh, Y. W. (2014).
  Mondrian forests: Efficient online random forests.
  In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27*, pages 3140–3148. Curran Associates, Inc.

- [Lee and Whiteley, 2018] Lee, A. and Whiteley, N. (2018).
  Variance estimation in the particle filter.
  Biometrika, 105(3):609–625.

- [Lelièvre and Stoltz, 2016] Lelièvre, T. and Stoltz, G. (2016).
  Partial differential equations and stochastic methods in molecular dynamics.
  Acta Numerica, 25:681–880.