Sequential Monte Carlo and Rare Events

Adam M. Johansen

a.m.johansen@warwick.ac.uk
http://go.warwick.ac.uk/amjohansen/talks/

ST912 Statistical Frontiers Seminar
February 2022

Warwick
Statistics

Overview

Motivating Problem Estimating (very) Small Probabilities

Monte Carlo Solutions Sequential Monte Carlo algorithms

Feynman-Kac Formulae A mathematical description

Splitting-type Algorithms SMC for Rare Events

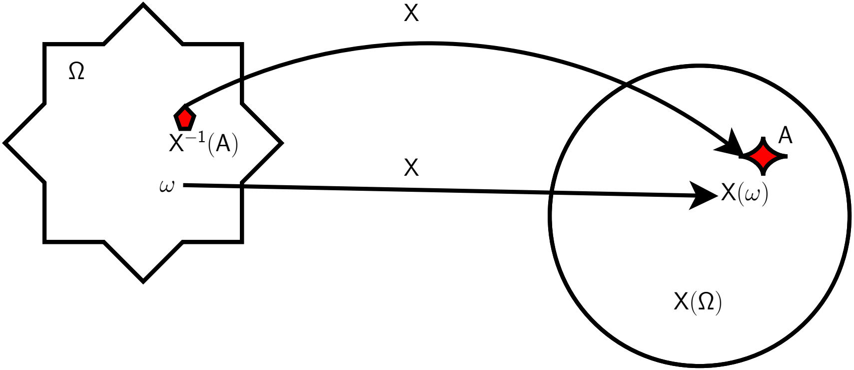
Unbiased Splitting Unbiased Algorithms in Continuous Time

Rare Event Estimation/Simulation

A Motivating Problem

Context

- Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$,
- and a random element $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$,
- what is $\mathbb{P}(X \in A) = \mathbb{P} \circ X^{-1}(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$,
- for some $A \in \mathcal{E}$ such that $\mathbb{P}(X \in A) \ll 1$?



Some Simple Examples

1. A really simple problem.
   - Let
     $$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$
   - What is $\mathbb{P}(X \in A)$ if $A = [a, \infty)$ for $a \gg 1$?
   - Simple semi-analytic solution $1 - \Phi(a)$.

2. A somewhat harder problem:
   - Let
     $$f(\mathbf{x}) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right).$$
   - What is $\mathbb{P}(X \in A)$ if $A = \otimes_{i=1}^{d}[a_i, b_i]$?
   - What can we say about $\mathrm{Law}(X)|_A$
3. Getting more interesting:
   - Let $dX_t = a(X_t)dt + b(X_t)dB_t$.
   - What is $\mathbb{P}(\zeta(X_{[0,T]}) \in A)$?
   - What is $\mathbb{P}(X_\sigma \in A)$ if $\sigma = \inf\{t : X_t \in A \cup R\}$?

Sampling Approaches

The Monte Carlo Method

- Given a probability density, f, and $\varphi : E \to \mathbb{R}$

$$I = \int_E \varphi(x) f(x) dx$$

- Simple Monte Carlo solution:
  - Sample $X_1, \ldots, X_N \overset{\text{i.i.d.}}{\sim} f$.
  - Estimate $\widehat{I} = \frac{1}{N} \sum_{i=1}^{N} \varphi(X_i)$.

Justified by the law of large numbers...

and the central limit theorem.

- Can also be viewed as approximating $\pi(dx) = f(x)dx$ with

$$\widehat{\pi}^N(dx) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}(dx).$$

Justified by Glivenko-Cantelli type results.

The Monte Carlo Method and Rare Events

- Use $\mathbb{P}(X \in A) \equiv \mathbb{E}[\mathbb{I}_A(X)] = I(\mathbb{I}_A)$.
- Then, directly:

$$\mathbb{P}(X \in A) \approx \hat{I}_n(\mathbb{I}_A) = \frac{|A \cap \{X_1, \ldots, X_n\}|}{n}.$$

Simple Monte Carlo and the Toy Problem

a	$\log(\hat{I}_{10^k}(\mathbb{I}_{[a,\infty)}))$							$\log$
	k	1	2	3	4	5	6	7	$1-\Phi(a)$
1		-2.30	-1.66	-1.80	-1.82	-1.83	-1.84	-1.84	-1.84
2			-3.91	-3.73	-3.76	-3.78	-3.79	-3.79	-3.78
3				-6.91	-6.81	-6.59	-6.60	-6.61	-6.61
4						-10.12	-10.26	-10.42	-10.36
5								-14.73	-15.06
6									-20.74

Simple calculations reveal:

- $\mathbb{E}[\hat{I}_n(\mathbb{I}_{[a,\infty)})] = \mathbb{P}(X \in [a,\infty))$
- $\mathrm{Var}[\hat{I}_n(\mathbb{I}_{[a,\infty)})] = \frac{1}{n}\mathbb{P}(X \in [a,\infty))(1-\mathbb{P}(X \in [a,\infty)))$
- So the relative standard deviation is $\sim (n\mathbb{P}(X \in [a,\infty)))^{-1/2}$.

Simple Monte Carlo and the Toy Problem

a	$\log(\hat{I}_{10^k}(\mathbb{I}_{[a,\infty)}))$							$\log$
	k	1	2	3	4	5	6	7	$1-\Phi(a)$
1		-2.30	-1.66	-1.80	-1.82	-1.83	-1.84	-1.84	-1.84
2			-3.91	-3.73	-3.76	-3.78	-3.79	-3.79	-3.78
3				-6.91	-6.81	-6.59	-6.60	-6.61	-6.61
4						-10.12	-10.26	-10.42	-10.36
5								-14.73	-15.06
6									-20.74

Simple calculations reveal:

- $\mathbb{E}[\hat{I}_n(\mathbb{I}_{[a,\infty)})] = \mathbb{P}(X \in [a,\infty))$
- $\mathrm{Var}[\hat{I}_n(\mathbb{I}_{[a,\infty)})] = \frac{1}{n}\mathbb{P}(X \in [a,\infty))(1-\mathbb{P}(X \in [a,\infty)))$
- So the relative standard deviation is $\sim (n\mathbb{P}(X \in [a,\infty)))^{-1/2}$.

Variance Reduction

- Want $\hat{p}_n$ such that $\hat{p}_n \approx \mathbb{P}(X \in A) =: p$:
  - Ideally, with $\mathbb{E}[\hat{p}_n] = p$.
  - Such that $\mathrm{Var}(\hat{p}_n) \ll p^2$.
  - For modest n.
- Controlling variance is the key issue.
  - Importance Sampling.
  - Splitting.
  - Interacting Particle Systems.
  - Sequential Monte Carlo.

The Importance–Sampling Identity

- Given g, such that
  - $f(x) > 0 \Rightarrow g(x) > 0$
  - and $f(x)/g(x) < \infty$,

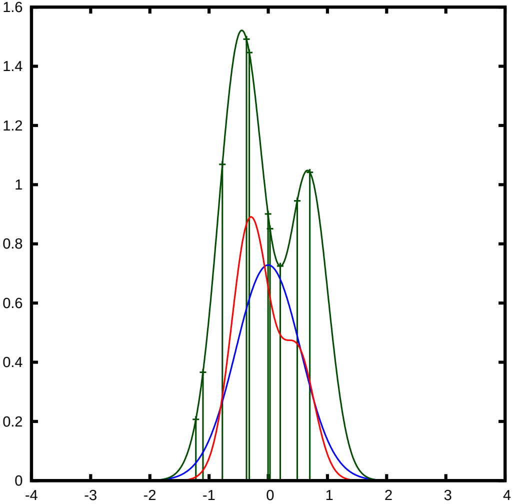
  define $w(x) = f(x)/g(x)$ and:

$$\int \varphi(x)f(x)dx = \int \varphi(x)f(x)g(x)/g(x)dx = \int \varphi(x)w(x)g(x)dx.$$

- This suggests the importance sampling estimator:
  - Sample $X_1, \ldots, X_N \overset{\text{i.i.d.}}{\sim} g$.
  - Estimate $\widehat{I} = \frac{1}{N}\sum_{i=1}^{N} w(X_i)\varphi(X_i)$.
- Can also be viewed as approximating $\pi(dx) = f(x)dx$ with

$$\widehat{\pi}^N(dx) = \frac{1}{N}\sum_{i=1}^{N} w(X_i)\delta_{X_i}(dx).$$

Importance Sampling Example



Importance Sampling Variance

The variance of this estimator is:

$$
\begin{aligned}
&\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} w(Y_i)\varphi(Y_i)\right] \\
=&\frac{1}{n}\mathrm{Var}\left[w(Y_1)\varphi(Y_1)\right] \\
=&\frac{1}{n}\left\{\mathbb{E}\left[(w(Y_1)\varphi(Y_1))^2\right] - \mathbb{E}\left[w(Y_1)\varphi(Y_1)\right]^2\right\} \\
=&\frac{1}{n}\left\{\int (w(y)\varphi(y))^2\, g(dy) - \left(\int w(y)\varphi(y) g(dy)\right)^2\right\} \\
=&\frac{1}{n}\left\{\int w(y)\varphi^2(y) f(dx) - \mathbb{E}[\varphi(X)]^2\right\}
\end{aligned}
$$

Optimal Importance Sampling

Proposition

Let $X \sim f$, where $f(dx) = f(x)dx$, with values in $(E, \mathcal{E})$ and let $\phi : \mathbb{R} \to (0, \infty)$ a function of interest. The proposal which minimizes the variance of the importance sampling estimator of $\mathbb{E}[\varphi(X)]$ is $g(x)dx$, where:

$$g(x) = \frac{f(x)\varphi(x)}{\int f(y)\varphi(y)dy}$$

Note: if $E \supset A \supset \mathrm{supp}\ \varphi(x)$, it suffices for $f|_A \ll g|_A$.

