Rare Event Simulation

and (Interacting) Particle Systems

Adam M. Johansen

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

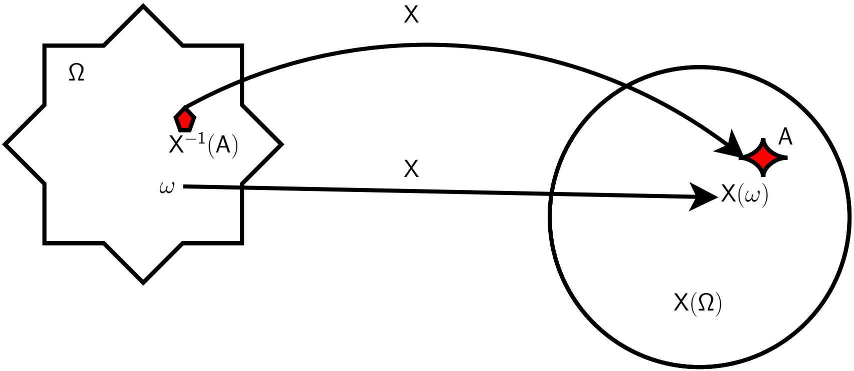
ST912 Statistical Frontiers Seminar
17th February 2017

Mathematics and Statistics
Centre for Doctoral Training

Warwick Statistics

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: Normal Probabilities

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
   $$f(\mathbf{x}) = \frac{1}{\sqrt{|2\pi\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right).$$
   - What is $\mathbb{P}(X \in A)$ if $A = \otimes_{i=1}^{d}[a_i, b_i]$?
   - What can we say about $\mathrm{Law}(X)|_A$

The Monte Carlo Method

- To approximate
$$I(\varphi) = \mathbb{E}[\varphi(X)]$$
with $X \sim \pi$.
- Sample $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \pi$ and use
$$\hat{I}_n(\varphi) = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i).$$
- SLLN:
$$\lim_{n \to \infty} \hat{I}_n(\varphi) \overset{\text{a.s.}}{=} \mathbb{E}[\varphi(X)]$$
- CLT:
$$\lim_{n \to \infty} \sqrt{n}(\hat{I}_n(\varphi) - I(\varphi)) \overset{D}{=} Z,$$
$Z \sim \mathcal{N}(0, \mathrm{Var}(\varphi(X)))$ provided $\mathrm{Var}(\varphi(X)) < \infty$.

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:

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

MASDOC

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 $\sim (n\mathbb{P}(X \in [a,\infty)))^{-1/2}$.

MASDOC

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 $\sim (n\mathbb{P}(X \in [a,\infty)))^{-1/2}$.

MASDOC

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.

Importance Sampling — A Change of Measure View

- If:
  - $X \sim f$
  - $Y \sim g$
  - $f \ll g$
  - $w(x) := \frac{df}{dg}(x)$
- Then:
  - $\mathbb{E}[\varphi(X)] \equiv \mathbb{E}[w(Y)\varphi(Y)]$
- So, if $Y_1, \ldots \overset{\text{iid}}{\sim} g$, then:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} w(Y_i)\varphi(Y_i) \overset{\text{a.s.}}{=} \mathbb{E}[\varphi(X)]$$

and this is an *unbiased* estimator for any n.

MASDOC

Importance Sampling Variance

The variance of this estimator is:

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

MASDOC

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 $E \supset A \supset \operatorname{supp} \varphi(x)$, 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 $g(x) = \exp(-(x-a))\mathbb{I}_{[a,\infty)}(x)$.

MASDOC

A Brief Diversion: *Sequential* Monte Carlo in 1 Slide

- Consider $\pi_p(dx_{1:p}) = \frac{1}{Z_p}\gamma_p(x_{1:p})dx_{1:p}$ for $p = 1, 2, \ldots$.
- Sample $X_1^1, \ldots, X_1^N \overset{\text{iid}}{\sim} M_1$.
- For $p = 2 : n$, sample ciid for $i = 1, \ldots, n$:

$$X_{1:p}^i \sim \frac{\sum_{j=1}^N G_{p-1}(x_{1:p-1}^j)\delta_{x_{1:p-1}^j}(dx_{1:p-1})M_p(x_{1:p-1}, dx_p)}{\sum_{j=1}^N G_{p-1}(x_{1:p-1}^j)}$$

- where $M_p(x_{1:p-1}, dx_p) = m_p(x_{1:p-1}, x_p)dx_p$. and $G_p(x_{1:p}) = \frac{\gamma_p(x_{1:p})}{\gamma_{p-1}(x_{1:p-1})m_p(x_{1:p-1}, x_p)} \propto \frac{\mathrm{d}\pi_p}{\mathrm{d}\pi_{p-1}M_p}(x_{1:p})$.
- Estimators include:

$$\widehat{Z}_p^N = \prod_{q=1}^{p} \frac{1}{N} G_q(X_{1:q}^i) \quad \widehat{I}_{SMC,p}^N(\varphi_p) = \frac{\sum_{i=1}^N G_p(X_{1:p}^i)\varphi(X_{1:p}^i)}{\sum_{i=1}^N G_p(X_{1:p}^i)}$$

There are *many* potential extensions (5).

So far

- Rare events probabilities are important,
- but not often available analytically.
- Monte Carlo appears to offer a solution;
- naïve approaches fail due to their high variance.
- Importance sampling can help dramatically;
- but can be very difficult (i.e. impossible) to implement effectively.
- Sequential Monte Carlo may offer some solutions.

Next

- Two classes of problems;
- potential solutions in these settings.

