Computer Intensive Methods in Mathematical Statistics

Johan Westerborn

Department of mathematics
KTH Royal Institute of Technology
johawes@kth.se

Lecture 5
Sequential Monte Carlo methods I
31 March 2017

Plan of today's lecture

Outline

Last time: variance reduction

- Last time we discussed how to reduce the variance of the standard MC sampler by introducing some auxiliary variables correlating with X. More specifically, we used
  1. a control variate Y such that $\mathbb{E}(Y) = m$ is known and considered
  $$Z = \phi(X) + \alpha(Y - m),$$
  where α was tuned optimally to $\alpha_* = -\mathbb{C}(\phi(X), Y)/\mathbb{V}(Y)$.
  2. antithetic variables V and V' such that $\mathbb{E}(V) = \mathbb{E}(V') = \tau$ and $\mathbb{C}(V, V') < 0$ and considered
  $$W = \frac{V + V'}{2}.$$

Last time: antithetic sampling

- The following theorem turned out to be useful when constructing antithetic variables.

Theorem

Let U be a random variable and let $\varphi : \mathbb{R} \to \mathbb{R}$ be a monotone function. Moreover, assume that there exists a non-increasing transform $T : \mathbb{R} \to \mathbb{R}$ such that $U \stackrel{d.}{=} T(U)$. Then $V = \varphi(U)$ and $V' = \varphi(T(U))$ are identically distributed and

$$\mathbb{C}(V, V') = \mathbb{C}(\varphi(U), \varphi(T(U))) \leq 0.$$

Last time: antithetic sampling (cont.)

- As any distribution function F, and, consequently, its inverse F^{-1} is non-decreasing, the previous result can be naturally applied as follows.
- Letting $\begin{cases} U \sim \mathrm{U}(0,1) \\ T(u) = 1 - u \\ \varphi = F^{-1} \end{cases}$ yields for $\begin{cases} X = F^{-1}(U) \\ X' = F^{-1}(1-U) \end{cases}$,

$$X \stackrel{\mathrm{d.}}{=} X' \text{ (with distribution function } F) \quad \text{and} \quad \mathbb{C}(X, X') \leq 0.$$

- This yields a rather generic way of generating negatively correlated random variables in the case where F^{-1} is known.

Last time: antithetic sampling (cont.)

- As any distribution function F, and, consequently, its inverse F^{-1} is non-decreasing, the previous result can be naturally applied as follows.
- Letting $\begin{cases} U \sim \mathrm{U}(0,1) \\ T(u) = 1 - u \\ \varphi = F^{-1} \end{cases}$ yields for $\begin{cases} X = F^{-1}(U) \\ X' = F^{-1}(1-U) \end{cases}$,

$$X \overset{\mathrm{d.}}{=} X' \text{ (with distribution function } F\text{)} \quad \text{and} \quad \mathbb{C}(X, X') \le 0.$$

- This yields a rather generic way of generating negatively correlated random variables in the case where F^{-1} is known.

Last time: antithetic sampling (cont.)

- As any distribution function F, and, consequently, its inverse F^{-1} is non-decreasing, the previous result can be naturally applied as follows.
- Letting $\begin{cases} U \sim \mathrm{U}(0,1) \\ T(u) = 1 - u \\ \varphi = F^{-1} \end{cases}$ yields for $\begin{cases} X = F^{-1}(U) \\ X' = F^{-1}(1-U) \end{cases}$,

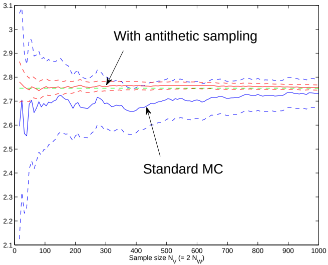
$$X \stackrel{\text{d.}}{=} X' \text{ (with distribution function } F) \quad \text{and} \quad \mathbb{C}(X, X') \leq 0.$$

- This yields a rather generic way of generating negatively correlated random variables in the case where F^{-1} is known.

Last time: antithetic sampling

- As an example we estimated $\tau = \int_0^{\pi/2} \exp(\cos^2(x))\, dx$

with $\begin{cases} X \sim \mathrm{U}(0, \pi/2) \\ V = \frac{\pi}{2} \exp(\cos^2(X)), \\ V' = \frac{\pi}{2} \exp(\sin^2(X)), \\ W = \frac{V + V'}{2}. \end{cases}$



Control variates reconsidered

- A problem with the control variate approach is that the optimal α, i.e.,

$$\alpha_* = -\frac{\mathbb{C}(\phi(X), Y)}{\mathbb{V}(Y)},$$

is generally not known explicitly.

