Monte Carlo and Empirical Methods for Stochastic Inference (MASM11/FMS091)

Magnus Wiktorsson
Centre for Mathematical Sciences
Lund University, Sweden

Lecture 5
Sequential Monte Carlo methods I
February 3, 2015
Plan of today’s lecture

1. Variance reduction reconsidered

2. Sequential MC problems

3. 3 Examples of SMC problems
   - Prelude: Markov chains
   - Example 1: Simulation of extreme events
   - Example 2: Estimation in general HMMs
   - Example 3: Estimation of SAWs
1. Variance reduction reconsidered

2. Sequential MC problems

3. Examples of SMC problems
   - Prelude: Markov chains
   - Example 1: Simulation of extreme events
   - Example 2: Estimation in general HMMs
   - Example 3: Estimation of SAWs
Last time we discussed how to reduce the variance of the standard MC sampler by introducing correlation between the variables of the sample. More specifically, we used

1. a control variate $Y$ such that $E(Y) = m$ is known:

$$Z = \phi(X) + \beta(Y - m),$$

where $\beta$ was tuned optimally to $\beta^* = -C(\phi(X), Y)/V(Y)$.

2. antithetic variables $V$ and $V'$ such that $E(V) = E(V') = \tau$ and $C(V, V') < 0$:

$$W = \frac{V + V'}{2}.$$
Last time: Variance reduction

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

**Theorem**

Let \( V = \varphi(U) \), where \( \varphi : \mathbb{R} \rightarrow \mathbb{R} \) is a monotone function. Moreover, assume that there exists a non-increasing transform \( T : \mathbb{R} \rightarrow \mathbb{R} \) such that \( U \overset{d.}{=} T(U) \). Then \( V = \varphi(U) \) and \( V' = \varphi(T(U)) \) are identically distributed and

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

An important application of this theorem is the following: Let \( F \) be a distribution function. Then, letting \( U \sim \mathcal{U}(0, 1) \), \( T(u) = 1 - u \), and \( \varphi = F^{-1} \) yields, for \( X = F^{-1}(U) \) and \( X' = F^{-1}(1 - U) \),

\[
X \overset{d.}{=} X' \text{ (with distribution function } F) \quad \text{and} \quad \mathbb{C}(X, X') \leq 0.
\]
Variance reduction reconsidered
Sequential MC problems
3 Examples of SMC problems

Last time: Variance reduction

\[ \tau = 2 \int_0^{\pi/2} \exp(\cos^2(x)) \, dx, \]

\[ \begin{align*}
V &= 2 \frac{\pi}{2} \exp(\cos^2(X)), \\
V' &= 2 \frac{\pi}{2} \exp(\sin^2(X)), \\
W &= \frac{V + V'}{2}.
\end{align*} \]
A problem with the control variate approach is that the optimal $\beta$, i.e.

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

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, $\beta^*$ using the drawn samples, and
4. use this to optimally construct $(Z^i)_{i=1}^N$.

This yields a so-called batch estimator of $\beta^*$. However, this procedure is computationally somewhat complex.
The estimators

\[ C_N \overset{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \phi(X_i)(Y^i - m) \]

\[ V_N \overset{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} (Y^i - m)^2 \]

of \(C(\phi(X), Y)\) and \(V(Y)\), respectively, can be implemented recursively according to

\[ C_{\ell+1} = \frac{\ell}{\ell + 1} C_\ell + \frac{1}{\ell + 1} \phi(X_{\ell+1})(Y^{\ell+1} - m) \]

and

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

with \(C_0 = V_0 = 0\).
An online approach to optimal control variates (cont.)

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

$$Z_{\ell+1} = \phi(X_{\ell+1}) + \beta_{\ell}(Y^{\ell+1} - m),$$

$$\tau_{\ell+1} = \frac{\ell}{\ell + 1} \tau_{\ell} + \frac{1}{\ell + 1} Z_{\ell+1},$$

where $\beta_0 \overset{\text{def}}{=} 1$, $\beta_{\ell} \overset{\text{def}}{=} C_{\ell}/V_{\ell}$ for $\ell > 0$, and $\tau_0 \overset{\text{def}}{=} 0$ yielding an online estimator. One may then establish the following convergence results.