Problem Formulation

Here we consider algorithms which are applicable to two types of *rare event*, both of which are defined in terms of the canonical Markov chain:

$$\left(\Omega = \prod_{n=0}^{\infty} E_n, \mathcal{F} = \prod_{n=0}^{\infty} \mathcal{F}_n, (X_n)_{n \in \mathbb{N}}, \mathbb{P}_{\eta_0} \right),$$

where the law $\mathbb{P}_{\eta_0}$ is defined by its finite dimensional distributions, for $n \in \mathbb{N}$:

$$\mathbb{P}_{\eta_0} \circ X_{0:n}^{-1}(dx_{0:n}) = \eta_0(dx_0) \prod_{p=1}^{n} M_p(x_{p-1}, dx_p).$$

Static Rare Events

We term the first type of rare events which we consider *static rare events*:

- They are defined as the probability that the first $P+1$ elements of the canonical Markov chain lie in a rare set, $\mathcal{T}$.
- That is, we are interested in

$$\mathbb{P}_{\eta_0}\left(x_{0:P} \in \mathcal{T}\right)$$

and the associated conditional distribution:

$$\mathbb{P}_{\eta_0}\left(x_{0:P} \in dx_{0:P} \left| x_{0:P} \in \mathcal{T}\right.\right)$$

- We assume that the rare event is characterised as a level set of a suitable potential function:

$$V : \mathcal{T} \to [\hat{V}, \infty), \text{ and } V : E_{0:P} \setminus \mathcal{T} \to (-\infty, \hat{V}).$$

MASDOC

A Simple Example: Normal Random Walks

- A toy example:
  - $\eta_0(dx_0) = \mathcal{N}(dx_0; 0, 1)$.
  - $M_n(x_{n-1}, dx_n) = \mathcal{N}(dx_n; x_{n-1}, 1)$.
  - $V(x_{0:P}) = x_P$.
  - $\mathcal{T} = V^{-1}([\hat{V}, \infty))$.
- So:
  - $X_P \sim \mathcal{N}(0, P+1)$
  - $$\begin{aligned}\mathbb{P}(X_{0:P} \in \mathcal{T}) =& \mathbb{P}\left(V(X_{0:P}) \in V(\mathcal{T})\right) \\ =& \mathbb{P}(X_P \geq \hat{V}) = 1 - \Phi(\hat{V}/\sqrt{P+1})\end{aligned}$$

MASDOC

A Slightly Harder Problem

- $\eta_0(x_0) = \mathcal{N}(x_0; 0, 1)$.
- $M_n(x_{n-1}, x_n) = \mathcal{N}(x_n; x_{n-1}, 1)$.
- $V(x_{0:P}) = \max_{0 \leq i \leq P} x_i$.
- $\mathcal{T} = V^{-1}([\hat{V}, \infty))$.

MASDOC

A Real Problem: Polarization Mode Dispersion

This examples was considered by (4, 9). We have a sequence of polarization vectors, r_n which evolve according to the equation:

$$r_n = R(\theta_n, \phi_n) r_{n-1} + \frac{1}{2} S(\theta_n)$$

where:

- $\phi_n \sim \mathcal{U}[-\pi, \pi]$
- $\cos(\theta_n) \sim \mathcal{U}[-1, 1]$, $s_n = \mathrm{sgn}(\theta_n) \sim \mathcal{U}\{-1, +1\}$,
- $S(\theta) = (\cos(\theta), \sin(\theta), 0)$ and $R(\theta, \phi)$ is the matrix which describes a rotation of ϕ about axis $S(\theta)$.

The rare events of interest are system trajectories for which $|r_P| > D$ where D is some threshold.

MASDOC

Feynman-Kac Formulæ and Interacting Particle Systems. . . or SMC

- Want a *black box* method.
- Using only:
  - Samples from η_0 and $M_n(x_{n-1}, \cdot)$
  - Pointwise evaluation of V
- Assume *spatial homogeneity* and that $V : x_{0:P} \mapsto V(x_P)$.
- Recall the core of an SMC algorithm, iteratively, sample:

$$X_n^1, \ldots, X_n^N \overset{\text{iid}}{\sim} \frac{\sum_{i=1}^N G_{n-1}(X_{n-1}^i) M_n(X_{n-1}^i, \cdot)}{\sum_{j=1}^N G_{n-1}(X_{n-1}^j)}.$$

- Provides *unbiased*, *consistent* estimates of:

$$\mathbb{E}_{\eta_0}\left[\prod_{p=0}^{P-1} G(X_p)\right] \qquad \mathbb{E}_{\eta_0}\left[\prod_{p=0}^{P-1} G(X_p)\varphi(X_P)\right]$$

Cf. (5) for more details.

The Approach of Del Moral and Garnier

One SMC approach to the static problem makes use of:

- η_0 and $M_n(x_{n-1}, dx_n)$ from the original model
- $G_n(x_n) = \exp(\beta V(x_n))$

Or the more flexible path space formulation, $\widetilde{E}_n = \otimes_{p=0}^{N} E_p$:

- $\widetilde{\eta}_0 = \eta_0$ as before and

  $$\tilde{M}_n(x_{n-1,0:n-1}, dx_{n,0:n}) = \delta_{x_{n-1,0:n-1}}(dx_{n,0:n-1}) M_n(x_{n-1,n-1} dx_{n,n})$$

- and either

  $$G_n(x_{0:n}) = \exp(\beta V(x_n))$$
  $$\text{or } G_n(x_{0:n}) = \exp\left(\alpha(V(x_n) - V(x_{n-1}))\right).$$

- A *black box* but with parameters.