- Thus, it was suggested to
  1. draw $(X^i)_{i=1}^N$,
  2. draw $(Y^i)_{i=1}^N$,
  3. estimate, via MC, α_* using the drawn samples, and
  4. use this to construct optimally $(Z^i)_{i=1}^N$.

  This yields a so-called batch estimator of α_*. However, this procedure is, computationally, somewhat involved.

An online approach to optimal control variates

- The estimators

$$C_N \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \phi(X^i)(Y^i - m) \quad \text{and} \quad V_N \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N}(Y^i - m)^2$$

of $\mathbb{C}(\phi(X), Y)$ and $\mathbb{V}(\phi(X))$, respectively, can however be implemented recursively according to

$$C_{\ell+1} = \frac{\ell}{\ell+1} C_\ell + \frac{1}{\ell+1}\phi(X^{\ell+1})(Y^{\ell+1} - m)$$

$$V_{\ell+1} = \frac{\ell}{\ell+1} V_\ell + \frac{1}{\ell+1}(Y^{\ell+1} - m)^2$$

with $C_0 = V_0 = 0$.

An online approach to optimal control variates (cont.)

- Inspired by this we set for $\ell = 0, 1, 2, \ldots, N-1$,

$$
\begin{aligned}
Z_{\ell+1} &= \phi(X^{\ell+1}) + \alpha_\ell(Y^{\ell+1} - m), \\
\tau_{\ell+1} &= \frac{\ell}{\ell+1}\tau_\ell + \frac{1}{\ell+1}Z_{\ell+1},
\end{aligned}
\qquad (*)
$$

where $\alpha_0 \stackrel{\text{def}}{=} 1$, $\alpha_\ell \stackrel{\text{def}}{=} -C_\ell / V_\ell$ for $\ell > 0$, and $\tau_0 \stackrel{\text{def}}{=} 0$ yielding an online estimator.

An online approach to optimal control variates (cont.)

- One may then establish the following (using martingale convergence results).

Theorem

Let τ_N be obtained through $(*)$. Then, as $N \to \infty$,

(i) $\tau_N \to \tau$ *(a.s.),*

(ii) $\sqrt{N}(\tau_N - \tau) \xrightarrow{d.} \mathrm{N}(0, \sigma_*^2)$,

where $\sigma_*^2 \stackrel{def}{=} \mathbb{V}(\phi(X))\{1 - \rho(\phi(X), Y)^2\}$ *is the optimal variance.*

Example: the tricky integral again

- Last time we estimated

$$\tau = \int_0^{\pi/2} \exp(\cos^2(x))\,dx = \int_0^{\pi/2} \underbrace{\frac{\pi}{2}\exp(\cos^2(x))}_{=\phi(x)} \underbrace{\frac{2}{\pi}}_{=f(x)}\,dx$$
$$= \mathbb{E}_f(\phi(X))$$

using

$$Z = \phi(X) + \alpha^*(Y - m),$$

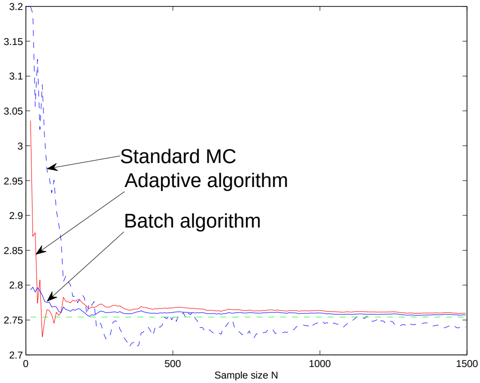
where $Y = \cos^2(X)$ is a control variate with $m = 1/2$.

- However, the optimal coefficient α^* is not known explicitly. We implement the online learning strategy!