Importance Sampling and the Toy Problem

a	k 1	2	3	4	5	6	7	$\log 1-\Phi(a)$
	$\log(\hat{I}_{10^k}(\mathbb{I}_{[a,\infty)}))$							
1	-1.72	-1.84	-1.83	-1.84	-1.84	-1.84	-1.84	-1.84
2	-3.63	-3.78	-3.79	-3.78	-3.78	-3.78	-3.78	-3.78
3	-6.43	-6.59	-6.63	-6.60	-6.61	-6.61	-6.61	-6.61
4	-10.16	-10.34	-10.40	-10.35	-10.36	-10.36	-10.36	-10.36
5	-14.85	-15.04	-15.12	-15.06	-15.07	-15.06	-15.06	-15.06
6	-20.51	-20.72	-20.81	-20.73	-20.73	-20.74	-20.74	-20.74
7	-27.16	-27.37	-27.46	-27.38	-27.39	-27.38	-27.38	-27.38
8	-34.79	-35.01	-35.10	-35.02	-35.01	-35.01	-35.01	-35.01
9	-43.41	-43.64	-43.73	-43.63	-43.63	-43.63	-43.62	-43.63

Using $g(x) = \exp(-(x-a))\mathbb{I}_{[a,\infty)}(x)$.

Self-Normalised Importance Sampling

- Often, f is known only up to a normalising constant.
- If $v(x) = cf(x)/g(x) = cw(x)$, then

$$\frac{\mathbb{E}_g(v\varphi)}{\mathbb{E}_g(v\mathbf{1})} = \frac{\mathbb{E}_g(cw\varphi)}{\mathbb{E}_g(cw\mathbf{1})} = \frac{c\mathbb{E}_f(\varphi)}{c\mathbb{E}_f(\mathbf{1})} = \mathbb{E}_f(\varphi).$$

- Estimate the numerator and denominator with the same sample:

$$\widehat{I} = \frac{\sum\limits_{i=1}^{N} v(X_i)\varphi(X_i)}{\sum\limits_{i=1}^{N} v(X_i)}.$$

- Biased for finite samples, but consistent.
- Typically reduces variance.

Discrete Time Markov Processes and Rare Events

- Often rare events are described in terms of a process.
- In discrete time, say

$$X_1 \sim p_1(x_1)dx_1 \qquad X_n|X_{1:n-1} \sim p_n(x_n|x_{n-1})dx_n$$

  where each X_i takes values in a space E.
- Questions might be of the form what is $\mathbb{P}(\xi(X_{1:n}) \in A)$ for $\xi : E^n \to \mathbb{R}$ and $A \subset \mathbb{R}$:
  - E.g. If $E = \mathbb{R}$, $p_1(x_1) \propto \exp(-x_1^2/2)$ and $p_n(x_n|x_{n-1}) \propto \exp(-(x_n - x_{n-1}^2)/2)$ and $A = [a, \infty)$.
  - If $p_1, \ldots$ characterizes the evolution of differential group delay in a fibre optic cable $A = [a, \infty)$ and

$$\xi(x_{1:n}) = \max\left\{ || \sum_{p=1}^{q} \xi'(x_q) || : q \leq n \right\}$$

    where ξ' is essentially a projection of the underlying state one arrives at a real engineering problem.
- We'll return to continuous time processes later.

Importance Sampling for DTMPs

- Sample $\{X_{1:n}^{(i)}\}$ at time n from $q_n(x_{1:n})$, define

$$w_n(x_{1:n}) \propto \frac{p_n(x_{1:n})}{q_n(x_{1:n})}$$

  where $p_n(x_{1:n}) = p_1(x_1) \prod_{m=2}^{n} p_(m x_m | x_{m-1})$.
- set $W_n^{(i)} = w_n(X_{1:n}^{(i)}) / \sum_j w_n(X_{1:n}^{(j)})$,
- then $\{W_n^{(i)}, X_n^{(i)}\}$ is a consistently weighted sample.
- This seems inefficient.

Sequential Importance Sampling (SIS) I

- Importance weight

$$
\begin{aligned}
w_n(x_{1:n}) \propto & \frac{p_1(x_1)\prod_{m=2}^{n} p_m(x_m|x_{m-1})}{q_n(x_{1:n})} \\
= & \frac{p_1(x_1)}{q_n(x_1)} \prod_{m=2}^{n} \frac{p_m(x_m|x_{m-1})}{q_n(x_m|x_{1:m-1})}
\end{aligned}
$$

- Given $\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\}$ targetting $p_{n-1}(x_{1:n-1})$
  - Let $q_n(x_{1:n-1}) = q_{n-1}(x_{1:n-1})$,
  - sample $X_n^{(i)} \overset{\text{i.i.d.}}{\sim} q_n(\cdot|X_{1:n-1}^{(i)})$ or even $q_n(\cdot|X_{n-1}^{(i)})$.

Sequential Importance Sampling (SIS) II

- And update the weights:

$$\begin{aligned} w_n(x_{1:n}) =& w_{n-1}(x_{1:n-1})\frac{p(x_n|x_{n-1})}{q_n(x_n|x_{n-1})} \\ W_n^{(i)} =& w_n(X_{1:n}^{(i)}) \\ =& w_{n-1}(X_{1:n-1}^{(i)})\frac{p(X_n^{(i)}|X_{n-1}^{(i)})}{q_n(X_n^{(i)}|X_{n-1}^{(i)})} \\ =& W_{n-1}^{(i)}\frac{p(X_n^{(i)}|X_{n-1}^{(i)})}{q_n(X_n^{(i)}|X_{n-1}^{(i)})} \end{aligned}$$

- If $\int p(x_{1:n}|y_{1:n})dx_n \approx p(x_{1:n-1}|y_{1:n-1})$ this makes sense.
- We only need to store $\{W_n^{(i)}, X_{n-1:n}^{(i)}\}$.
- Same computation every iteration.

Importance Sampling on Huge Spaces Doesn't Work

- It's said that IS *breaks the curse of dimensionality*:

$$\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N} w(X_i)\varphi(X_i) - \int \varphi(x)f(x)dx\right] \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathbb{V}\mathrm{ar}_g\left[w\varphi\right])$$

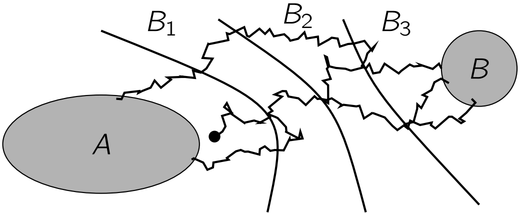
- This is true.
- But it's not *enough*.
- $\mathbb{V}\mathrm{ar}_g\left[w\varphi\right]$ increases (often exponentially) with dimension.
- **Eventually**, an SIS estimator (of $p(x_{1:n})$) **will** fail.
- But $p(x_n) = \int p(x_{1:n})dx_{1:n-1}$ is a *fixed-dimensional* distribution...which has implications which we will revisit.

Multilevel Splitting

Returning to continuous time processes

Splitting

An insight dating back to the 1950s:



Allowing $\sigma_{B_i} = \inf\{t : X_t \in B_i \cup A\}$:

$$\mathbb{P}(X_{\sigma_B} \in B) = \mathbb{P}(X_{\sigma_{B_1}} \in B_1) \prod_{i=2}^{m} \mathbb{P}(X_{\sigma_{B_i}} \in B_i | X_{\sigma_{B_{i-1}}} \in B_{i-1})$$

where $B_1 \subset B_2 \subset \cdots \subset B_m = B$ and A is positive recurrent for X.
We can estimate each term separately.

The Discrete Skeleton of MLS

Algorithmically, this idealised algorithm reduces a continuous-time problem to a discrete-time on:

- Let λ denote the initial distribution: $X(0) \sim \lambda$.
- Define $U_i = \left(\sigma_{B_i}, X_{B_i}(\sigma_i)\right)$, $i = 1, \dots, m$
- Let $M_i : (\mathbb{R}_{\geq 0} \times \mathbb{R}^d) \times \mathcal{S} \to [0, 1]$ denote the Markov kernels of this discrete-time process [1].
- Define $G_i : \mathbb{R}_{\geq 0} \times \mathbb{R}^d \to \{0, 1\}$ as:

$$G_i(t, x) = \begin{cases} 1, & \text{if } x \in B_i, \\ 0, & \text{otherwise.} \end{cases}$$

[1] Where $\mathcal{S}$ be the Borel sigma algebra associated with $\mathbb{R}_{\geq 0} \times \mathbb{R}^d$

The Algorithm

Algorithm Idealised Multilevel Splitting

Given λ; $G_1, \ldots, G_m$; $M_1, \ldots, M_m$; N_0; and $R_1, \ldots, R_{m-1}$:

1. For $j = 1, \ldots, N_0$, draw independently:
$$X_1^j(0) \sim \lambda \text{ and } U_1^j \sim M_1\left(\left(0, X_1^j(0)\right), \cdot\right).$$
2. Let $S_1 = \{U_1^j : G_1(U_1^j) = 1\}$ be the survivors, and $N_1 = |S_1|$.
3. For $i = 2, \ldots, m$:
   - 3.1 If $N_{i-1} = 0$, return $\hat{p} = 0$.
   - 3.2 Given $S_{i-1} = \{\bar{U}_{i-1}^j\}_{j=1}^{N_{i-1}}$, for all $(j,k) \in \{(j', k') : 1 \leq j' \leq N_{i-1}, 1 \leq k' \leq R_{i-1}\}$ sample independently $U_i^{j,k} \sim M_i(\bar{U}_{i-1}^j, \cdot)$.
   - 3.3 Let $S_i = \{U_i^{j,k} : G_i(U_i^{j,k}) = 1\}$, and $N_i = |S_i|$.
4. Return $\hat{p} = \frac{N_m}{N_0 \prod_{i=1}^{m-1} R_i}$.

Requires the choice of $R_1, \ldots, R_{m-1}$ and samples from $M_1, \ldots, M_m$.

Beyond Idealised Multilevel Splitting

Fixed-effort Splitting Addresses the first issue.

- Rather than fixing R_i, fix N_i.
- Sample N_i times with replacement from the survivors from the previous iteration.
- Or something motivated by similar considerations.
- Can be analysed directly, but it can also be viewed as a particular instance of sequential Monte Carlo.