- Underlying Identity:

$$\begin{aligned}\mathbb{P}(X_P \in V^{-1}([\hat{V},\infty)) =& \mathbb{P}(V(X_P) \geq \hat{V}) \\ =& \mathbb{E}\left[\mathbb{I}_{[\hat{V},\infty)}(V(X_P))\right] \\ =& \mathbb{E}\left[\prod_{n=0}^{p-1} G(X_n) \cdot \mathbb{I}_{[\hat{V},\infty)}(V(X_P)) \prod_{n=0}^{p-1} \frac{1}{G(X_n)}\right].\end{aligned}$$

- Basic Algorithm:
  - Sample $X_0^1, \ldots, X_0^N \overset{\text{iid}}{\sim} \eta_0$. $Y_0^1, \ldots, Y_0^N = 1$.
  - For $p = 1$ to P, for $i = 1$ to N:
    - Compute $z_{p-1} = \frac{1}{N}\sum_{j=1}^N G_{p-1}(X_{p-1}^j)$.
    - Sample $A_p^i \sim \frac{1}{Nz_{p-1}}\sum_{j=1}^N G_{p-1}(X_{p-1}^j)\delta_j(\cdot)$.
    - Sample $X_p^i \sim M_p\left(X_{p-1}^{A_p^i}, \cdot\right)$.
    - Set $Y_p^i = Y_{p-1}^{A_p^i} / G_{p-1}\left(X_{p-1}^{A_{p-1}^i}\right)$.
  - Report:

$$\hat{p} = \prod_{p=0}^{P-1} z_p \cdot \frac{1}{N}\sum_{i=1}^{n} \mathbb{I}_{[\hat{V},\infty)}(X_p^i) Y_p^i.$$

MASDOC

SMC Samplers

Actually, SMC techniques can be used to sample from *any* sequence of distributions...

- Given a sequence of *target* distributions, $\{\pi_n\}$, on measurable spaces $(E_n, \mathcal{E}_n)$
- Construct a synthetic sequence $\{\tilde{\pi}_n\}$ on the product spaces $\bigotimes_{p=0}^{n}(E_p, \mathcal{E}_p)$ by introducing *arbitrary* auxiliary Markov kernels, $L_p : E_{p+1} \otimes \mathcal{E}_p \to [0, 1]$:

$$\tilde{\pi}_n(dx_{1:n}) = \pi_n(dx_n) \prod_{p=0}^{n-1} L_p\left(x_{p+1}, dx_p\right),$$

which each admit one of the target distributions as their final time marginal.

SMC Outline — One Iteration

- Given a sample $\{X_{1:n-1}^{(i)}\}_{i=1}^{N}$ targeting $\tilde{\pi}_{n-1}$, for $i = 1$ to N:
  - sample $X_n^{(i)} \sim K_n(X_{n-1}^{(i)}, \cdot)$,
  - calculate

$$W_n(X_{1:n}^{(i)}) = \frac{\tilde{\pi}_n(X_{1:n}^{(i)})}{\tilde{\pi}_{n-1}(X_{1:n-1})K_n(X_{n-1}^{(i)}, X_n^{(i)})}$$

$$= \frac{\pi_n(X_n^{(i)}) \prod_{p=1}^{n-1} L_p(X_{p+1}^{(i)}, X_p^{(i)})}{\pi_{n-1}(X_{n-1}^{(i)}) \prod_{p=1}^{n-2} L_p(X_{p+1}^{(i)}, X_p^{(i)}) K_n(X_{n-1}^{(i)}, X_n^{(i)})}$$

$$= \frac{\pi_n(X_n^{(i)}) L_{n-1}(X_n^{(i)}, X_{n-1}^{(i)})}{\pi_{n-1}(X_{n-1}^{(i)}) K_n(X_{n-1}^{(i)}, X_n^{(i)})}.$$

- Resample, yielding: $\{X_{1:n}^{(i)}\}_{i=1}^{N}$ targeting $\tilde{\pi}_n$.

MASDOC

Alternative SMC Summary

At each iteration, given a set of weighted samples $\{X_{n-1}^{(i)}, W_{n-1}^{(i)}\}_{i=1}^{N}$ targeting π_{n-1}:

- Sample $X_n^{(i)} \sim K_n(X_{n-1}^{(i)}, \cdot)$.
- $\left\{(X_{n-1}^{(i)}, X_n^{(i)}), W_{n-1}^{(i)}\right\}_{i=1}^{N} \sim \pi_{n-1}(X_{n-1})K_n(X_{n-1}, X_n)$.
- Set weights $W_n^{(i)} = W_{n-1}^{(i)} \frac{\pi_n(X_n)L_{n-1}(X_n, X_{n-1})}{\pi_{n-1}(X_{n-1})K_n(X_{n-1}, X_n)}$.
- $\left\{(X_{n-1}, X_n), W_n^{(i)}\right\}_{i=1}^{N} \sim \pi_n(X_n)L_{n-1}(X_n, X_{n-1})$ and, marginally, $\left\{X_n^{(i)}, W_n^{(i)}\right\}_{i=1}^{(i)} \sim \pi_n$.
- Resample to obtain an unweighted particle set.
- Hints that we'd like $L_{n-1}(x_n, x_{n-1}) = \frac{\pi_{n-1}(x_{n-1})K_n(x_{n-1}, x_n)}{\int \pi_{n-1}(x'_{n-1})K_n(x'_{n-1}, x_n)}$.

Key Points of SMC

- An iterative technique for sampling from a sequence of *similar* distributions.
- By use of intermediate distributions, we can obtain well behaved weighted samples from intractable distributions,
- and estimate associated normalising constants.
- Can be interpreted as a mean field approximation of a Feynman-Kac flow (3).

Static Rare Events: Path-space Approach

- Begin by sampling a set of paths from the law of the Markov chain.
- Iteratively obtain samples from a sequence of distributions which moves "smoothly" towards one which places the majority of its mass on the rare set.
- We construct our sequence of distributions via a potential function and a sequence of *inverse temperatures parameters*:

$$\pi_t(dx_{0:P}) \propto \mathbb{P}_{\eta_0}(dx_{0:P}) g_{t/T}(x_{0:P})$$

$$g_\theta(x_{0:p}) = \left(1 + \exp\left(-\alpha(\theta)\left(V(x_{0:P}) - \hat{V}\right)\right)\right)^{-1}$$

- Estimate the normalising constant of the final distribution and correct via importance sampling.

MASDOC

Path Sampling — An Alternative Approach to Estimating Normalizing Constants

- An integral expression for the log normalising constant of sufficiently regular distributions.
- Given a sequence of densities $p(x|\theta) = q(x|\theta)/z(\theta)$:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log z(\theta) = \mathbb{E}_\theta \left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log q(\cdot|\theta) \right] \qquad (\star)$$

  where the expectation is taken with respect to $p(\cdot|\theta)$.
- Consequently, we obtain:

$$\log \left(\frac{z(1)}{z(0)} \right) = \int_0^1 \mathbb{E}_\theta \left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log q(\cdot|\theta) \right]$$

- See $\star\star$ or (6) for details. Also (9, 13) in SMC context. In our case, we use our particle system to approximate *both* integrals.