Example: the tricky integral again, MATLAB code

```
cos2 = @(x) cos(x).^2;
phi = @(x) (pi/2)*exp(cos2(x));
m = 1/2;
X = (pi/2)*rand;
Y = cos2(X);
c = phi(X)*(Y - m);
v = (Y - m)^2;
tau_CV = phi(X) + (Y - m);
alpha = - c/v;
for k = 2:N,
    X = (pi/2)*rand;
    Y = cos2(X);
    Z = phi(X) + alpha*(Y - m);
    tau_CV = (k - 1)*tau_CV/k + Z/k;
    c = (k - 1)*c/k + phi(X)*(Y - m)/k;
    v = (k - 1)*v/k + (Y - m)^2/k;
    alpha = - c/v;
end
```

Example: the tricky integral again



Outline

Sequential MC problems

- We will now (and for the coming two lectures) extend the principal aim to the problem of estimating sequentially sequences $(\tau_n)_{n \geq 0}$ of expectations

$$\tau_n = \mathbb{E}_{f_n}(\phi(X_{0:n})) = \int_{\mathsf{X}_n} \phi(x_{0:n}) f_n(x_{0:n}) \, dx_{0:n}$$

  over spaces X_n of increasing dimension.

- The densities $(f_n)_{n \geq 0}$ are supposed to be known up to normalizing constants only; i.e., for every $n \geq 0$,

$$f(x_{0:n}) = \frac{z_n(x_{0:n})}{c_n},$$

  where c_n is an unknown constant and z_n is a known positive function on X_n.

Sequential MC problems

- We will now (and for the coming two lectures) extend the principal aim to the problem of estimating sequentially sequences $(\tau_n)_{n \geq 0}$ of expectations

$$\tau_n = \mathbb{E}_{f_n}(\phi(X_{0:n})) = \int_{\mathsf{X}_n} \phi(x_{0:n}) f_n(x_{0:n}) \, dx_{0:n}$$

  over spaces X_n of increasing dimension.
- The densities $(f_n)_{n \geq 0}$ are supposed to be known up to normalizing constants only; i.e., for every $n \geq 0$,

$$f(x_{0:n}) = \frac{z_n(x_{0:n})}{c_n},$$

  where c_n is an unknown constant and z_n is a known positive function on X_n.

Outline

Outline

Prelude: Markov chains

- A Markov chain on $\mathsf{X} \subseteq \mathbb{R}^d$ is a family of random variables (= stochastic process) $(X_k)_{k \geq 0}$ taking values in X such that

$$\mathbb{P}(X_{k+1} \in \mathsf{A} \mid X_0, X_1, \ldots, X_k) = \mathbb{P}(X_{k+1} \in \mathsf{A} \mid X_k)$$

  for all $\mathsf{A} \subseteq \mathsf{X}$. We call the chain time homogeneous if the conditional distribution of X_{k+1} given X_k does not depend on k.
- The distribution of X_{k+1} given $X_k = x$ determines completely the dynamics of the process, and the density q of this distribution is called the transition density of (X_k). Consequently,

$$\mathbb{P}(X_{k+1} \in \mathsf{A} \mid X_k = x_k) = \int_{\mathsf{A}} q(x_{k+1} \mid x_k) \, dx_{k+1}.$$

Prelude: Markov chains

- A Markov chain on $\mathsf{X} \subseteq \mathbb{R}^d$ is a family of random variables (= stochastic process) $(X_k)_{k \geq 0}$ taking values in X such that

$$\mathbb{P}(X_{k+1} \in \mathsf{A} \mid X_0, X_1, \ldots, X_k) = \mathbb{P}(X_{k+1} \in \mathsf{A} \mid X_k)$$

  for all $\mathsf{A} \subseteq \mathsf{X}$. We call the chain time homogeneous if the conditional distribution of X_{k+1} given X_k does not depend on k.
- The distribution of X_{k+1} given $X_k = x$ determines completely the dynamics of the process, and the density q of this distribution is called the transition density of (X_k). Consequently,

$$\mathbb{P}(X_{k+1} \in \mathsf{A} \mid X_k = x_k) = \int_{\mathsf{A}} q(x_{k+1} \mid x_k) \, dx_{k+1}.$$

Markov chains (cont.)

- Let $f_n(x_0, x_1, \ldots, x_n)$ be the joint density of $X_0, X_1, \ldots, X_n$.

Theorem

Let (X_k) be Markov with initial distribution χ and transition density q. Then

(i) $$f_n(x_0, x_1, \ldots, x_n) = \chi(x_0) \prod_{k=0}^{n-1} q(x_{k+1} \mid x_k) \quad (n \geq 0),$$

(ii) $$f_n(x_n \mid x_0) = \int \cdots \int \prod_{k=0}^{n-1} q(x_{k+1} \mid x_k) \, dx_1 \cdots dx_{n-1} \quad (n > 0).$$

- Equation (ii) is referred to as the Chapman-Kolmogorov equation.