Multilevel Splitting with Couplings Addresses the second.

- Avoid sampling from $M_1, \ldots, M_m$.
- Instead sample from a more tractable transition.
- Utilizes ε-strong simulation from the law of diffusion processes.
- And an additional modification to make the algorithm tractable.

SMC: Sequential Importance Resampling

Resampling

- We can produce unweighted samples from weighted ones.
- Given $\{W_i, X_i\}_{i=1}^N$ an unbiased resampling $\{\tilde{X}_i\}_{i=1}^N$ is such that

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \varphi(\tilde{X}_i)\middle|\{W_i, X_i\}_{i=1}^N\right] = \sum_{i=1}^N W_i\varphi(X_i)$$

  for any continuous bounded φ.
- Simplest option: sample from empirical distribution

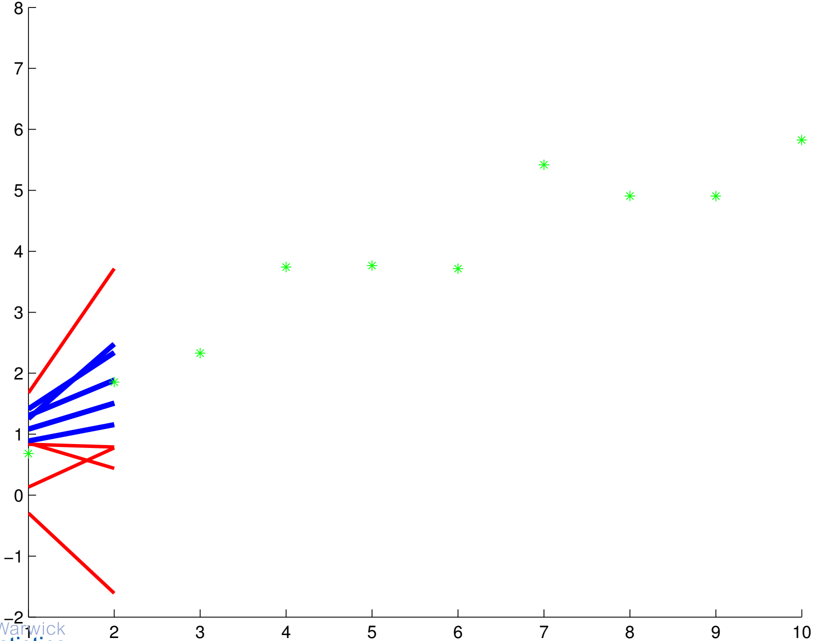
$$\tilde{X}_1, \ldots, \tilde{X}_N \overset{iid}{\sim} \sum_{j=1}^N W_j \delta_{X_j}(\cdot)$$

- Other approaches reduce the *additional* variance.

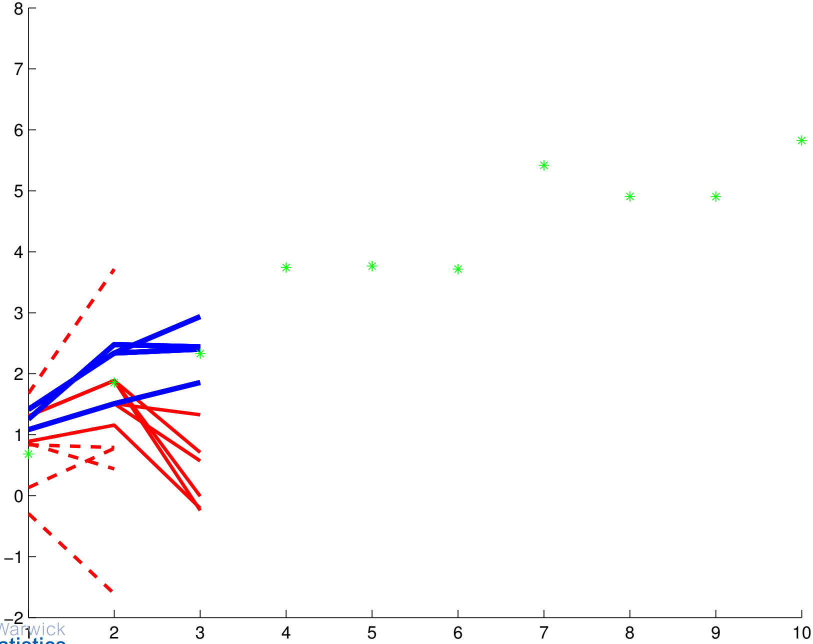
The SIR[esampling] Algorithm

- Problem: variance of the weights in SIS builds up over time.
- Solution? Given $\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\}$:
  1. **Resample**, to obtain $\{\frac{1}{N}, \widetilde{X}_{1:n-1}^{(i)}\}$.
  2. Sample $X_n^{(i)} \sim q_n(\cdot|\widetilde{X}_{n-1}^{(i)})$.
  3. Set $X_{1:n-1}^{(i)} = \widetilde{X}_{1:n-1}^{(i)}$.
  4. Set $W_n^{(i)} \propto p_n(X_n^{(i)}|X_{n-1}^{(i)})/q_n(X_n^{(i)}|X_{n-1}^{(i)})$ with $\sum_i W_n^{(i)} = 1$.
- And continue as with SIS.
- Actually, we only need to be able to evaluate up to a normalizing constant: see step 4.
- There is a cost, but this really works...
  at least for some problems.