Static Rare Events: Framework

Initialization proceeds via importance sampling:

At $t = 0$.
for $i = 1$ to N **do**
  Sample $X_0^{(i)} \sim \nu$ for some importance distribution ν.
  Set $W_0^{(i)} \propto \frac{\pi_0(X_0^{(i)})}{\nu(X_1^{(i)})}$ such that $\sum_{j=1}^{N} W_0^{(j)} = 1$.
end for

Samples are obtained from our sequence of distributions using SMC techniques.

for $t = 1$ to T **do**

  if ESS $<$ threshold **then** resample $\left\{ W_{t-1}^{(i)}, X_{t-1}^{(i)} \right\}_{i=1}^{N}$.

  If desired, apply a Markov kernel, $\tilde{K}_{t-1}$ of invariant distribution π_{t-1} to improve sample diversity.

  for $i = 1$ to N **do**

    Sample $X_t^{(i)} \sim K_t(X_{t-1}^{(i)}, \cdot)$.

    Weight $W_t^{(i)} \propto \hat{W}_{t-1}^{(i)} w_t^{(i)}$ where the *incremental importance weight*, $w_t^{(i)}$ is defined through

$$w_t^{(i)} = \frac{\pi_t(X_t^{(i)}) L_{t-1}(X_t^{(i)}, X_{t-1}^{(i)})}{\pi_{t-1}(X_{t-1}^{(i)}) K_t(X_{t-1}^{(i)}, X_t^{(i)})}, \text{ and } \sum_{j=1}^{N} W_t^{(j)} = 1.$$

  end for

end for

Finally, an estimate can be obtained:

Approximate the path sampling identity to estimate the normalising constant:

$$\hat{Z}_1 = \frac{1}{2}\exp\left[\sum_{t=1}^{T}\left(\alpha(t/T) - \alpha((t-1)/T)\right)\frac{\hat{E}_{t-1}+\hat{E}_t}{2}\right]$$

$$\hat{E}_t = \frac{\sum_{j=1}^{N} W_t^{(j)} \frac{V\left(X_t^{(j)}\right)-\hat{V}}{1+\exp\left(\alpha_t\left(V\left(X_t^{(j)}\right)-\hat{V}\right)\right)}}{\sum_{j=1}^{N} W_t^{(j)}}$$

Estimate the rare event probability using importance sampling:

$$p^\star = \hat{Z}_1 \frac{\sum_{j=1}^{N} W_T^{(j)}\left(1+\exp(\alpha(1)(V\left(X_T^{(j)}\right)-\hat{V}))\right)\mathbb{I}_{(\hat{V},\infty]}\left(V\left(X_T^{(j)}\right)\right)}{\sum_{j=1}^{N} W_T^{(j)}}.$$

MASDOC

Example: Gaussian Random Walk

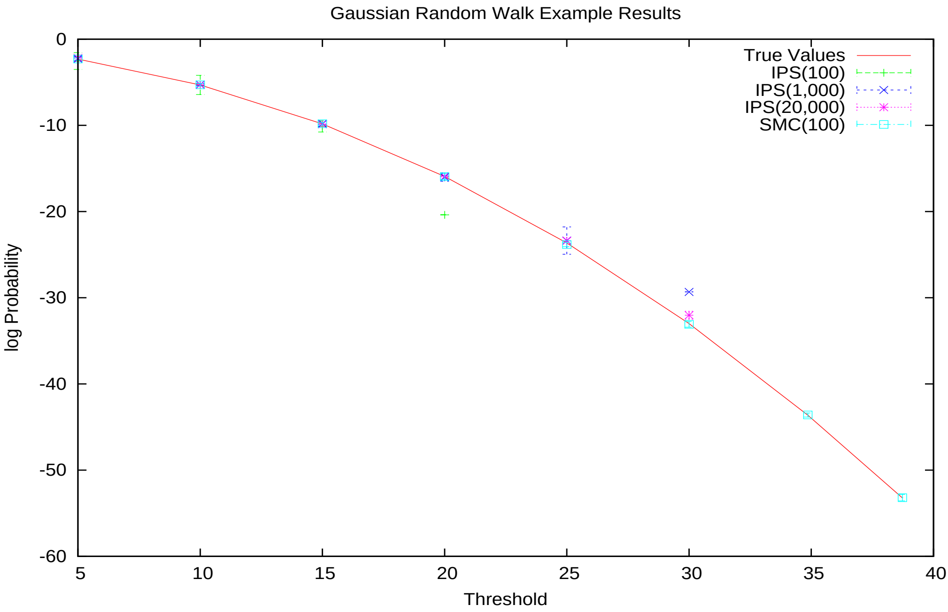
- A toy example: $M_n(R_{n-1}, R_n) = \mathcal{N}(R_n | R_{n-1}, 1)$.
- $V(r_{1:P}) = r_P$
- $\mathcal{T} = V^{-1}([\hat{V}, \infty))$.
- Proposal kernel:

$$K_n(X_{n-1}, dx_n) = \sum_{j=-S}^{S} w_{n+1}(X_{n-1}, X_n) \prod_{i=1}^{P} \delta_{X_{n-1,i}+ij\delta}(dx_{n,i}),$$