**Theorem**

Let $\tau_N$ be obtained through $(\ast)$. Then, as $N \to \infty$,

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

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

where $\sigma_*^2 \overset{\text{def}}{=} \nabla(\phi(X))\{1 - \rho(\phi(X), Y)^2\}$ is the optimal variance.
Example: the tricky integral again

We estimate

\[ \tau = \int_{-\pi/2}^{\pi/2} \exp(\cos^2(x)) \, dx \overset{\text{sym}}{=} 2 \int_0^{\pi/2} \frac{\pi}{2} \exp(\cos^2(x)) \left( \frac{2}{\pi} \right) dx = \phi(x) + f(x) = \mathbb{E}_f(\phi(X)) \]

using

\[ Z = \phi(X) + \beta^*(Y - m), \]

where \( Y = \cos^2(X) \) is a control variate with

\[ m = \mathbb{E}(Y) = \int_0^{\pi/2} \cos^2(x) \frac{2}{\pi} \, dx = \{\text{use integration by parts}\} = \frac{1}{2}. \]

However, the optimal coefficient \( \beta^* \) is not known explicitly.
Example: the tricky integral again

```matlab
cos2 = @(x) cos(x).^2;
phi = @(x) 2(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);
beta = - c/v;
for k = 2:N,
    X = (pi/2)*rand;
    Y = cos2(X);
    Z = phi(X) + beta*(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;
    beta = - c/v;
end
```
Example: the tricky integral again
1. Variance reduction reconsidered

2. Sequential MC problems

3. 3 Examples of SMC problems
   - Prelude: Markov chains
   - Example 1: Simulation of extreme events
   - Example 2: Estimation in general HMMs
   - Example 3: Estimation of SAWs
Sequential MC problems

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

\[
\tau_n = \mathbb{E}_{f_n}(\phi(X_{0:n})) = \int_{X_n} \phi(x_{0:n}) f_n(x_{0:n}) \, dx_{0:n}
\]

over spaces \(X_n\) of increasing dimension, where again the densities \((f_n)_{n \geq 0}\) are 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 \(X_n\).

As we will see, such sequences appear in many applications in statistics and numerical analysis.
1. Variance reduction reconsidered

2. Sequential MC problems

3. 3 Examples of SMC problems
   - Prelude: Markov chains
   - Example 1: Simulation of extreme events
   - Example 2: Estimation in general HMMs
   - Example 3: Estimation of SAWs
Variance reduction reconsidered
Sequential MC problems
3 Examples of SMC problems

Prelude: Markov chains
Example 1: Simulation of extreme events
Example 2: Estimation in general HMMs
Example 3: Estimation of SAWs

We are here → •

1. Variance reduction reconsidered

2. Sequential MC problems

3. 3 Examples of SMC problems
   • Prelude: Markov chains
   • Example 1: Simulation of extreme events
   • Example 2: Estimation in general HMMs
   • Example 3: Estimation of SAWs
A Markov chain on $X \subseteq \mathbb{R}^d$ is a family of random variables (= stochastic process) $(X_k)_{k \geq 0}$ taking values in $X$ such that

$$
P(X_{k+1} \in B|X_0, X_1, \ldots, X_k) = P(X_{k+1} \in B|X_k)
$$

for all $B \subseteq X$. 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,

$$
P(X_{k+1} \in B|X_k = x_k) = \int_B q(x_{k+1}|x_k) \, dx_{k+1}.
$$
Markov chains (cont.)

The following theorem provides the joint density $f_n(x_0, x_1, \ldots, x_n)$ of $X_0, X_1, \ldots, X_n$.

