Monte Carlo and Empirical Methods for Stochastic Inference (MASM11/FMSN50)

Magnus Wiktorsson
Centre for Mathematical Sciences
Lund University, Sweden

Lecture 6
Sequential Monte Carlo methods II
February 7, 2019

Plan of today's lecture

1. Last time: Sequential MC problems
2. Random number generation reconsidered
3. Sequential Monte Carlo (SMC) methods
    - Overview
    - Sequential importance sampling (SIS)

We are here $\longrightarrow \bullet$

Last time: Sequential MC problems

In the sequential MC framework, we aim at sequentially estimating sequences $(\tau_n)_{n\geq 0}$ of expectations

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

over spaces X_n of increasing dimension, where the densities (f_n) are known up to normalizing constants only, i.e., for every $n \geq 0$,

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

where c_n is an unknown constant.

Last time: Markov chains

Some applications involved the notion of 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

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

The density q of the distribution of X_{k+1} given $X_k = x_k$ is called the transition density of (X_k). Consequently,

$$\mathbb{P}(X_{k+1} \in \mathsf{B} | X_k = x_k) = \int_{\mathsf{B}} q(x_{k+1} | x_k) \, \mathrm{d}x_{k+1}.$$

As a first example we considered an AR(1) process:

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

where α is a constant and (ϵ_k) are i.i.d. variables.

Last time: 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 $X_0 \sim \chi$. 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 (The Chapman-Kolmogorov equation)

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

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

Last time: Rare event analysis (REA) for Markov chains

Let (X_k) be a Markov chain. Assume that we want to compute, for $n = 0, 1, 2, \ldots$

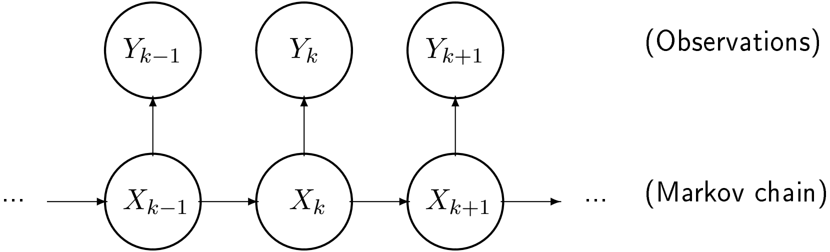
$$\begin{aligned}\tau_n = \mathbb{E}(\phi(X_{0:n}) | X_{0:n} \in \mathsf{B}) &= \int_{\mathsf{B}} \phi(x_{0:n}) \frac{f_n(x_{0:n})}{\mathbb{P}(X_{0:n} \in \mathsf{B})} \, \mathrm{d}x_{0:n} \\ &= \int_{\mathsf{B}} \phi(x_{0:n}) \frac{\chi(x_0) \prod_{k=0}^{n-1} q(x_{k+1}|x_k)}{\mathbb{P}(X_{0:n} \in \mathsf{B})} \, \mathrm{d}x_{0:n},\end{aligned}$$

where B is a possibly “rare” event and $\mathbb{P}(X_{0:n} \in \mathsf{B})$ is generally unknown. We thus face a sequential MC problem $(*)$ with

$$\begin{cases} z_n(x_{0:n}) \leftarrow \chi(x_0) \prod_{k=0}^{n-1} q(x_{k+1}|x_k), \\ c_n \leftarrow \mathbb{P}(X_{0:n} \in \mathsf{B}). \end{cases}$$

Last time: Estimation in general HMMs

Graphically:



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

Last time: Estimation in general HMMs

In an HMM, the smoothing distribution $f_n(x_{0:n}|y_{0:n})$ is 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

$$\begin{aligned} L_n(y_{0:n}) &= \textit{density of the observations } y_{0:n} \\ &= \int \cdots \int \chi(x_0)p(y_0|x_0)\prod_{k=1}^{n} p(y_k|x_k)q(x_k|x_{k-1})\, dx_0 \cdots dx_n. \end{aligned}$$

Last time: Estimation in general HMMs

Assume that we want to compute, online for $n = 0, 1, 2, \ldots$,

$$\tau_n = \mathbb{E}(\phi(X_{0:n}) | Y_{0:n} = y_{0:n})$$
$$= \int \cdots \int \phi(x_{0:n}) f_n(x_{0:n} | y_{0:n}) \, \mathrm{d}x_0 \cdots \mathrm{d}x_n$$
$$= \int \cdots \int \phi(x_{0:n}) \frac{\chi(x_0) p(y_0 | x_0) \prod_{k=1}^{n} p(y_k | x_k) q(x_k | x_{k-1})}{L_n(y_{0:n})} \, \mathrm{d}x_0 \cdots \mathrm{d}x_n,$$

where $L_n(y_{0:n})$ (= obscene integral) is generally unknown. We thus face a sequential MC problem $(*)$ with

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

We are here $\longrightarrow \bullet$

1. Last time: Sequential MC problems
2. Random number generation reconsidered
3. Sequential Monte Carlo (SMC) methods
   - Overview
   - Sequential importance sampling (SIS)