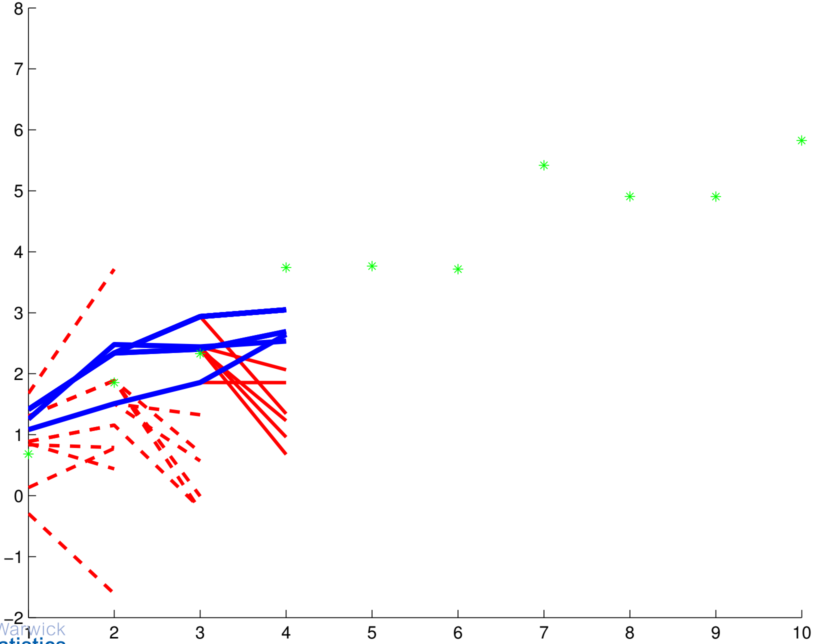
Iteration 2



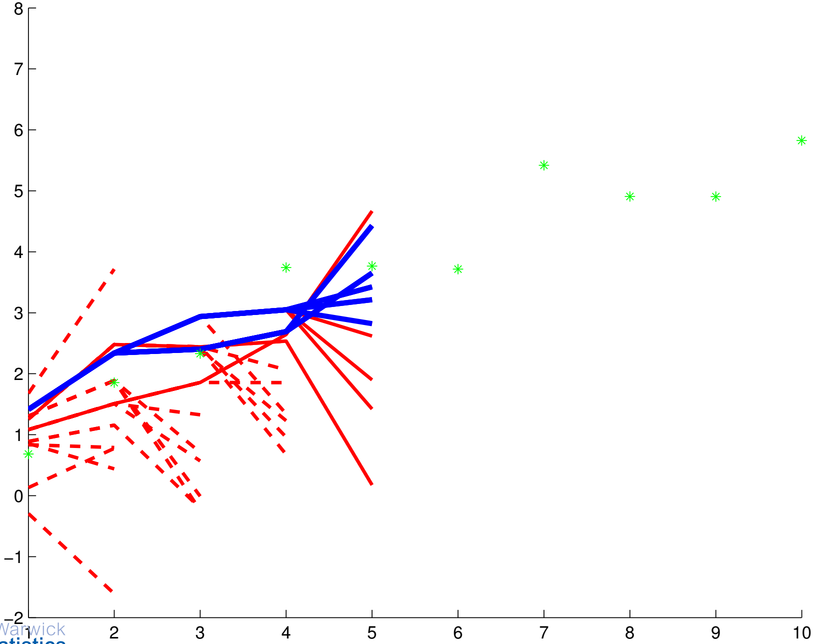
Iteration 3



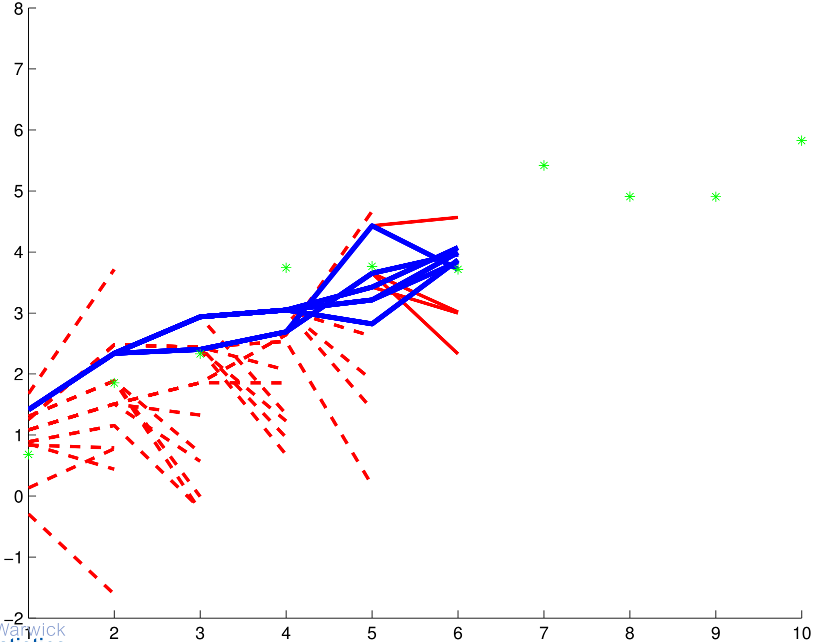
Iteration 4



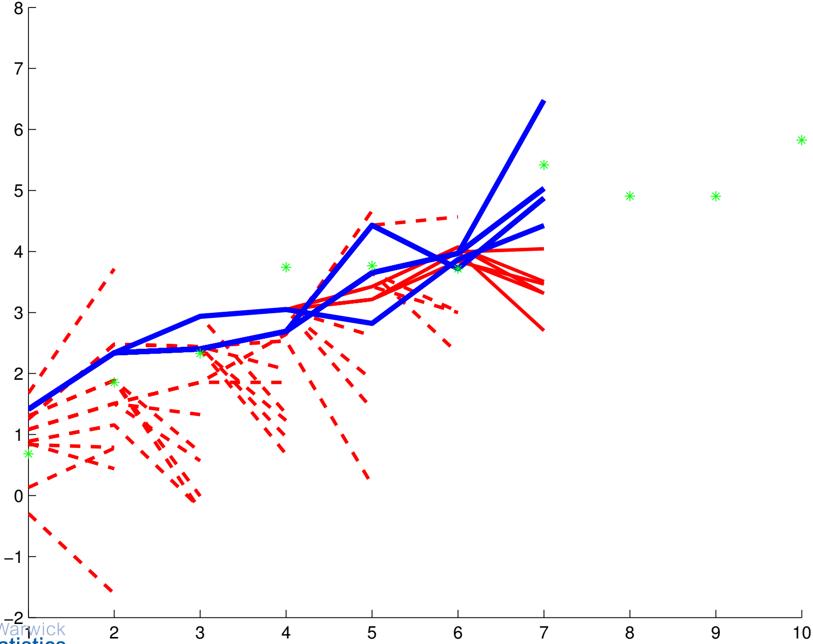
Iteration 5



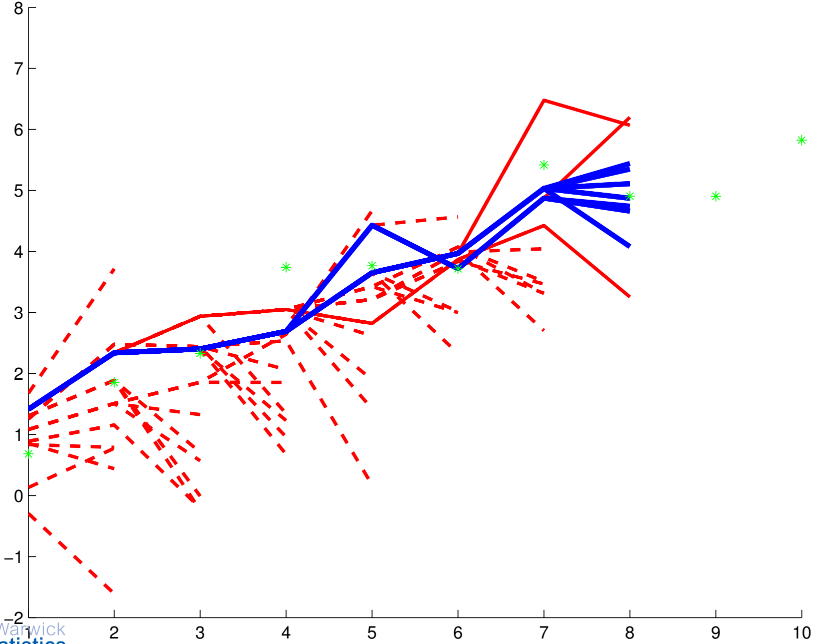
Iteration 6



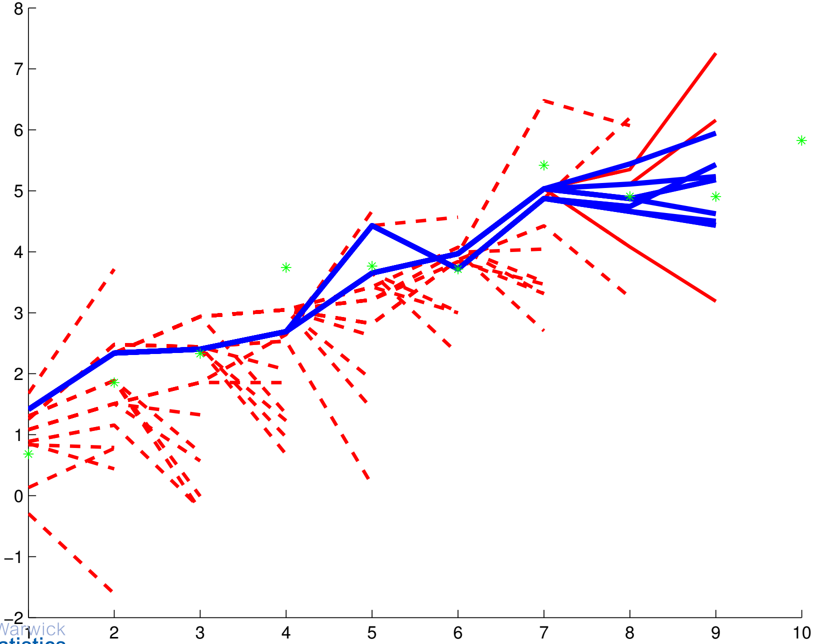
Iteration 7



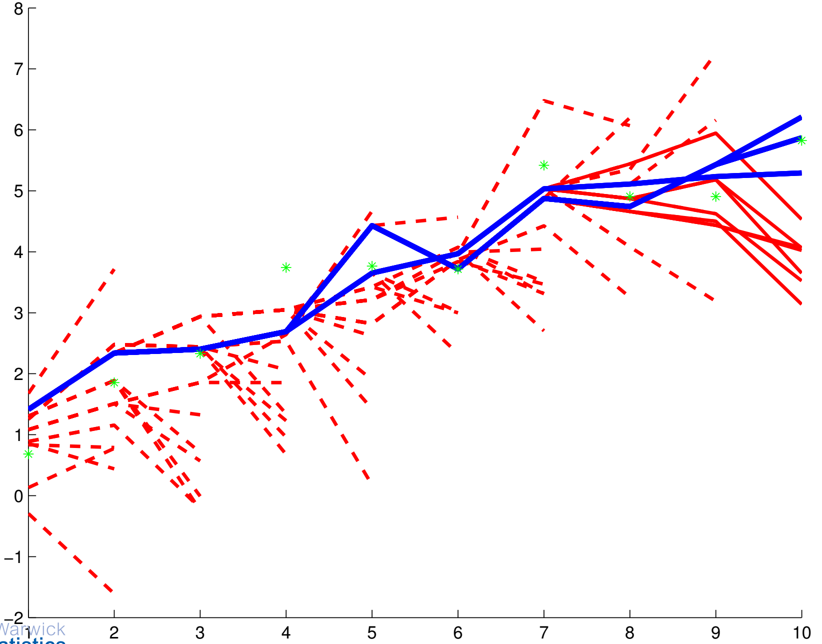
Iteration 8



Iteration 9



Iteration 10



What Are the Target and Proposal for Splitting?

- Let

$$q_1(\tau_1, x_1) = \int \lambda(dx_0) M_1((0, x_0), \cdot)$$
$$q_i = \int M_i((\sigma_i, x_i), \cdot) \quad i = 2, \ldots, m$$

- Let $G_i(\tau_i, x_i) = \mathbb{I}_{B_i}(x_i)$
- Set

$$\gamma_i(d(\tau_j, x_j)_{j=1}^i) = q_1(\tau_1, x_1) G_1(\tau_1, x_1) \prod_{j=1}^{i} q_i((\tau_{i-1}, x_{i-1}), d(\tau_i, x_i))$$
$$p_i \propto \gamma_i$$

- In sampling we only need γ_i and can recover estimates of its normalizing constant.

Feynman-Kac Formulæ

A Probabilistic Perspective

This entire section can be skipped and the rest of the presentation should remain accessible, it's here only for those who are interested in the probabilistic foundations of these algorithms.

Feynman-Kac Formulæ

- A natural description for measure-valued stochastic processes.
- Model for:
  - Particle motion in absorbing environments.
  - Classes of branching particle system.
  - Simple genetic algorithms.
  - Particle filters and related algorithms.

Elements of this framework:

- Probabilistic Construction
- Semigroup[oid] Structure
- McKean Interpretations
- Particle Approximations
- Selected Results

Probabilistic Construction

Following Del Moral (2004)

The Canonical Markov Chain

- Consider the filtered probability space:

$$(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\in\mathbb{N}}, \mathbb{P}_\mu)$$

- Let $\{X_n\}_{n\in\mathbb{N}}$ be a Markov chain such that for any $n \in \mathbb{N}$:

$$\mathbb{P}_\mu(X_{1:n} \in dx_{1:n}) = \mu(dx_1)\prod_{i=2}^{n} M_i(x_{i-1}, dx_i)$$

$$X_i : \Omega \to E_i \qquad \mu \in \mathcal{P}(E_1) \qquad M_i : E_{i-1} \to \mathcal{P}(E_i)$$

- $(E_i, \mathcal{E}_i)$ are measurable spaces.
- The X_i are $\mathcal{E}_i/\mathcal{F}_i$-measurable.
- Using Kolmogorov's/Tulcea's extension theorem there exists a unique process-valued extension.