  where the weight of individual moves is given by

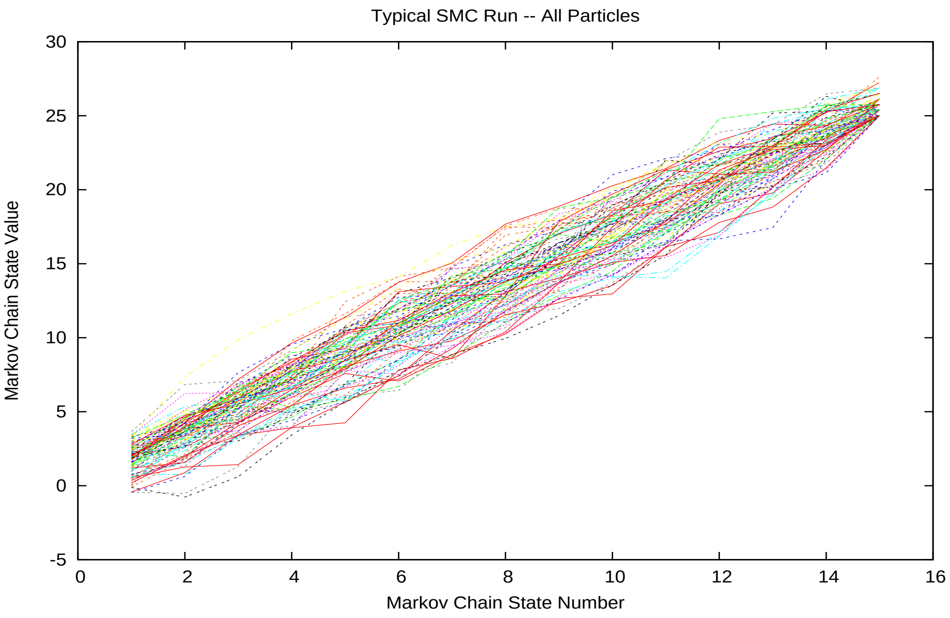
$$w_n(X_{n-1}, X_n) \propto \pi_n(X_n).$$

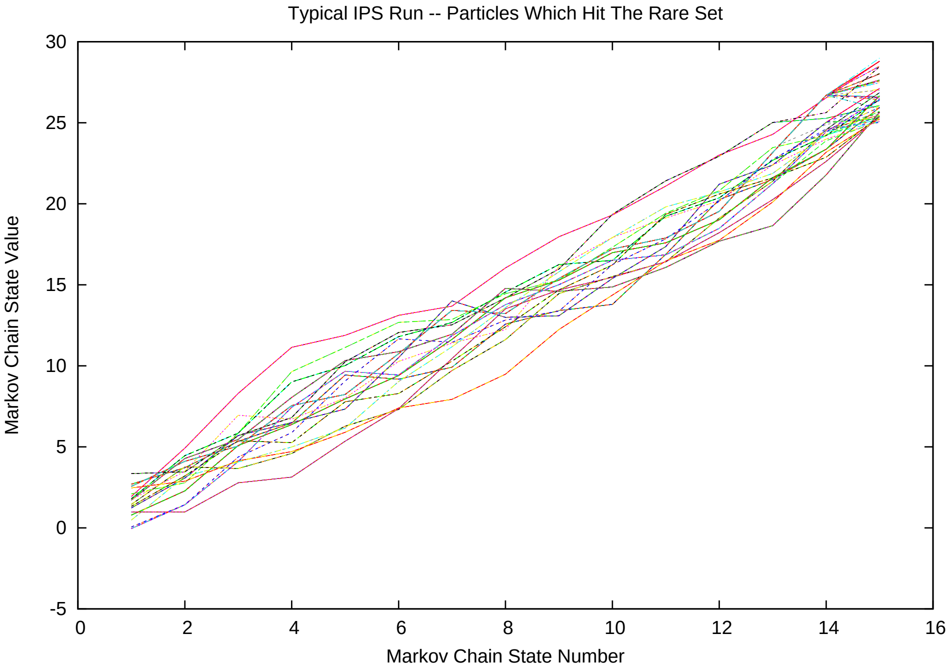
- Linear annealing schedule.
- Number of distributions $T \propto \hat{V}^{3/2}$ (T=2500 when $\hat{V} = 25$).



Gaussian Random Walk Example Results
True Values
IPS(100)
IPS(1,000)
IPS(20,000)
SMC(100)
log Probability
Threshold
0
-10
-20
-30
-40
-50
-60
5
10
15
20
25
30
35
40

MASDOC


Typical SMC Run -- All Particles
Markov Chain State Value
Markov Chain State Number
30
25
20
15
10
5
0
-5
0
2
4
6
8
10
12
14
16


Typical IPS Run -- Particles Which Hit The Rare Set
Markov Chain State Value
Markov Chain State Number
30
25
20
15
10
5
0
-5
0
2
4
6
8
10
12
14
16

Example: Polarization Mode Dispersion

This examples was considered by (4). We have a sequence of polarization vectors, r_n which evolve according to the equation:

$$r_n = R(\theta_n, \phi_n) r_{n-1} + \frac{1}{2}\Omega(\theta_n)$$

where:

- $\phi_n \sim \mathcal{U}[-\pi, \pi]$
- $\cos(\theta_n) \sim \mathcal{U}[-1, 1]$, $s_n = \operatorname{sgn}(\theta_n) \sim \mathcal{U}\{-1, +1\}$,
- $\Omega(\theta) = (\cos(\theta), \sin(\theta), 0)$ and $R(\theta, \phi)$ is the matrix which describes a rotation of ϕ about axis $\Omega(\theta)$.

The rare events of interest are system trajectories for which $|r_P| > D$ where D is some threshold.