Conditional methods

Say that we want to generate a random vector from a given bivariate density $p(x, y)$. If we know how to draw from the conditional distribution $p(y|x)$ and the marginal $p(x)$ this can be done naturally using the following scheme.

draw $Z_1 \sim p(x)$
draw $Z_2 \sim p(y|x = Z_1)$
return (Z_1, Z_2)

Conditional methods

This can be naturally extended to n-variate densities $p(x_1, \dots, x_n)$:

draw $Z_1 \sim p(x_1)$
draw $Z_2 \sim p(x_2 | x_1 = Z_1)$
draw $Z_3 \sim p(x_3 | x_1 = Z_1, x_2 = Z_2)$
$\vdots$
draw $Z_{n-1} \sim p(x_{n-1} | x_1 = Z_1, x_2 = Z_2, \dots, x_{n-2} = Z_{n-2})$
draw $Z_n \sim p(x_n | x_1 = Z_1, x_2 = Z_2, \dots, x_{n-1} = Z_{n-1})$
return $(Z_1, \dots, Z_n)$

Theorem

The vector $(Z_1, \dots, Z_n)$ has indeed n-variate density function $p(x_1, \dots, x_n)$.

We are here ⟶ •

1. Last time: Sequential MC problems
2. Random number generation reconsidered
3. Sequential Monte Carlo (SMC) methods
   - Overview
   - Sequential importance sampling (SIS)

We are here $\longrightarrow \bullet$

1. Last time: Sequential MC problems
2. Random number generation reconsidered
3. Sequential Monte Carlo (SMC) methods
   - Overview
   - Sequential importance sampling (SIS)

Sequential Monte Carlo (SMC) methods

It is natural to aim at solving the problem using usual self-normalized IS.

However, the generated samples $(X_i^{0:n}, \omega_n(X_i^{0:n}))$ should be such that

- having $(X_i^{0:n}, \omega_n(X_i^{0:n}))$, the next sample $(X_i^{0:n+1}, \omega_{n+1}(X_i^{0:n+1}))$ is easily generated with a complexity that does not increase with n (online sampling).
- the approximation remains stable as n increases.

We call each draw $X_i^{0:n} = (X_i^0, \ldots, X_i^n)$ a particle. Moreover, we denote importance weights by

$$\omega_n^i \stackrel{\text{def}}{=} \omega_n(X_i^{0:n}).$$

We are here $\longrightarrow \bullet$

1. Last time: Sequential MC problems
2. Random number generation reconsidered
3. Sequential Monte Carlo (SMC) methods
   - Overview
   - Sequential importance sampling (SIS)

Sequential importance sampling (SIS)

We proceed recursively. Assume that we have generated particles $(X_i^{0:n})$ from $g_n(x_{0:n})$ so that

$$\sum_{i=1}^{N} \frac{\omega_n^i}{\sum_{\ell=1}^{N} \omega_n^\ell} \phi(X_i^{0:n}) \approx \mathbb{E}_{f_n}(\phi(X_{0:n})),$$

where, as usual, $\omega_n^i = \omega_n(X_i^{0:n}) = z_n(X_i^{0:n})/g_n(X_i^{0:n})$.

Key trick: Choose an instrumental distribution satisfying

$$g_{n+1}(x_{0:n+1}) = g_{n+1}(x_{n+1}|x_{0:n})g_{n+1}(x_{0:n})$$
$$g_{n+1}(x_{0:n+1}) = g_{n+1}(x_{n+1}|x_{0:n})g_n(x_{0:n}).$$

SIS (cont.)

Now assume that we have drawn $X_{0:n} \sim g_n(x_{0:n})$. Then, as

$$g_{n+1}(x_{0:n+1}) = g_{n+1}(x_{n+1}|x_{0:n})g_{n+1}(x_{0:n}) = g_{n+1}(x_{n+1}|x_{0:n})g_n(x_{0:n}),$$

the conditional method allows us to generate a draw $X_{0:n+1}$ from $g_{n+1}(x_{0:n+1})$ using the following procedure:

draw $X_{n+1} \sim g_{n+1}(x_{n+1}|x_{0:n} = X_{0:n})$
let $X_{0:n+1} \leftarrow (X_{0:n}, X_{n+1})$

This can be repeated recursively, yielding online sampling from the sequence (g_n).

SIS (cont.)

Consequently, $X_i^{0:n+1}$ and ω_{n+1}^i can be generated by

- keeping the previous $X_i^{0:n}$,
- simulating $X_i^{n+1} \sim g_{n+1}(x_{n+1}|X_i^{0:n})$,
- setting $X_i^{0:n+1} = (X_i^{0:n}, X_i^{n+1})$, and
- computing

$$\begin{aligned}\omega_{n+1}^i &= \frac{z_{n+1}(X_i^{0:n+1})}{g_{n+1}(X_i^{0:n+1})} \\ &= \frac{z_{n+1}(X_i^{0:n+1})}{z_n(X_i^{0:n})g_{n+1}(X_i^{n+1}|X_i^{0:n})} \times \frac{z_n(X_i^{0:n})}{g_n(X_i^{0:n})} \\ &= \frac{z_{n+1}(X_i^{0:n+1})}{z_n(X_i^{0:n})g_{n+1}(X_i^{n+1}|X_i^{0:n})} \times \omega_n^i.\end{aligned}$$

SIS (cont.)