Some Operator Notation

Given two measurable spaces, $(E, \mathcal{E})$ and $(F, \mathcal{F})$, a measure μ on $(E, \mathcal{E})$ and a Markov kernel, $K : E \to \mathcal{P}(F)$, define:

$$\mu(\varphi_E) := \int \mu(dx)\varphi_E(x)$$

$$\mu K(\varphi_F) := \int \mu(dx)K(x, dy)\varphi_F(y) \qquad \mu K \in \mathcal{P}(F)$$

$$K(\varphi_F)(x) := \int K(x, dy)\varphi_F(y) \qquad K(\varphi_F) : E \to \mathbb{R}$$

with φ_E, φ_F suitably measurable functions.

Given two functions, $g, h : E \to \mathbb{R}$, define $g \cdot h : E \to \mathbb{R}$ via $(g \cdot h)(x) = g(x)h(x)$.

Given $e : E \to \mathbb{R}$ and $f : F \to \mathbb{R}$, let $(e \otimes f)(x, y) := e(x)f(y)$.

The Feynman-Kac Formulæ

- Given $\mathbb{P}_\mu$ and **potential functions**:

$$\{G_i\}_{i \in \mathbb{N}} \qquad\qquad G_i : E_i \rightarrow [0, \infty)$$

- Define two path measures weakly:

$$\mathbb{Q}_n(\varphi_{1:n}) = \frac{\mathbb{E}\left[\varphi_{1:n}(X_{1:n}) \prod_{i=1}^{n-1} G_i(X_i)\right]}{\mathbb{E}\left[\prod_{i=1}^{n-1} G_i(X_i)\right]}$$

$$\widehat{\mathbb{Q}}_n(\varphi_{1:n}) = \frac{\mathbb{E}\left[\varphi_{1:n}(X_{1:n}) \prod_{i=1}^{n} G_i(X_i)\right]}{\mathbb{E}\left[\prod_{i=1}^{n} G_i(X_i)\right]}$$

where $\varphi_{1:n} : \otimes_{i=1}^{n} E_i \rightarrow \mathbb{R}$.

Example (Filtering via FK Formulæ: Prediction)

- Let $\mu(x_1) = f(x_1)$, $M_n(x_{n-1}, dx_n) = f(x_n|x_{n-1})dx_n$.
- Let $G_n(x_n) = g(y_n|x_n)$.
- Then:

$$\begin{aligned}
\mathbb{Q}_n(\varphi_{1:n}) =& \mathbb{E}\left[\varphi_{1:n}(X_{1:n})\prod_{i=1}^{n-1} G_i(X_i)\right] \Big/ \mathbb{E}\left[\prod_{i=1}^{n-1} G_i(X_i)\right] \\
=& \mathbb{E}\left[\varphi_{1:n}(X_{1:n})\prod_{i=1}^{n-1} g(y_i|X_i)\right] \Big/ \mathbb{E}\left[\prod_{i=1}^{n-1} g(y_i|X_i)\right] \\
=& \frac{\int \left[f(x_1)\prod_{i=2}^{n} f(x_i|x_{i-1})\right]\left[\prod_{j=1}^{n-1} g(y_j|x_j)\right]\varphi_{1:n}(x_{1:n})dx_{1:n}}{\int \left[f(x_1)\prod_{i=2}^{n} f(x_i|x_{i-1})\right]\left[\prod_{j=1}^{n-1} g(y_j|x_j)\right]dx_{1:n}} \\
=& \int p(x_{1:n}|y_{1:n-1})\varphi_{1:n}(x_{1:n})dx_{1:n}
\end{aligned}$$

Example (Filtering via FK Formulæ: Update/Filtering)

- Whilst:

$$
\begin{aligned}
\widehat{\mathbb{Q}}_n(\varphi_{1:n}) =& \mathbb{E}\left[\varphi_{1:n}(X_{1:n})\prod_{i=1}^{n} G_i(X_i)\right] \Big/ \mathbb{E}\left[\prod_{i=1}^{n} G_i(X_i)\right] \\
=& \mathbb{E}\left[\varphi_{1:n}(X_{1:n})\prod_{i=1}^{n} g(y_i|X_i)\right] \Big/ \mathbb{E}\left[\prod_{i=1}^{n} g(y_i|X_i)\right] \\
=& \frac{\int \left[f(x_1)\prod_{i=2}^{n} f(x_i|x_{i-1})\right]\left[\prod_{j=1}^{n} g(y_j|x_j)\right]\varphi_{1:n}(x_{1:n})dx_{1:n}}{\int \left[f(x_1)\prod_{i=2}^{n} f(x_i|x_{i-1})\right]\left[\prod_{j=1}^{n} g(y_j|x_j)\right]dx_{1:n}} \\
=& \int p(x_{1:n}|y_{1:n})\varphi_{1:n}(x_{1:n})dx_{1:n}
\end{aligned}
$$

Feynman-Kac Marginal Measures

We are typically interested in marginals:

"Predicted"

$$\gamma_n(\varphi_n) = \mathbb{E}\left[\varphi_n(X_n)\prod_{i=1}^{n-1} G_i(X_i)\right]$$

$$\eta_n(\varphi_n) = \frac{\mathbb{E}\left[\varphi_n(X_n)\prod_{i=1}^{n-1} G_i(X_i)\right]}{\mathbb{E}\left[\prod_{i=1}^{n-1} G_i(X_i)\right]} = \gamma_n(\varphi_n)/\gamma_n(\mathbf{1})$$

"Updated"

$$\widehat{\gamma}_n(\varphi_n) = \mathbb{E}\left[\varphi_n(X_n)\prod_{i=1}^{n} G_i(X_i)\right]$$

$$\widehat{\eta}_n = \frac{\mathbb{E}\left[\varphi_n(X_n)\prod_{i=1}^{n} G_i(X_i)\right]}{\mathbb{E}\left[\prod_{i=1}^{n} G_i(X_i)\right]} = \widehat{\gamma}_n(\varphi_n)/\widehat{\gamma}_n(\mathbf{1})$$

Key property:

$$\eta_n(A_n) = \int_{E_1 \times \ldots E_{n-1} \times A_n} \mathbb{Q}_n(dx_{1:n})$$

$$\widehat{\eta}_n(A_n) = \int_{E_1 \times \ldots E_{n-1} \times A_n} \widehat{\mathbb{Q}}_n(dx_{1:n})$$

A Glimpse of the Theory

A Dynamic Systems View:
How do the marginal distributions evolve?

Don’t worry about the details in these slides.

Some Recursive Relationships

- The unnormalized marginals obey:

$$\widehat{\gamma}_n(\varphi_n) = \gamma_n(\varphi_n \cdot G_n) \qquad \gamma_n(\varphi_n) = \widehat{\gamma}_{n-1} M_n(\varphi_n)$$

- Whilst the normalized marginals satisfy:

$$\widehat{\eta}_n(\varphi_n) = \frac{\widehat{\gamma}_n(\varphi_n)}{\widehat{\gamma}_n(\mathbf{1})} = \frac{\gamma_n(\varphi_n \cdot G_n)}{\gamma_n(G_n)} = \frac{\eta_n(\varphi_n \cdot G_n)}{\eta_n(G_n)}$$

$$\eta_n(\varphi_n) = \frac{\gamma_n(\varphi_n)}{\gamma_n(\mathbf{1})} = \frac{\widehat{\gamma}_{n-1} M_n(\varphi_n)}{\widehat{\gamma}_{n-1} M_n(\mathbf{1})} = \frac{\widehat{\eta}_{n-1} M_n(\varphi_n)}{\widehat{\eta}_{n-1} M_n(\mathbf{1})} = \widehat{\eta}_{n-1} M_n(\varphi_n)$$

- So:

$$\widehat{\eta}_n = \frac{\widehat{\eta}_{n-1} M_n(\varphi_n \cdot G_n)}{\widehat{\eta}_{n-1} M_n(G_n)}$$

The Boltzmann-Gibbs Operator

- Given $\nu \in \mathcal{P}(E)$ and $G : E \to \mathbb{R}$:

$$\begin{aligned} \Psi_G : \quad & \mathcal{P}(E) \to \mathcal{P}(E) \\ \Psi_G : \quad & \nu \to \Psi_G(\nu) \end{aligned}$$

- The **Boltzmann-Gibbs** Operator Ψ_G is defined weakly by:

$$\forall \varphi \in \mathcal{C}_b : \qquad \Psi_G(\nu)(\varphi) = \frac{\nu(G \cdot \varphi)}{\nu(G)}$$

- or equivalently, for all measurable sets A:

$$\begin{aligned} \Psi_G(A) &= \frac{\nu(G \cdot \mathbb{I}_A)}{\nu(G)} \\ &= \frac{\int_A \nu(dx) G(x)}{\int_E \nu(dx') G(x')} \end{aligned}$$

Example (Boltzmann-Gibbs Operators and Bayes' Rule)

- Let $\mu(dx) = f(x)\lambda(dx)$ be a prior measure.
- Let $G(x) = g(y|x)$ be the likelihood.
- Then:

$$\Psi_G(\mu)(\varphi) = \frac{\mu(G \cdot \varphi)}{\mu(G)} = \frac{\int \mu(dx) G(x) \varphi(x)}{\int \mu(dx') G(x')}$$

$$= \frac{\int f(x) g(y|x) \varphi(x) \lambda(dx)}{\int f(x') g(y|x') \lambda(dx')}$$

$$= \int f(x|y) \varphi(x) \lambda(dx)$$

with

$$f(x|y) := \frac{f(x) g(y|x)}{\int f(x) g(y|x) \lambda(dx)}$$

- So: $\Psi_{g(y|\cdot)}$: Prior $\rightarrow$ Posterior.