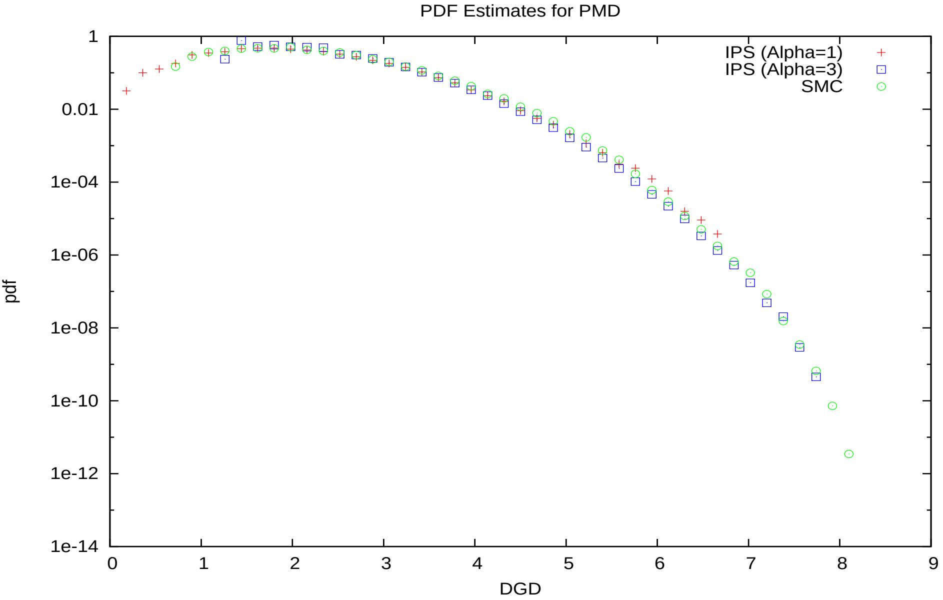
MASDOC

- We use a π_n-invariant MCMC proposal for K_n and the associated time-reversal kernel for L_{n-1}.
- This leads to:

$$W_n(X_{n-1}, X_n) = \frac{\mathrm{d}\pi_n}{\mathrm{d}\pi_{n-1}}(X_{n-1})$$

  allowing sampling and resampling to be exchanged (cf. (8)).
- (Precisely, we employed a Metropolis-Hastings kernel with a proposal which randomly selects two indices uniformly between 1 and n and proposes replacing the ϕ and c values between those two indices with values drawn from the uniform distribution over $[-\pi, \pi] \times [-1, 1]$. This proposal is then accepted with the usual Metropolis acceptance probability.)

Example: Polarization Mode Dispersion - The PDF



MASDOC

Dynamic Rare Events

The other class of rare events in which we are interested are termed *dynamic rare events*:

- They correspond to the probability that a Markov process hits some rare set, $\mathcal{T}$, before its first entrance to some recurrent set $\mathcal{R}$.
- That is, given the stopping time $\tau = \inf \{p : X_p \in \mathcal{T} \cup \mathcal{R}\}$, we seek

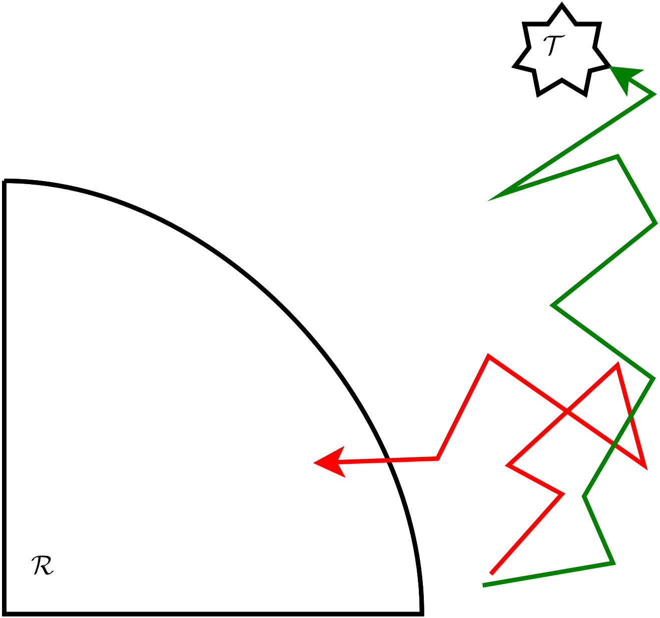
$$\mathbb{P}_{\eta_0}\left(X_\tau \in \mathcal{T}\right)$$

and the associated conditional distribution:

$$\mathbb{P}_{\eta_0}\left(\tau = t, X_{0:t} \in dx_{0:t} | X_\tau \in \mathcal{T}\right)$$

MASDOC

Dynamic Rare Events: Illustration



Importance Sampling

- Use a second Markov process on the same space:

$$\left(\Omega = \prod_{n=0}^{\infty} E_n, \mathcal{F} = \prod_{n=0}^{\infty} \mathcal{F}_n, (X_n)_{n\in\mathbb{N}}, \widetilde{\mathbb{P}}_{\widetilde{\eta}_0} \right),$$

where the law $\widetilde{\mathbb{P}}_{\widetilde{\eta}_0}$ is defined by its finite dimensional distributions:

$$\widetilde{\mathbb{P}}_{\widetilde{\eta}_0} \circ X_{0:N}^{-1}(dx_{0:N}) = \widetilde{\eta}_0(dx_0) \prod_{i=1}^{N} \widetilde{M}_i(x_{i-1}, dx_i).$$

- So that:

$$\frac{d\mathbb{P}_{\eta_0} \circ X_{0:N}}{d\widetilde{\mathbb{P}}_{\widetilde{\eta}_0} \circ X_{0:N}}(x_{0:N}) = \frac{d\eta_0}{d\widetilde{\eta}_0}(x_0) \prod_{p=1}^{N} \frac{dM_n(x_{n-1}, \cdot)}{d\widetilde{M}_n(x_{n-1}, \cdot)}(x_n)$$