**Theorem**

Let $(X_k)$ be Markov with initial distribution $\chi$. Then for $n > 0$,

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

**Corollary (Chapman-Kolmogorov equation)**

Let $(X_k)$ be Markov. Then for $n > 1$,

$$f_n(x_n|x_0) = \int \cdots \int \left( \prod_{k=0}^{n-1} q(x_{k+1}|x_k) \right) dx_1 \cdots dx_{n-1}.$$
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 $\alpha$ 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,

$$P(X_{k+1} \leq x_{k+1} | X_k = x_k) = P(\alpha X_k + \epsilon_{k+1} \leq x_{k+1} | X_k = x_k)$$

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

implying that

$$q(x_{k+1} | x_k) = \frac{\partial}{\partial x_{k+1}} P(X_{k+1} \leq x_{k+1} | X_k = x_k)$$

$$= \frac{\partial}{\partial x_{k+1}} P(\epsilon_{k+1} \leq x_{k+1} - \alpha x_k) = f(x_{k+1} - \alpha x_k).$$
1. Variance reduction reconsidered

2. Sequential MC problems

3. Examples of SMC problems
   - Prelude: Markov chains
   - Example 1: Simulation of extreme events
   - Example 2: Estimation in general HMMs
   - Example 3: Estimation of SAWs
Simulation of rare events for Markov chains

Let \((X_k)\) be a Markov chain on \(X = \mathbb{R}\) and consider the rectangle \(B = B_0 \times B_1 \times \cdots B_n \subseteq \mathbb{R}^n\), where every \(B_\ell = (a_\ell, b_\ell)\) is an interval. Here \(B\) can be a possibly extreme event.

Say that we wish to compute, sequentially as \(n\) increases, some expectation under the conditional distribution \(f_{n|B}\) of the states \(X_{0:n} = (X_0, X_2, \ldots, X_n)\) given \(X_{0:n} \in B\), i.e.

\[
\tau_n = \mathbb{E}_{f_n}(\phi(X_{0:n})|X_{0:n} \in B) = \mathbb{E}_{f_{n|B}}(\phi(X_{0:n}))
\]

\[
= \int_{B} \phi(x_{0:n}) \frac{f(x_{0:n})}{\mathbb{P}(X_{0:n} \in B)} \, dx_{0:n}.
\]

Here the unknown probability \(c_n = \mathbb{P}(X_{0:n} \in B)\) of the rare event \(B\) is often the quantity of interest.
Simulation of rare events for Markov chains (cont.)

As

\[ c_n = \mathbb{P}(X_{0:n} \in B) = \int \mathbb{1}_B(x_{0:n}) f(x_{0:n}) \, dx_{0:n} \]

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

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

\[ c_n^N = \frac{N_B}{N}. \]

**Problem:** if \( c_n = 10^{-9} \) we may expect to produce a billion draws before obtaining a single draw belonging to \( B \)! As we will see, SMC methods solve the problem efficiently.
1. Variance reduction reconsidered

2. Sequential MC problems

3. 3 Examples of SMC problems
   - Prelude: Markov chains
   - Example 1: Simulation of extreme events
   - Example 2: Estimation in general HMMs
   - Example 3: Estimation of SAWs
A hidden Markov model (HMM) comprises two stochastic processes:

1. A Markov chain \((X_k)_{k \geq 0}\) with transition density \(q\):

   \[X_{k+1}|X_k = x_k \sim q(x_{k+1}|x_k).\]

   The Markov chain is not seen by us (hidden) 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 \(Y_k|(X_k)_{k \geq 0} \overset{d}{=} Y_k|X_k\) will be denoted by \(p(y_k|x_k)\).
Graphically:

\[ Y_{k-1} \rightarrow X_{k-1} \rightarrow X_k \rightarrow X_{k+1} \rightarrow \ldots \]

\[ Y_k \rightarrow X_k \rightarrow X_{k+1} \rightarrow \ldots \]

\[ Y_k | X_k = x_k \sim p(y_k | x_k) \quad \text{(Observation density)} \]

\[ X_{k+1} | X_k = x_k \sim q(x_{k+1} | x_k) \quad \text{(Transition density)} \]

\[ X_0 \sim \chi(x_0) \quad \text{(Initial distribution)} \]
Example HMM: A stochastic volatility model