Markov Semigroups

- A **semigroup** $\mathcal{S}$ comprises:
  - A set, S.
  - An associative binary operation, $\cdot$.
- A Markov Chain with homogeneous transition M has **dynamic semigroup** M_n:
  - $M_0(x, A) = \delta_x(A)$.
  - $M_1(x, A) = M(x, A)$.
  - $M_n(x, A) = \int M(x, dy) M_{n-1}(y, A)$.
  - $(M_n \cdot M_m)(x, A) = \int M_n(x, dy) M_m(y, A) = M_{n+m}(x, A)$.
- A *linear* semigroup.
- Key property:

$$\mathbb{P}\left(X_{n+m} \in A | X_m = x\right) = M_n(x, A).$$

Markov Semigroupoids

- A **semigroupoid**, $\mathcal{S}'$ comprises:
  - A set, S.
  - A *partial* associative binary operation, $\cdot$.
- A Markov Chain with inhomogeneous transitions M_n has **dynamic semigroupoid** $M_{p:q}$:
  - $M_{p:p}(x, A) = \delta_x(A)$.
  - $M_{p:p+1}(x, A) = M_{p+1}(x, A)$.
  - $M_{p:q}(x, A) = \int M_{p+1}(x, dy) M_{p+1:q}(y, A)$.
  - $(M_{p:q} \cdot M_{q:r})(x, A) = \int M_{p:q}(x, dy) M_{q:r}(y, A) = M_{p:r}(x, A)$.
- A *linear* semigroupoid.
- Key property:

$$\mathbb{P}(X_{n+m} \in A | X_m = x) = M_{m,n+m}(x, A).$$

An Unnormalized Feynman-Kac Semigroupoid

- We previously established:

$$\gamma_n = \widehat{\gamma}_{n-1} M_n \qquad \widehat{\gamma}_n(\varphi_n) = \gamma_n(\varphi_n \cdot G_n)$$

- Defining

$$Q_p(x_{p-1}, dx_p) = G_{p-1}(x_{p-1}) M_p(x_{p-1}, dx_p)$$

  we obtain $\gamma_n = \gamma_{n-1} Q_n$.
- We can construct the dynamic semigroupoid $Q_{p:q}$:
  - $Q_{p:p}(x, A) = \delta_x(A)$.
  - $Q_{p:p+1}(x, A) = Q_{p+1}(x, A)$.
  - $Q_{p:q}(x, A) = \int Q_{p+1}(x, dy) Q_{p+1:q}(y, A)$.
  - $(Q_{p:q} \cdot Q_{q:r})(x, A) = \int Q_{p:q}(x, dy) Q_{q:r}(y, A) = Q_{p:r}(x, A)$.
- *Just a Markov semigroupoid for general measures:*
  $\forall p \leq q : \gamma_q = \gamma_p Q_{p:q}$.

Warwick Statistics

A Normalised Feynman-Kac Semigroupoid

- We previously established:

$$\eta_n = \widehat{\eta}_{n-1} M_n \qquad \widehat{\eta}_n(\varphi) = \frac{\eta_n(\varphi_n \cdot G_n)}{\eta_n(G_n)}$$

- From the definition of Ψ_{G_n}: $\widehat{\eta}_n = \Psi_{G_n}(\eta_n)$.
- Defining $\Phi_n : \mathcal{P}(E_{n-1}) \to \mathcal{P}(E_n)$ as:

$$\Phi_n : \eta_{n-1} \to \Psi_{G_{n-1}}(\eta_{n-1}) M_n$$

  we have the recursion $\eta_n = \Phi_n(\eta_{n-1})$ and the nonlinear semigroupoid, $\Phi_{p:q}$:
  - $\Phi_{p:p}(x, A) = \delta_x(A)$.
  - $\Phi_{p:p+1}(x, A) = \Phi_{p+1}(x, A)$.
  - $\Phi_{p:q}(x, A) = \Phi_{p+1:q}(\Phi_{p+1}(\eta_p))$ for $q > p + 1$.
  - $(\Phi_{p:q} \cdot \Phi_{q:r})(x, A) = \int \Phi_{q:r}(y, A) \Phi_{p:q}(x, dy) = \Phi_{p:r}(x, A)$.
- Again: $\forall p \leq q : \ \eta_q = \eta_p \Phi_{p:q}$.

McKean Interpretations

Microscopic mass transport.

McKean Interpretations of Feynman-Kac Formulæ

- Families of Markov kernels consistent with FK Marginals.
- A collection $\{K_{n,\eta}\}_{n\in\mathbb{N},\eta\in\mathcal{P}(E_{n-1})}$ is a **McKean Interpretation** if:

$$\forall n \in \mathbb{N}: \ \eta_n = \Phi_n(\eta_{n-1}) = \eta_{n-1} K_{n,\eta_{n-1}}.$$

- Not unique. . . and not linear.
- Selection/Mutation approach seems natural:
  - Choose $S_{n,\eta}$ such that $\eta S_{n,\eta} = \Psi_{G_n}(\eta)$.
  - Set $K_{n+1,\eta} = S_{n,\eta} M_{n+1}$.
- Still not unique:
  - $S_{n,\eta}(x_n, \cdot) = \Psi_{G_n}(\eta)$
  - $S_{n,\eta}(x_n, \cdot) = \epsilon_n G_n(x_n)\delta_{x_n}(\cdot) + (1 - \epsilon_n G_n(x_n))\Psi_{G_n}(\eta)(\cdot)$

Particle Interpretations

Stochastic discretisations.

Particle Interpretations of Feynman-Kac Formulæ I

Given a McKean interpretation, we can attach an N-particle model.

- Denote $\xi_n^{(N)} = (\xi_n^{(N,1)}, \xi_n^{(N,2)}, \ldots, \xi_n^{(N,N)}) \in E_n^N$.
- Allow
$$\left(\Omega^N, \mathcal{F}^N = (\mathcal{F}_n^N)_{n\in\mathbb{N}}, \xi^{(N)}, \mathbb{P}_{\eta_0}^N\right)$$
to indicate a particle-set-valued Markov chain.
- Let $\eta_{n-1}^{(N)} = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{n-1}^{(N,i)}}$.
- Allow the elementary transitions to be:

$$\mathbb{P}\left(\xi_n^{(N)} \in d\xi_n^{(N)} | \xi_{n-1}^{(N)}\right) = \prod_{p=1}^{N} K_{n,\eta_{n-1}^{(N)}}(\xi_{n-1}^{(N,p)}, d\xi_n^{(N,p)})$$

Particle Interpretations of Feynman-Kac Formulæ II

- Consider $K_{n,\eta} = S_{n-1,\eta} M_n$

$$\mathbb{P}\left(\xi_n^{(N)} \in d\xi_n^{(N)} | \xi_{n-1}^{(N)}\right) = \prod_{p=1}^{N} S_{n-1,\eta_{n-1}^{(N)}} M_n(\xi_{n-1}^{(N,p)}, d\xi_n^{(N,p)})$$

- Defining:

$$\mathcal{S}_{n-1}^{(N)}(\xi_{n-1}^{(N)}, d\widehat{\xi}_n^{(N)}) = \prod_{i=1}^{N} S_{n,\eta_{n-1}^{(N)}}(\xi_{n-1}^{(N,p)}, d\widehat{\xi}_{n-1}^{(N,p)})$$

$$\mathcal{M}_n^{(N)}(\widehat{\xi}_{n-1}^{(N)}, d\xi_n^{(N)}) = \prod_{i=1}^{N} M_n(\widehat{\xi}_{n-1}^{(N,p)}, \xi_n^{(N,p)})$$

it is clear that:

$$\mathbb{P}\left(\xi_n^{(N)} \in d\xi_n^{(N)} | \xi_{n-1}^{(N)}\right) = \int_{E_{n-1}^N} \mathcal{S}_{n-1,\eta_{n-1}^{(N)}}(\xi_{n-1}^{(N)}, d\widehat{\xi}_{n-1}^{(N)}) \mathcal{M}_n(\widehat{\xi}_{n-1}^{(N)}, d\xi_n^{(N)})$$

Selection, Mutation and Structure

- A suggestive structural similarity:

$$\begin{array}{ccccc} \eta_{n-1} \in \mathcal{P}(E_{n-1}) & \xrightarrow{S_{n-1,\eta_{n-1}}} & \widehat{\eta}_n \in \mathcal{P}(E_{n-1}) & \xrightarrow{M_n} & \eta_n \in \mathcal{P}(E_n) \\ \xi_{n-1}^{(N)} \in E_{n-1}^N & \xrightarrow{\text{Select}} & \widehat{\xi}_n^{(N)} \in E_{n-1}^N & \xrightarrow{\text{Mutate}} & \xi_n^{(N)} \in E_n^N \end{array}$$

- Selection:

$$S_{n-1,\eta_{n-1}^{(N)}} = \Psi_{G_{n-1}}(\eta_{n-1}^{(N)}) = \sum_{i=1}^{N} \frac{G_{n-1}(\xi_{n-1}^{(N,i)})}{\sum_{j=1}^{N} G_{n-1}(\xi_{n-1}^{(N,j)})} \delta_{\xi_{n-1}^{(N,i)}}$$

$$\widehat{\xi}_{n-1}^{(N,i)} \overset{\text{i.i.d.}}{\sim} \Psi_{G_{n-1}}(\eta_{n-1}^{(N)})$$

- Mutation (conditionally independent):

$$\xi_n^{(N,i)} \sim M_n(\widehat{\xi}_{n-1}^{(N,i)}, d\xi_n^{(N,i)})$$

- Semigroupoid

$$\mathbb{P}^N(\xi_n^{(N)} \in dx_n^{(N)} | \xi_{n-1}^{(N)}) = \prod_{i=1}^{N} \Phi_n(\eta_{n-1}^{(N)})(dx_n^{(N,i)})$$

Selected Results

Local Error Decomposition

$$
\begin{array}{cccccccc}
\eta_1 & \rightarrow & \eta_2 = \Phi_2(\eta_1) & \rightarrow & \eta_3 = \Phi_{1:3}(\eta_1) & \rightarrow & \ldots & \rightarrow \Phi_{1:n}(\eta_1) \\
\Downarrow & & & & & & & \\
\eta_1^N & \rightarrow & \Phi_2(\eta_1^N) & \rightarrow & \Phi_{1:3}(\eta_1^N) & \rightarrow & \ldots & \rightarrow \Phi_{1:n}(\eta_1^N) \\
 & & \Downarrow & & & & & \\
 & & \eta_2^N & \rightarrow & \Phi_3(\eta_2^N) & \rightarrow & \ldots & \rightarrow \Phi_{2:n}(\eta_2^N) \\
 & & & & \Downarrow & & & \\
 & & & & \eta_3^N & \rightarrow & \ldots & \rightarrow \Phi_{3:n}(\eta_3^N) \\
 & & & & & & \vdots & \\
 & & & & & & \Downarrow & \\
 & & & & & & \eta_{n-1}^N & \rightarrow \Phi_n(\eta_{n-1}^N) \\
 & & & & & & & \Downarrow \\
 & & & & & & & \eta_n^N
\end{array}
$$

Formally: $\eta_n^N - \eta_n = \sum_{p=1}^{n} \Phi_{p,n}(\eta_p^N) - \Phi_{p,n}(\Phi_p(\eta_{p-1}^N))$

Some Unbiasesness

Theorem (Del Moral 2004: Theorem 7.4.2)

Under mild regularity conditions, for any $n \in \mathbb{N}$ and some bounded, measurable φ_n:

$$\mathbb{E}\left[\eta^N(\varphi_n)\right] = \gamma_n(\varphi_n).$$

N.B. Typically the corresponding result does not hold for normalised measures:

$$\mathbb{E}\left[\eta_n^N(\varphi_n)\right] \neq \eta_n(\varphi_n).$$

although it does asymptotically with bias $\mathcal{O}(N^{-1})$.