Example: The AR(1) process

- As a first example we consider a first order autoregressive process (AR(1)) in $\mathbb{R}$. Set

$$X_0 = 0, \quad X_{k+1} = \alpha X_k + \epsilon_{k+1},$$

where α is a constant and the variables $(\epsilon_k)_{k \geq 1}$ of the noise sequence are i.i.d. with density function f.

- In this case,

$$\mathbb{P}(X_{k+1} \leq x_{k+1} \mid X_k = x_k) = \mathbb{P}(\alpha X_k + \epsilon_{k+1} \leq x_{k+1} \mid X_k = x_k)$$
$$= \mathbb{P}(\epsilon_{k+1} \leq x_{k+1} - \alpha x_k \mid X_k = x_k) = \mathbb{P}(\epsilon_{k+1} \leq x_{k+1} - \alpha x_k),$$

implying that

$$q(x_{k+1} \mid x_k) = \frac{\partial}{\partial x_{k+1}} \mathbb{P}(\epsilon_{k+1} \leq x_{k+1} - \alpha x_k) = f(x_{k+1} - \alpha x_k).$$

Example: The AR(1) process

- As a first example we consider a first order autoregressive process (AR(1)) in $\mathbb{R}$. Set

$$X_0 = 0, \quad X_{k+1} = \alpha X_k + \epsilon_{k+1},$$

where α is a constant and the variables $(\epsilon_k)_{k \geq 1}$ of the noise sequence are i.i.d. with density function f.

- In this case,

$$\mathbb{P}(X_{k+1} \leq x_{k+1} \mid X_k = x_k) = \mathbb{P}(\alpha X_k + \epsilon_{k+1} \leq x_{k+1} \mid X_k = x_k)$$
$$= \mathbb{P}(\epsilon_{k+1} \leq x_{k+1} - \alpha x_k \mid X_k = x_k) = \mathbb{P}(\epsilon_{k+1} \leq x_{k+1} - \alpha x_k),$$

implying that

$$q(x_{k+1} \mid x_k) = \frac{\partial}{\partial x_{k+1}} \mathbb{P}(\epsilon_{k+1} \leq x_{k+1} - \alpha x_k) = f(x_{k+1} - \alpha x_k).$$

Outline

General hidden Markov models (HMMs)

- A hidden Markov model (HMM) comprises
  1. a Markov chain $(X_k)_{k \geq 0}$ with transition density q, i.e.

  $$X_{k+1} \mid X_k = x_k \sim q(x_{k+1} \mid x_k),$$

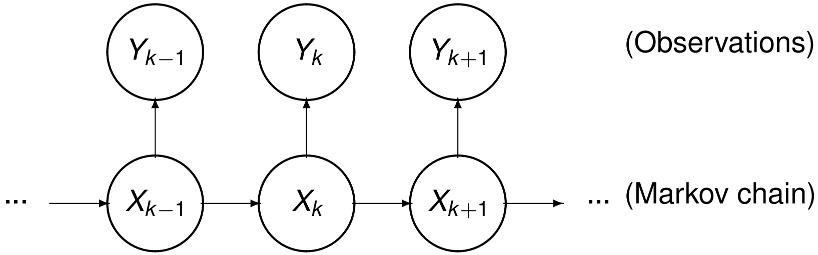
  which is hidden away from us but partially observed through
  2. an observation process $(Y_k)_{k \geq 0}$ such that conditionally on the chain $(X_k)_{k \geq 0}$,
     - (i) the Y_k's are independent with
     - (ii) conditional distribution of each Y_k depending on the corresponding X_k only.
- The density of the conditional distribution $Y_k \mid (X_k)_{k \geq 0} \stackrel{\text{d.}}{=} Y_k \mid X_k$ will be denoted by $p(y_k \mid x_k)$.

General hidden Markov models (HMMs)

- A hidden Markov model (HMM) comprises
  1. a Markov chain $(X_k)_{k \geq 0}$ with transition density q, i.e.
  $$X_{k+1} \mid X_k = x_k \sim q(x_{k+1} \mid x_k),$$
  which is hidden away from us but partially observed through
  2. an observation process $(Y_k)_{k \geq 0}$ such that conditionally on the chain $(X_k)_{k \geq 0}$,
     (i) the Y_k's are independent with
     (ii) conditional distribution of each Y_k depending on the corresponding X_k only.
- The density of the conditional distribution $Y_k \mid (X_k)_{k \geq 0} \stackrel{\text{d.}}{=} Y_k \mid X_k$ will be denoted by $p(y_k \mid x_k)$.