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

\[
\begin{align*}
X_{k+1} &= \alpha X_k + \sigma \epsilon_{k+1}, \\
Y_k &= \beta \exp \left( \frac{X_k}{2} \right) \epsilon_k,
\end{align*}
\]

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

- the values of the observation process \( (Y_k) \) are observed daily log-returns (from e.g. the Swedish OMXS30 index) and
- the hidden chain \( (X_k) \) is the unobserved log-volatility (modeled by a stationary AR(1) process).

The strength of this model is that it allows for volatility clustering, a phenomenon that is often observed in real financial time series.
Example HMM: A stochastic volatility model

A realization of the model looks like follows (here $\alpha = 0.975$, $\sigma = 0.16$, and $\beta = 0.63$).
Example HMM: A stochastic volatility model

The smoothing distribution

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

**Theorem (Smoothing distribution)**

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

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

$$L_n(y_{0:n}) = \text{density of the observations } y_{0:n}$$

$$= \int \cdots \int \chi(x_0)p(y_0|x_0)\prod_{k=1}^{n}p(y_k|x_k)q(x_k|x_{k-1}) \, dx_0 \cdots dx_n.$$
Estimation of smoothed expectations

Being a high-dimensional (say \( n \approx 1000 \) or \( 10,000 \)) integral over complicated integrands, \( L_n(y_{0:n}) \) is in general unknown. However by writing

\[
\tau_n = \mathbb{E}(\phi(X_{0:n})|Y_{0:n} = y_{0:n}) = \int \cdots \int \phi(x_{0:n}) f_n(x_{0:n}|y_{0:n}) \, dx_0 \cdots dx_n
\]

\[
= \int \cdots \int \phi(x_{0:n}) \frac{z_n(x_{0:n})}{c_n} \, dx_0 \cdots dx_n,
\]

with

\[
\begin{aligned}
z_n(x_{0:n}) &= \chi(x_0) p(y_0|x_0) \prod_{k=1}^{n} p(y_k|x_k) q(x_k|x_{k-1}), \\
c_n &= L_n(y_{0:n}),
\end{aligned}
\]

we may cast the problem of computing \( \tau_n \) into the framework of self-normalized IS. In particular we would like to update sequentially, in \( n \), the approximation as new data \((Y_k)\) appears.
Variance reduction reconsidered
Sequential MC problems
3 Examples of SMC problems

Prelude: Markov chains
Example 1: Simulation of extreme events
Example 2: Estimation in general HMMs
Example 3: Estimation of SAWs

1. Variance reduction reconsidered
2. Sequential MC problems
3. 3 Examples of SMC problems
   - Prelude: Markov chains
   - Example 1: Simulation of extreme events
   - Example 2: Estimation in general HMMs
   - Example 3: Estimation of SAWs
Self-avoiding walks (SAWs)

Denote by $\mathcal{N}(x)$ the set of neighbors of a point $x$ in $\mathbb{Z}^2$. Let

$$S_n \overset{\text{def}}{=} \{x_0:n \in \mathbb{Z}^{2n} : x_0 = 0, |x_{k+1} - x_k| = 1, x_k \neq x_\ell, \forall 0 \leq \ell < k \leq n\}$$

be the set of $n$-step self-avoiding walks in $\mathbb{Z}^2$.
In addition, let

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

SAWs are used in

- 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 combinatorial problem!
An MC approach to SAWs

Trick: Let $f_n(x_{0:n})$ be the uniform distribution on $S_n$:

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

We may thus cast the problem of computing the number $c_n$ (= the normalizing constant of $f_n$) into the framework of self-normalized IS based on some convenient instrumental distribution $g_n$ on $\mathbb{Z}^{2n}$.

In addition, solving this problem for $n = 1, 2, 3, \ldots, 508, 509, \ldots$ calls for sequential implementation of IS.

This will be the topic of HA2!