Importance Sampling

- Then:

$$
\begin{aligned}
\mathbb{P}_{\eta_0}(X_\tau \in \mathcal{T}) =& \mathbb{E}_{\eta_0}\left[\mathbb{I}_{\mathcal{T}}(X_\tau)\right] = \sum_{t=0}^{\infty} \mathbb{E}_{\eta_0}\left[\mathbb{I}_{\{t\}}(\tau)\mathbb{I}_{\mathcal{T}}(X_t)\right] \\
=& \sum_{t=0}^{\infty} \widetilde{\mathbb{E}}_{\eta_0}\left[\mathbb{I}_{\{t\}}(\tau)\mathbb{I}_{\mathcal{T}}(X_t)\frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}}(X_{0:\infty})\right] \\
=& \sum_{t=0}^{\infty} \widetilde{\mathbb{E}}_{\eta_0}\left[\widetilde{\mathbb{E}}_{\eta_0}\left[\mathbb{I}_{\{t\}}(\tau)\mathbb{I}_{\mathcal{T}}(X_t)\frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}}(X_{0:\infty})\middle|\mathcal{F}_t\right]\right] \\
=& \sum_{t=0}^{\infty} \widetilde{\mathbb{E}}_{\eta_0}\left[\mathbb{I}_{\{t\}}(\tau)\mathbb{I}_{\mathcal{T}}(X_t)\frac{d\mathbb{P}\circ X_{0:t}^{-1}}{d\widetilde{\mathbb{P}}\circ X_{0:t}^{-1}}(X_{0:t})\right] \\
=& \widetilde{\mathbb{E}}_{\eta_0}\left[\mathbb{I}_{\mathcal{T}}(X_\tau)\frac{d\mathbb{P}\circ X_{0:\tau}^{-1}}{d\widetilde{\mathbb{P}}\circ X_{0:\tau}^{-1}}(X_{0:\tau})\right]
\end{aligned}
$$

- But *how* do you choose $\widetilde{\eta}_0$ and $\widetilde{M}_n$?

Splitting

- If V increases towards $\mathcal{T}$, could consider:

$$\mathcal{T}_1 = V^{-1}([V_1,\infty)) \supset \mathcal{T}_2 = V^{-1}([V_2,\infty)) \supset \cdots \supset \mathcal{T}_L = \mathcal{T}$$

- and the non-decreasing first hitting times:

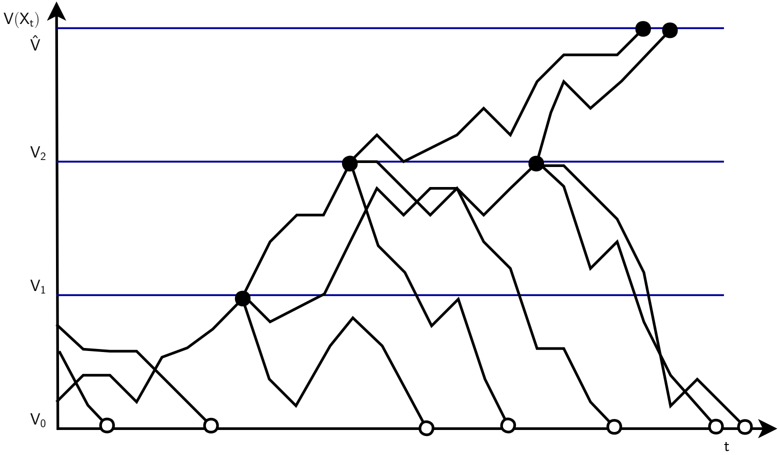
$$\tau_i = \inf\{t : X_t \in \mathcal{R} \cup \mathcal{T}_i\}$$

- which yield the decomposition:

$$\mathbb{P}(X_\tau \in \mathcal{T}) = \mathbb{P}(X_{\tau_1} \in \mathcal{T}_1)\prod_{l=2}^{L}\mathbb{P}(X_{\tau_l} \in \mathcal{T}_l | \{X_{\tau_{l-1}} \in \mathcal{T}_{l-1}\})$$

- and the *multilevel splitting* method [dating back to (10)].

Multilevel Splitting



A Simple Splitting Algorithm

- Sample $X^1_{1:\tau^1_1}, \ldots, X^{N_1}_{1:\tau^1_1}$ iid from $\mathbb{P}$.
- Compute $G(X^i_{1:\tau^i_1}) = \mathbb{I}_{\mathcal{T}_1}(X^i_{\tau^i_1})$. Let $S_1 = \sum_{i=1}^{N_1} G(X^i_{1:\tau^i_1})$.
- For $l = 2$ to L:
  - Set $N_l = rS_{l-1}$.
  - Let $(\hat{X}^i_{\hat{\tau}_{l-1}})_{i=1:N_l}$ comprise r copies of each $X_{\tau^i_{l-1}} \in \mathcal{T}_{l-1}$.
  - For $i = 1, \ldots, N_l$: Sample $X^i_{\hat{\tau}^i_{l-1}:\tau^i_l} \sim \mathbb{P}(\cdot|\{X^i_{\tau_{l-1}} = \hat{X}_{\hat{\tau}^i_{l-1}}\})$.
  - Compute $G(X^i_{\hat{\tau}^i_{l-1}:\tau^i_l}) = \mathbb{I}_{\mathcal{T}_l}(X^i_{\tau^i_l})$. Let $S_l = \sum_{i=1}^{N_1} G(X^i_{1:\tau^i_l})$.
- Compute

$$\mathbb{P}(\widehat{X_\tau \in \mathcal{T}}) = \prod_{l=1}^{L} \frac{S_l}{N_l}.$$

Adaptive Multilevel Splitting 1 — Choosing r_l

- Describes a *Branching Process*.
- If $\mathbb{E}[rS_l] \neq N_l$ essentially:
  - Particle system dies eventually.
  - N_l grows exponentially fast.
- See (7, 11) for some preliminary analysis.
- See (12) for *two-stage* schemes in which a preliminary run specifies r_L.

MASDOC

Adaptive Multilevel Splitting 2 — Choosing the Levels

- Let $\tau_1^i = \inf\{t : X_t \in \mathcal{R}\}$.
- Sample $X^1_{1:\tau_1^i}, \ldots, X^{N_1}_{1:\tau_1^i}$ iid from $\mathbb{P}$.
- Compute $\check{X}_1^i = \max\{X_1^i, \ldots, X_{\tau_1^i}\}$. Set $V_1 = \check{X}_1^{(\lfloor \alpha N_1 \rfloor)}$.
- While $V_l < \hat{V}$. $l \leftarrow l + 1$.
  - Set $N_l = N_1$.
  - Let $(\hat{X}^i_{\hat{\tau}_{l-1}})_{i=1:N_l}$ comprise r copies of each $X_{\tau^i_{l-1}}$ which reached V_{l-1} up to the time when it reached it.
  - For $i = 1, \ldots, N_l$: Sample $X^i_{\hat{\tau}^i_{l-1}:\tau^i_l} \sim \mathbb{P}(\cdot | \{X^i_{\tau_{l-1}} = \hat{X}_{\hat{\tau}^i_{l-1}}\})$.