Law of Large Numbers and Weak Convergence

Theorem (Del Moral 2004: Theorem 7.4.4)

Under regularity conditions, for any $n \geq 1$, $p \geq 1$, $\varphi_n \in \mathcal{C}_b(E_n)$:

$$\sqrt{N}\mathbb{E}\left[|\eta_n^N(\varphi_n) - \eta_n(\varphi_n)|^p\right]^{1/p} \leq c_{p,n}||\varphi_n||_\infty$$

By a Borel-Cantelli argument:

$$\lim_{N\to\infty} \eta_n^N(\varphi_n) \overset{a.s.}{\to} \eta_n(\varphi_n).$$

Central Limit Theorem

Proposition (Del Moral 2004: Proposition 9.4.2)

Under regularity conditions, for any $n \geq 1$:

$$\sqrt{N}(\eta_n^N(\varphi_n) - \eta_n(\varphi_n)) \xrightarrow{d} \mathcal{N}\left(0, \sigma_n^2(\varphi_n)\right)$$

where

$$\sigma_n^2(\varphi_n) = \sum_{q=1}^{n} \eta_q \left[(\bar{Q}_{q,n}(\varphi_n - \eta_n(\varphi_n)))^2 \right]$$

where

$$\bar{Q}_{q,n}(\varphi_n)(x_q) = Q_{q,n}(\varphi_n)(x_q) / \eta_q Q_{q,n}(\mathbf{1}).$$

Exact Estimation of Rare Events in Continuous Time

Recent work with James Hodgson
and Murray Pollock

So What's The Difficulty?

A fundamental problem remains. We *cannot* typically sample from the discrete time kernel associated with a stochastic process.

- We **can** exactly sample a broad class of scalar volatility SDEs at finite numbers of time points using exact simulation methods.
- We **can** constrain such paths to arbitrary finite tolerances using ε-strong methods.
- We **cannot** identify the stopping times σ_i even with these methods.
- We **can** identify whether $X_{\sigma_i} \in B_i$ and the value of X_{T_i} for some types of random $T_i \geq \sigma_i$.
- And that's the basis of the following method.

ε-strong Methods

An ε-strong algorithm jointly constructs of X together with a family of processes $\tilde{X}^{\varepsilon}$ indexed by $\varepsilon > 0$ over $[s, t]$ such that:

1. $\sup_{r\in[s,t]} \|X(r) - \tilde{X}^{\varepsilon}(r)\| \overset{a.s.}{\leq} \varepsilon$ for an appropriate norm $\|\cdot\|$;
2. $\tilde{X}^{\varepsilon}$ is piece-wise constant and left-continuous on $[s, t]$ with a.s. finitely many jump;
3. $\tilde{X}^{\varepsilon}$ can be simulated exactly; and
4. Given $\varepsilon_1 > \varepsilon_2 > \cdots > \varepsilon_m > 0$, for $1 \leq \ell_1 < \ell_2 \leq m$ it holds a.s. $\forall r \in [s, t]$ that

   $$\{x : \|\tilde{X}^{\varepsilon_{\ell_2}}(r) - x\| \leq \varepsilon_{\ell_2}\} \subset \{x : \|\tilde{X}^{\varepsilon_{\ell_1}}(r) - x\| \leq \varepsilon_{\ell_1}\},$$

   and moreover it is possible to sample explicitly $\tilde{X}^{\varepsilon_{\ell_2}}$ conditional on $\tilde{X}^{\varepsilon_{\ell_1}}$.

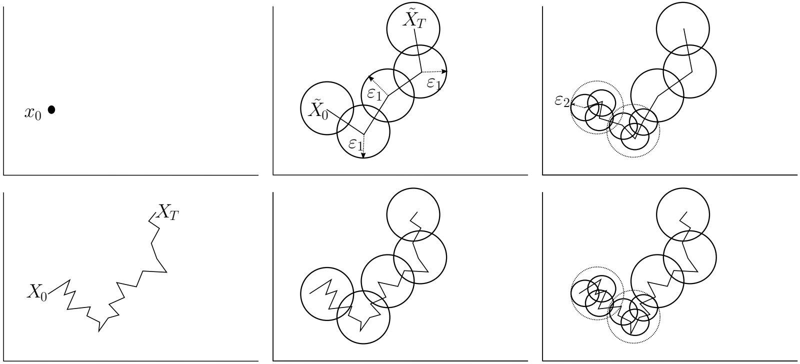


Figure: An schematic illustration of ε-strong simulation. The top row shows shows the ε-strong process $\tilde{X}$ developing as conditional samples are made first with tolerance ε_1, followed with $\varepsilon_2 < \varepsilon_1$. The bottom row shows the fixed target path X, and how the ε-strong constraints relate to it. Pale circles indicate superseded constraints from the previous step.

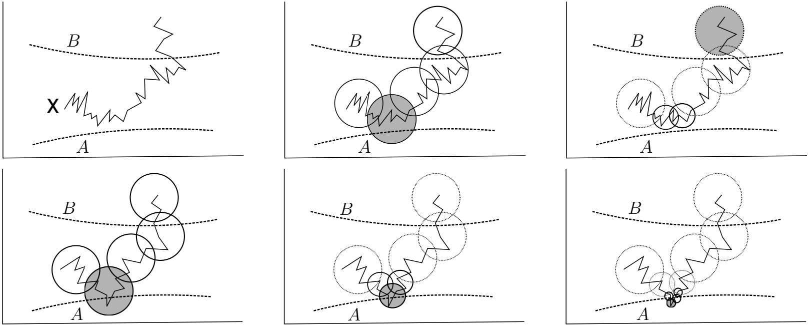


Figure: Determining crossings with ε-strong algorithms.
The first row shows: a realisation of X over a finite time horizon, an initial ε-strong simulation and a refinement which is sufficient to show the process crossing into B.
The second row shows an alternative sample path consistent with the same initial ε-strong simulation, an inconclusive refinement and a further refinement sufficient to conclude that the process has crossed into A.

Warwick Statistics

Halfway There...

- To implement MLS we need to be able to sample from:

$$M_i((\sigma_{i-1}, x_{i-1}), \cdot)$$

  and evaluate

$$G_i(\sigma_i, x_i) = \mathbb{I}_{B_i}(x_i).$$

- We **can** now sample and evaluate $G_i(\sigma_i, x_i)$ *without* knowing σ_i or x_i explicitly.
- We **cannot** sample from M_i.
- We can determine **if** crossings occur, but not **when** or **where**.
- The next insight is that we do not need to.

An Idea

- Splitting relies on sampling at step i several times from the law of the process given the point at which it hit B_{i-1}.
- Traditionally, it does this independently.
- It doesn't have to do it independently.
- Any coupling of the sample paths with the correct marginals would be valid.
- We could use a coupling which makes the simulation problem tractable.

The Underlying Picture

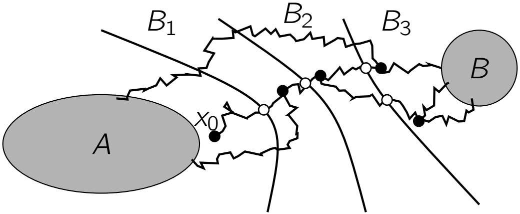


Figure: An illustration of Idealised Splitting with Couplings for a single particle system. The particle begins at the node labelled x_0. Level crossings are indicated by empty nodes, whereas splittings occur at the filled nodes. Between any empty node and the following filled node, the particle trajectories are coupled identically.

The Gory Details I

- Define the bounding random time:

$$\tilde{\sigma}_i = T \cdot \min\{m \in \mathbb{N} : mT \geq \sigma_i\},$$

- Similarly, let $\tilde{\tau}_i$ be the corresponding upper bound on the first hitting time of B_i.
- Rather than split these paths into *independent* copies at times τ_i, from time τ_i until time $\tilde{\tau}_i$, the "split" paths are set to be identically equal, and after this time they evolve conditionally independently given $X_{\tilde{\tau}_i}$.
- For $i = 1, \ldots, m$, let $\tilde{M}_i$ denote the transition kernels for the discrete time quadruple process $V_i = (\sigma_i, \tilde{\sigma}_i, X(\sigma_i), \tilde{X}(\tilde{\sigma}_i))$.
- Define also $\tilde{G}_i(V_i) = \mathbb{I}_{B_i}(X(\sigma_i))$.
- Call the estimator for p resulting from this algorithm $\tilde{p}$.