General HMMs (cont.)

- Graphically:



$$
\begin{aligned}
Y_k \mid X_k = x_k &\sim p(y_k \mid x_k) && \text{(Observation density)} \\
X_{k+1} \mid X_k = x_k &\sim q(x_{k+1} \mid x_k) && \text{(Transition density)} \\
X_0 &\sim \chi(x_0) && \text{(Initial distribution)}
\end{aligned}
$$

A brief look at the S&P500 index

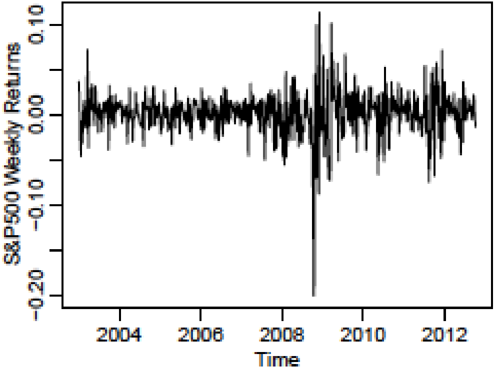


Figure: Weekly log-returns of S&P500 from January 2, 2003 to September 28, 2012.

Example HMM: stochastic volatility

- The following dynamical system is used in financial economy (see e.g. Jacuquier *et al.*, 1994). Let

$$\begin{cases} X_{k+1} = \alpha X_k + \sigma \epsilon_{k+1}, \\ Y_k = \beta \exp\left(\frac{X_k}{2}\right) \varepsilon_k, \end{cases}$$

where $\alpha \in (0,1)$, $\sigma > 0$, and $\beta > 0$ are constants and $(\epsilon_k)_{k \geq 1}$ and $(\varepsilon_k)_{k \geq 0}$ are sequences of i.i.d. standard normal-distributed noise variables.

Example HMM: stochastic volatility

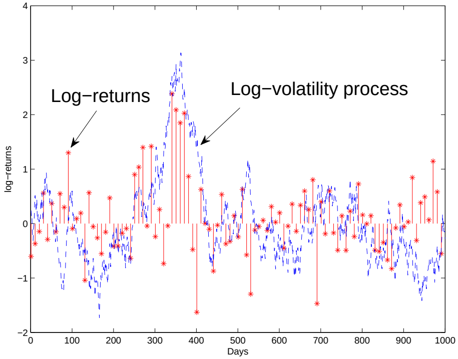
- In this model,
  - the values of the observation process (Y_k) are observed daily log-returns and
  - the hidden chain (X_k) is the unobserved log-volatility (modeled by a stationary AR(1) process).
- The strength of the model is that it allows for volatility clustering.

Example HMM: stochastic volatility

- In this model,
  - the values of the observation process (Y_k) are observed daily log-returns and
  - the hidden chain (X_k) is the unobserved log-volatility (modeled by a stationary AR(1) process).
- The strength of the model is that it allows for volatility clustering.

Example HMM: stochastic volatility (cont.)

- A typical realization of the the model looks like follows (here $\alpha = 0.975$, $\sigma = 0.16$, and $\beta = 0.63$).



Smoothing of hidden states

- When operating on HMMs, one is most often interested in the smoothing distribution $f_n(x_{0:n} \mid y_{0:n})$, i.e. the conditional distribution of a set $X_{0:n}$ of hidden states given $Y_{0:n} = y_{0:n}$.

Theorem (smoothing distribution)

$$f_n(x_{0:n} \mid y_{0:n}) = \frac{\chi(x_0)p(y_0 \mid x_0)\prod_{k=1}^{n} p(y_k \mid x_k)q(x_k \mid x_{k-1})}{L_n(y_{0:n})},$$

where $L_n(y_{0:n})$ is the likelihood function given by

$$L_n(y_{0:n}) = \int \chi(x_0)p(y_0 \mid x_0)\prod_{k=1}^{n} p(y_k \mid x_k)q(x_k \mid x_{k-1})\, dx_{0:n}.$$