    With $\tau_l^i = \inf\{t : X_t \in \mathcal{R}\}$.
  - Compute $\check{X}_l^i = \max\{X^i_{\hat{\tau}^i_{l-1}}, \ldots, X_{\tau_l^i}\}$. Set $V_l = \check{X}_l^{(\lfloor \alpha N_l \rfloor)}$.
- Compute

$$\mathbb{P}(\widehat{X_\tau \in \mathcal{T}}) = (1 - \alpha)^{l-1} \cdot \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(\tau^i_{\mathcal{T}} < \tau^i_l)$$

where $\tau^i_{\mathcal{T}} = \inf\{t \in \{\hat{\tau}^i_{l-1}, \ldots, \tau^i_l\} : X^i_t \in \mathcal{T}\}$

Related adaptive methods can be unbiased (1).

Sequential Monte Carlo: *Interacting* Particle Systems

- Sample $X^1_{1:\tau^1_1}, \dots, X^N_{1:\tau^1_1}$ iid from $\mathbb{P}$.
- Compute $G_1(X^i_{\tau^i_1}) = \mathbb{I}_{\mathcal{T}_1}(X^i_{\tau^i_1})$. Let $S_1 = \sum_{i=1}^N G_1(X^i_{\tau^i_1})$.
- For $l = 2$ to L:
  - Sample $(\hat{X}^i_{\hat{\tau}_{l-1}})_{i=1:N_l}$ iid from $\frac{1}{S_{l-1}} \sum_{j=1}^N G_{l-1}(X^j_{\tau_{l-1}})\delta_{X_{\tau^j_{l-1}}}$
  - For $i = 1, \dots, N$: Sample $X^i_{\hat{\tau}^i_{l-1}:\tau^i_l} \sim \mathbb{P}(\cdot|\{X^i_{\tau_{l-1}} = \hat{X}_{\hat{\tau}^i_{l-1}}\})$.
  - Compute $G_l(X^i_{\tau^i_l}) = \mathbb{I}_{\mathcal{T}_l}(X^i_{\tau^i_l})$. Let $S_l = \sum_{i=1}^{N_1} G_l(X^i_{\tau^i_l})$.
- Compute

$$\widehat{\mathbb{P}(X_\tau \in \mathcal{T})} = \prod_{l=1}^{L} \frac{S_l}{N_l}.$$

See, for example (2).

MASDOC

Some References I

[1] C.-E. Bréhier, M. Gazeau, L. Goudenège, T. Lelièvre, and M. Rousset. Unbiasedness of some generalized adaptive multilevel splitting schemes. *Annals of Applied Probability*, 26(6):3559–3601, 2016.

[2] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Statistics and Computing*, 22(3):795–808, 2012.

[3] P. Del Moral. *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Probability and Its Applications. Springer Verlag, New York, 2004.

[4] P. Del Moral and J. Garnier. Genealogical particle analysis of rare events. *Annals of Applied Probability*, 15 (4):2496–2534, November 2005.

[5] A. Doucet and A. M. Johansen. 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. Oxford University Press, 2011.

[6] A. Gelman and X.-L. Meng. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13(2):163–185, 1998.

[7] P. Glasserman, P. Heidelberger, P.Shahabuddin, and T. Zajic. Multilevel splitting for estimating rare event probabilities. *Operations Research*, 47(4):585–600, 1999.

[8] A. M. Johansen and A. Doucet. Auxiliary variable sequential Monte Carlo methods. Research Report 07:09, University of Bristol, Department of Mathematics – Statistics Group, University Walk, Bristol, BS8 1TW, UK, July 2007. A substantially shorter version focusing on particle filtering appeared in Statistics and Probability Letters.

[9] A. M. Johansen, P. Del Moral, and A. Doucet. Sequential Monte Carlo samplers for rare events. In *Proceedings of the 6th International Workshop on Rare Event Simulation*, pages 256–267, Bamberg, Germany, October 2006.

[10] H. Kahn and T. E. Harris. Estimation of particle transmission by random sampling. *National Bureau of Standards Applied Mathematics Series*, 12:27–30, 1951.

[11] A. Lagnoux. Rare event simulation. *Probability in the Engineering and Informational Sciences*, 20(1): 43–66, 2006.

[12] A. Lagnoux. A two-step branching splitting model under cost constraint. *Journal of Applied Probability*, 46 (2):429–452, 2009.

[13] Y. Zhou, A. M. Johansen, and J. A. D. Aston. Towards automatic model comparison: An adaptive sequential Monte Carlo approach. *Journal of Computational and Graphical Statistics*, 2015. In press.

Path Sampling Identity

Given a probability density, $p(x|\theta) = q(x|\theta)/z(\theta)$:

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\log z(\theta) &= \frac{1}{z(\theta)}\frac{\partial}{\partial\theta}z(\theta) \\
&= \frac{1}{z(\theta)}\frac{\partial}{\partial\theta}\int q(x|\theta)dx \\
&= \int \frac{1}{z(\theta)}\frac{\partial}{\partial\theta}q(x|\theta)dx \qquad (\star\star) \\
&= \int \frac{p(x|\theta)}{q(x|\theta)}\frac{\partial}{\partial\theta}q(x|\theta)dx \\
&= \int p(x|\theta)\frac{\partial}{\partial\theta}\log q(x|\theta)dx = \mathbb{E}_{p(\cdot|\theta)}\left[\frac{\partial}{\partial\theta}\log q(\cdot|\theta)\right]
\end{aligned}
$$

wherever $\star\star$ is permissible. Back to $\star$.

MASDOC