Algorithm Idealised Splitting with Coupling

Given λ together with $\tilde{G}_i$, $\tilde{M}_i$ for $i = 1, \ldots, m$, an initial number of particles N_0, and splitting ratios $R_1, \ldots, R_{m-1}$:

1. For $j = 1, \ldots, N_0$:
   1.1 Draw $X_1^j(0) \sim \lambda$, $V_1^j \sim \tilde{M}_1\left(\left(0, 0, X_1^j(0), X_1^j(0)\right), \cdot\right)$.
2. Let $S_1 = \{V_1^j : \tilde{G}_1(V_1^j) = 1\}$ be a list of the the surviving paths, and $N_1 = |S_1|$.
3. For $i = 2, \ldots, m$:
   3.1 If $N_{i-1} = 0$, return $\tilde{p} = 0$.
   3.2 Otherwise given $S_{i-1} = \{V_{i-1}^j\}_{j=1}^{N_{i-1}}$, for each $(j, k) \in \{(j', k') : 1 \le j' \le N_{i-1}, 1 \le k' \le R_{i-1}\}$:
      3.2.1 Sample $V_i^{(j,k)} \sim \tilde{M}_i(V_{i-1}^j, \cdot)$
   3.3 Let $S_i = \{V_i^{j,k} : \tilde{G}_i(V_i^{j,k}) = 1\}$, and set $N_i = |S_i|$.
4. Estimate

$$\tilde{p} = \frac{N_m}{N_0 \prod_{i=1}^{m-1} R_i}.$$

Remark

It is not possible to implement this Algorithm as written: we cannot simulate full paths of X, nor make splits at times τ_i. But the construction of MLS with couplings means that an algorithm which simply splits paths at the tractable time $\tilde{\tau}_i$ instead produces identical estimators for p_i.

Proposition (Hodgson, J. & Pollock (in press))

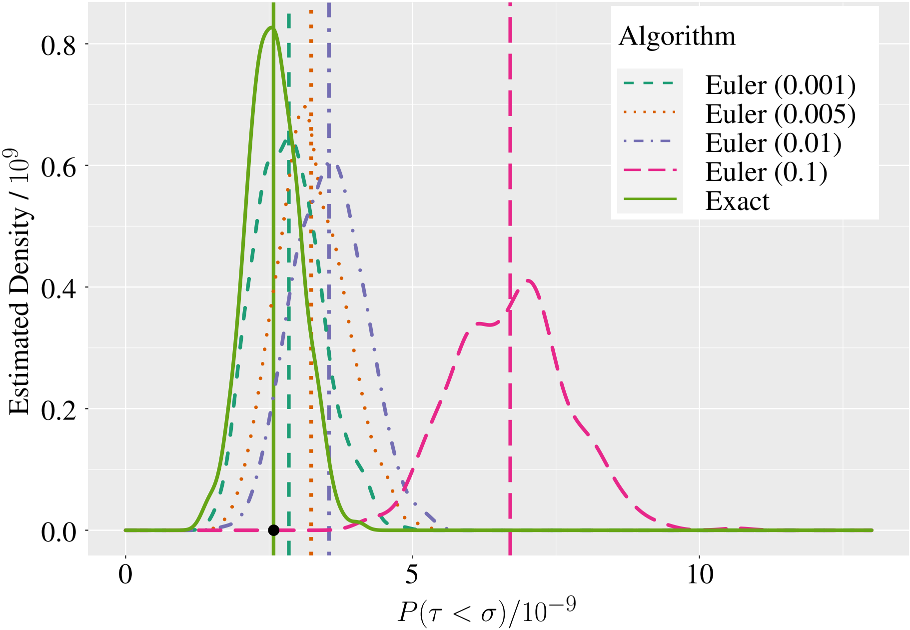
$\tilde{p}$ is an unbiased estimator for p: $\mathbb{E}[\tilde{p}] = p$.

Proof.

Mirrors the proof for standard MLS, using the fact that the *marginal* law of particles between barriers is indentical to that under the simpler scheme. □

A Toy Example: Univariate Brownian Motion

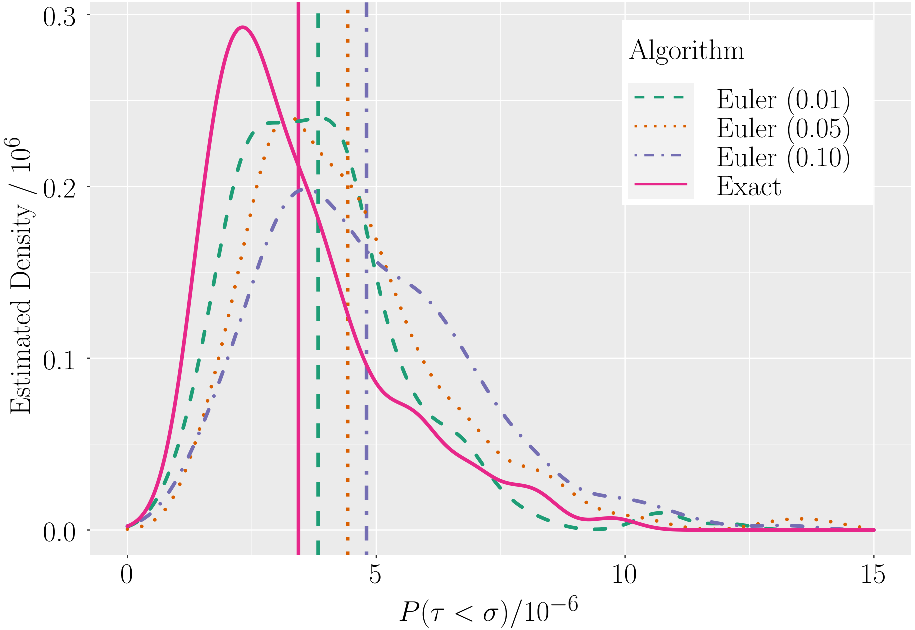
- A setting in which the exact solution is known.
- We choose $A = (-\infty, 0]$, $B = [3^{18}, \infty)$, $B_i = [3^i, \infty)$ for $i = 1, \ldots, 17$, with initial point $x_0 = 1$.
- As is well-known that for real $0 < a < b$, the probability that a Brownian path started at a reaches b before 0 is a/b: $p = 3^{-18} \approx 2.58 \times 10^{-9}$.



Algorithm
Euler (0.001)
Euler (0.005)
Euler (0.01)
Euler (0.1)
Exact
Estimated Density / 10^9
0.0
0.2
0.4
0.6
0.8
0
5
10
$P(\tau < \sigma)/10^{-9}$

Slightly More Challenging: Bivariate Brownian Motion

- The random process is again taken to be Brownian motion initialised at $W_0 = (\frac{1}{2}, \frac{1}{2})$.
- The reaction co-ordinate is chosen to be $\xi(x, y) = \min(x, y)$, and the levels are chosen to be $A = \xi^{-1}((-\infty, 0))$, $B = \xi^{-1}((2^{\frac{21}{2}}, \infty))$, $B_i = \xi^{-1}(2^{\frac{1}{2}(i+1)}, \infty))$ for $i = 1, \ldots, 18$.
- We are not aware of any simple means by which the rare event probability can be analytically obtained in this case.


Algorithm
Euler (0.01)
Euler (0.05)
Euler (0.10)
Exact
Estimated Density / 10^6
$P(\tau < \sigma)/10^{-6}$
0.0
0.1
0.2
0.3
0
5
10
15

Summary

- SMC provides a mechanism for approximating (sequences of) probability distributions via importance sampling and resampling.
- There is still scope to further develop (and understand) SMC methodology.
- There are still unsolved problems in rare event simulation and estimation.
- My own current interests include:
  - Divide-and-conquer approaches to efficient distributed implementation.
  - The interaction with Generalized Bayesian Inference.
  - Automatic optimization of SMC algorithms.

Some References I

- **Some approachable SMC References**

  - Doucet, A. and Johansen, A.M.: A tutorial on particle filtering and smoothing: Fiteen years later. In D. Crisan and B. Rozovsky, editors, *The Oxford Handbook of Nonlinear Filtering*, pages 656–704. 2011. See Preprint.
  - Chopin, N. and Papaspiliopoulos, O.: An introduction to Sequential Monte Carlo. Springer. 2020. See: SpringerLink.

- **A More Technical Treatment of Feynman-Kac Formulæ**

  - Del Moral, P.: Feynman Kac Formulæ: Genealogical and Interacting Particle Systems. Springer, 2004. See: SpringerLink.

- **Multilevel Splitting**

  - Origins: Kahn, H., Harris, T.E.: Estimation of particle transmission by random sampling. National Bureau of Standards Applied Mathematics Series 12, 27–30 (1951). See: `https://dornsifecms.usc.edu/assets/sites/520/docs/kahnharris.pdf`.
  - Unbiasedness: Amrein, M., Künsch, H.: A Variant of Importance Splitting for Rare Event Estimation: Fixed Number of Successes. ACM Transactions on Modelling and Computer Simulation 21(2), Article 13, 20pp. (2011). See: doi:10.1145/1899396.1899401.

Some References II

- **ε-strong Simulation for SDEs:**
  - M. Pollock, A. M. Johansen and G. O. Roberts, On Exact and ε-strong Simulation of (Jump) Diffusions. Bernoulli, 22(2):794–856, 2016. See: 10.3150/14-BEJ676.
- **Exact Rare Event Estimation in Continuous Time:**
  - Hodgson, J., Johansen, A.M. and Pollock, M.: Unbiased simulation of rare events in continuous time. *Methodology and Computing in Applied Probability*, 2021. In press. See: doi:10.1007/s11009-021-09886-2.
- **Some Other Uses of SMC in Rare Event Problems:**
  - Cérou, F., Del Moral, P., Le Gland, F., Lezaud, P.: Genetic genealogical models in rare event analysis. ALEA: Latin American Journal of Probability and Mathematical Statistics 1, 181–203 (2006). See doi:10.1007/s11222-011-9231-6.
  - Del Moral, P., Garnier, J.: Genealogical particle analysis of rare events. Annals of Applied Probability 15(4), 2496–2534 (2005). See doi:10.1214/105051605000000566.
  - Johansen, A., Del Moral, P., Doucet, A.: Sequential Monte Carlo samplers for rare events. In: Proceedings of the 6th International Workshop on Rare Event Simulation, pp. 256–267. Bamberg, Germany (2006). See Preprint.