Estimation of smoothed expectations

- Being a high-dimensional (say $n \approx 10{,}000$) integral over complicated integrands, $L_n(y_{0:n})$ is in general unknown.
- However by writing

$$\tau_n = \mathbb{E}(\phi(X_{0:n}) \mid Y_{0:n} = y_{0:n}) = \int \phi(x_{0:n}) f_n(x_{0:n} \mid y_{0:n})\, dx_{0:n}$$
$$= \int \phi(x_{0:n}) \frac{z_n(x_{0:n})}{c_n}\, dx_{0:n},$$

with

$$\begin{cases} z_n(x_{0:n}) = \chi(x_0) p(y_0 \mid x_0) \prod_{k=1}^{n} p(y_k \mid x_k) q(x_k \mid x_{k-1}), \\ c_n = L_n(y_{0:n}), \end{cases}$$

we may cast the problem of computing τ_n into the framework of SMC problems.

Estimation of smoothed expectations

- Being a high-dimensional (say $n \approx 10{,}000$) integral over complicated integrands, $L_n(y_{0:n})$ is in general unknown.
- However by writing

$$\begin{aligned}\tau_n = \mathbb{E}(\phi(X_{0:n}) \mid Y_{0:n} = y_{0:n}) &= \int \phi(x_{0:n}) f_n(x_{0:n} \mid y_{0:n})\, dx_{0:n} \\ &= \int \phi(x_{0:n}) \frac{z_n(x_{0:n})}{c_n}\, dx_{0:n},\end{aligned}$$

with

$$\begin{cases} z_n(x_{0:n}) = \chi(x_0) p(y_0 \mid x_0) \prod_{k=1}^n p(y_k \mid x_k) q(x_k \mid x_{k-1}), \\ c_n = L_n(y_{0:n}), \end{cases}$$

we may cast the problem of computing τ_n into the framework of SMC problems.

Estimation of smoothed expectations

- In particular we would like to update the approximation sequentially in n, i.e. online, as new data (Y_k) become available.
- Of particular interest is the filter distribution, which is the marginal of the smoothing distribution with respect to the current state X_n:

$$\tau_n = \mathbb{E}(\phi(X_n) \mid Y_{0:n} = y_{0:n}) = \int \phi(x_n) f_n(x_{0:n} \mid y_{0:n}) \, dx_{0:n}.$$

- Computing the smoothing/filtering distributions is essential when calibrating the model parameters (inference) as well as using the model for prediction.

Estimation of smoothed expectations

- In particular we would like to update the approximation sequentially in n, i.e. online, as new data (Y_k) become available.
- Of particular interest is the filter distribution, which is the marginal of the smoothing distribution with respect to the current state X_n:

$$\tau_n = \mathbb{E}(\phi(X_n) \mid Y_{0:n} = y_{0:n}) = \int \phi(x_n) f_n(x_{0:n} \mid y_{0:n}) \, dx_{0:n}.$$

- Computing the smoothing/filtering distributions is essential when calibrating the model parameters (inference) as well as using the model for prediction.

Outline

Simulation of rare events for Markov chains

- Let (X_k) be a Markov chain on $\mathsf{X} = \mathbb{R}$ and consider some rectangle $\mathsf{A} = \mathsf{A}_0 \times \mathsf{A}_1 \times \cdots \mathsf{A}_n \subseteq \mathbb{R}^n$, where $\mathsf{A}_\ell = (a_\ell, b_\ell)$. Here A can be a possibly rare event.
- Here the unknown probability $c_n = \mathbb{P}(X_{0:n} \in \mathsf{A})$ of the rare event A is often the quantity of interest.
- Let $f_{n|\mathsf{A}}$ be the conditional density of the states $X_{0:n} = (X_0, X_2, \ldots, X_n)$ given $X_{0:n} \in \mathsf{A}$ and consider

$$\tau_n = \mathbb{E}_{f_n}(\phi(X_{0:n}) \mid X_{0:n} \in \mathsf{A}) = \mathbb{E}_{f_{n|\mathsf{A}}}(\phi(X_{0:n}))$$

$$= \int_{\mathsf{A}} \phi(x_{0:n}) \underbrace{\frac{f(x_{0:n})}{\mathbb{P}(X_{0:n} \in \mathsf{A})}}_{= f_{n|\mathsf{A}}(x_{0:n}) = z_n(x_{0:n})/c_n} \, dx_{0:n}.$$

Simulation of rare events for Markov chains

- Let (X_k) be a Markov chain on $\mathsf{X} = \mathbb{R}$ and consider some rectangle $\mathsf{A} = \mathsf{A}_0 \times \mathsf{A}_1 \times \cdots \mathsf{A}_n \subseteq \mathbb{R}^n$, where $\mathsf{A}_\ell = (a_\ell, b_\ell)$. Here A can be a possibly rare event.
- Here the unknown probability $c_n = \mathbb{P}(X_{0:n} \in \mathsf{A})$ of the rare event A is often the quantity of interest.
- Let $f_{n|\mathsf{A}}$ be the conditional density of the states $X_{0:n} = (X_0, X_2, \ldots, X_n)$ given $X_{0:n} \in \mathsf{A}$ and consider

$$\begin{aligned}\tau_n = \mathbb{E}_{f_n}(\phi(X_{0:n}) \mid X_{0:n} \in \mathsf{A}) = \mathbb{E}_{f_{n|\mathsf{A}}}(\phi(X_{0:n}))\\ = \int_{\mathsf{A}} \phi(x_{0:n}) \underbrace{\frac{f(x_{0:n})}{\mathbb{P}(X_{0:n} \in \mathsf{A})}}_{= f_{n|\mathsf{A}}(x_{0:n}) = z_n(x_{0:n})/c_n} \, dx_{0:n}.\end{aligned}$$

Simulation of rare events for Markov chains (cont.)

- As
$$c_n = \mathbb{P}(X_{0:n} \in \mathsf{A}) = \int \mathbb{1}_{\mathsf{A}}(x_{0:n}) f(x_{0:n})\, dx_{0:n}$$
a first—naive—approach could of course be to use standard MC and simply
  1. simulate the Markov chain N times, yielding $(X^i_{0:n})_{i=1}^N$,
  2. count the number N_{A} of trajectories that fall into A, and
  3. estimate c_n using the standard MC estimator
$$c_n^N = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\mathsf{A}}(X^i_{0:n}) = \frac{N_{\mathsf{A}}}{N}.$$
- Problem: if $c_n = 10^{-9}$ we may expect to produce a billion draws before obtaining a single draw belonging to A ☹.
- SMC methods save the day!

Simulation of rare events for Markov chains (cont.)

- As

$$c_n = \mathbb{P}(X_{0:n} \in \mathsf{A}) = \int \mathbb{1}_{\mathsf{A}}(x_{0:n}) f(x_{0:n}) \, dx_{0:n}$$

a first—naive—approach could of course be to use standard MC and simply

1. simulate the Markov chain N times, yielding $(X_{0:n}^i)_{i=1}^N$,
2. count the number N_{A} of trajectories that fall into A, and
3. estimate c_n using the standard MC estimator

$$c_n^N = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\mathsf{A}}(X_{0:n}^i) = \frac{N_{\mathsf{A}}}{N}.$$

- Problem: if $c_n = 10^{-9}$ we may expect to produce a billion draws before obtaining a single draw belonging to A ☹.
- SMC methods save the day!

Outline

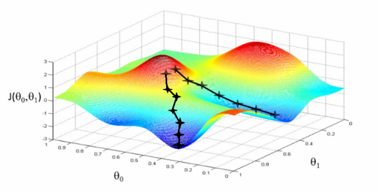
Generalized SMC problems

- Interestingly, it is generally not required that the spaces (X_n) are of increasing dimension.
- Indeed, a sequence of arbitrary densities $f_n^*(x_n)$, defined on arbitrary spaces E_n and known up to normalizing constants, can typically be extended to a sequence of densities

$$f_n(x_{1:n}) = f_n^*(x_n) \prod_{k=1}^{n-1} r_k(x_k \mid x_{k+1}), \quad n > 0,$$

defined on the augmented spaces $\mathsf{X}_n = \mathsf{E}_1 \times \cdots \times \mathsf{E}_n$ via auxiliary Markov transition densities (r_k).
- In this construction, $f_n^*(x_n)$ is the marginal of $f_n(x_{1:n})$ w.r.t. x_n.

Example: global maximisation



- When finding the global maximum of $f(x)$ over some space E, consider the Boltzmann distributions

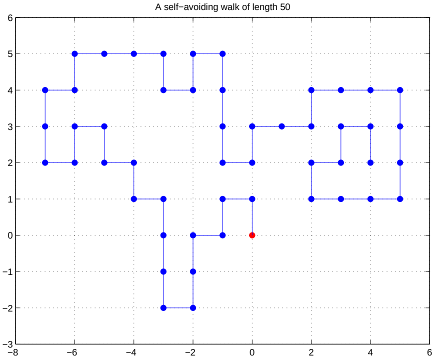
$$f_n^*(x_n) = \frac{1}{c_n} e^{f(x_n)/T_n}$$

on $E_n = E$, where the 'temperatures' (T_n) vanish with n.
- Optimization of $f(x)$ can now be performed by sampling from the sequence $(f_n^*(x_n))$.

Outline

Self-avoiding walks (SAWs)

- Let $S_n \stackrel{\text{def}}{=} \{x_{0:n} \in (\mathbb{Z}^2)^n : x_0 = \mathbf{0}, |x_{k+1} - x_k| = 1, x_k \neq x_\ell, \forall 0 \leq \ell < k \leq n\}$ be the set of n-step self-avoiding walks in $\mathbb{Z}^2$.



Application of SAWs

- In addition, let

$$c_n = |S_n| = \text{The number of possible SAWs of length } n.$$

- SAWs are used in, e.g.,
  - polymer science for describing long chain polymers, with the self-avoidance condition modeling the excluded volume effect.
  - statistical mechanics and the theory of critical phenomena in equilibrium.
- However, computing c_n (and in analyzing how c_n depends on n) is known to be a very challenging (NP-hard) combinatoric problem!

Application of SAWs

- In addition, let

$$c_n = |S_n| = \text{The number of possible SAWs of length } n.$$

- SAWs are used in, e.g.,
  - polymer science for describing long chain polymers, with the self-avoidance condition modeling the excluded volume effect.
  - statistical mechanics and the theory of critical phenomena in equilibrium.
- However, computing c_n (and in analyzing how c_n depends on n) is known to be a very challenging (NP-hard) combinatoric problem!

Application of SAWs

- In addition, let

$$c_n = |S_n| = \text{The number of possible SAWs of length } n.$$

- SAWs are used in, e.g.,
  - polymer science for describing long chain polymers, with the self-avoidance condition modeling the excluded volume effect.
  - statistical mechanics and the theory of critical phenomena in equilibrium.
- However, computing c_n (and in analyzing how c_n depends on n) is known to be a very challenging (NP-hard) combinatoric problem!

An MC approach to SAWs

- Diabolic trick: let $f_n(x_{0:n})$ be the uniform distribution on S_n:

$$f_n(x_{0:n}) = \frac{1}{c_n} \underbrace{\mathbb{1}_{\mathsf{S}_n}(x_{0:n})}_{=z(x_{0:n})}, \quad x_{0:n} \in (\mathbb{Z}^2)^n.$$

- We may now cast the problem of computing the number c_n (= the normalizing constant of f_n) into the framework of SMC problems.
- In addition, solving this problem for $n = 1, 2, 3, \ldots, 508, 509, \ldots$ calls for sequential implementation of IS.

An MC approach to SAWs

- Diabolic trick: let $f_n(x_{0:n})$ be the uniform distribution on S_n:

$$f_n(x_{0:n}) = \frac{1}{c_n} \underbrace{\mathbb{1}_{S_n}(x_{0:n})}_{=z(x_{0:n})}, \quad x_{0:n} \in (\mathbb{Z}^2)^n.$$

- We may now cast the problem of computing the number c_n (= the normalizing constant of f_n) into the framework of SMC problems.
- In addition, solving this problem for $n = 1, 2, 3, \ldots, 508, 509, \ldots$ calls for sequential implementation of IS.

An MC approach to SAWs

- Diabolic trick: let $f_n(x_{0:n})$ be the uniform distribution on S_n:

$$f_n(x_{0:n}) = \frac{1}{c_n} \underbrace{\mathbb{1}_{\mathsf{S}_n}(x_{0:n})}_{=z(x_{0:n})}, \quad x_{0:n} \in (\mathbb{Z}^2)^n.$$

- We may now cast the problem of computing the number c_n (= the normalizing constant of f_n) into the framework of SMC problems.
- In addition, solving this problem for $n = 1, 2, 3, \ldots, 508, 509, \ldots$ calls for sequential implementation of IS.

Outline

Next week

- The coming two lectures will be devoted completely to SMC methods.
- The last of these two lectures launches HA1.
- Next week, E2 deals with
  - asymptotic properties of importance sampling estimators and
  - antithetic sampling.