Voilà, the sample $(X_i^{0:n+1}, \omega_{n+1}^i)$ can now be used to approximate $\mathbb{E}_{f_{n+1}}(\phi(X_{0:n+1}))$!

So, by running the SIS algorithm, we have updated an approximation

$$\sum_{i=1}^N \frac{\omega_n^i}{\sum_{\ell=1}^N \omega_n^\ell} \phi(X_i^{0:n}) \approx \mathbb{E}_{f_n}(\phi(X_{0:n}))$$

to an approximation

$$\sum_{i=1}^N \frac{\omega_{n+1}^i}{\sum_{\ell=1}^N \omega_{n+1}^\ell} \phi(X_i^{0:n+1}) \approx \mathbb{E}_{f_{n+1}}(\phi(X_{0:n+1}))$$

by only adding a component X_i^{n+1} to $X_i^{0:n}$ and sequentially updating the weights.

SIS: Pseudo code

for $i = 1 \to N$ **do**
  draw $X_i^0 \sim g_0$
  set $\omega_0^i = \frac{z_0(X_i^0)}{g_0(X_i^0)}$
end for
return (X_i^0, ω_0^i)
for $k = 0, 1, 2, \ldots$ **do**
  for $i = 1 \to N$ **do**
    draw $X_i^{k+1} \sim g_{k+1}(x_{k+1} | X_i^{0:k})$
    set $X_i^{0:k+1} \leftarrow (X_i^{0:k}, X_i^{k+1})$
    set $\omega_{k+1}^i \leftarrow \frac{z_{k+1}(X_i^{0:k+1})}{z_k(X_i^{0:k}) g_{k+1}(X_i^{k+1} | X_i^{0:k})} \times \omega_k^i$
  end for
  return $(X_i^{0:k+1}, \omega_{k+1}^i)$
end for

Example: REA reconsidered

We consider again the example of REA for Markov chains ($\mathsf{X} = \mathbb{R}$, $X_0 = x_0 = a$):

$$\tau_n = \mathbb{E}(\phi(X_{0:n}) | a \leq X_\ell, \forall \ell = 0, \ldots, n) = \int_{(a,\infty)^n} \phi(x_{0:n}) \underbrace{\frac{\prod_{k=1}^{n-1} q(x_{k+1}|x_k)}{\mathbb{P}(a \leq X_\ell, \forall \ell)}}_{= z_n(x_{0:n})/c_n} \, \mathrm{d}x_{1:n}.$$

Choose $g_{k+1}(x_{k+1}|x_{0:k})$ to be the conditional density of X_{k+1} given X_k and $X_{k+1} \geq a$:

$$g_{k+1}(x_{k+1}|x_{0:k}) = \{\text{cf. HA1, Problem 1}\} = \frac{q(x_{k+1}|x_k)}{\int_a^\infty q(z|x_k)\, \mathrm{d}z}.$$

Example: REA

This implies that (recall that we have conditioned on $X_0 = x_0 = a$)

$$g_n(x_{0:n}) = \prod_{k=0}^{n-1} \frac{q(x_{k+1}|x_k)}{\int_a^\infty q(z|x_k)\,\mathrm{d}z}.$$

In addition, the weights are updated according to

$$\begin{aligned}
\omega_{k+1}^i &= \frac{z_{k+1}(X_i^{0:k+1})}{z_k(X_i^{0:k})g_{k+1}(X_i^{k+1}|X_i^{0:k})} \times \omega_k^i \\
&= \frac{\prod_{\ell=0}^{k} q(X_i^{\ell+1}|X_i^\ell)}{\prod_{\ell=0}^{k-1} q(X_i^{\ell+1}|X_i^\ell) \times \frac{q(X_i^{k+1}|X_i^k)}{\int_a^\infty q(z|X_i^k)\,\mathrm{d}z}} \times \omega_k^i \\
&= \int_a^\infty q(z|X_i^k)\,\mathrm{d}z \times \omega_k^i.
\end{aligned}$$

Example: REA; Matlab implementation for AR(1) process with Gaussian noise

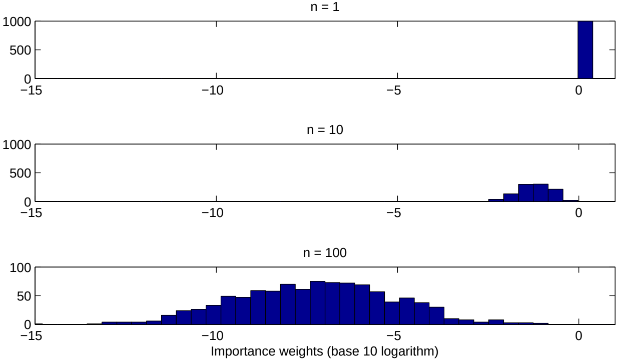
```
% design of instrumental distribution:

int = @(x) 1 - normcdf(a,alpha*x,sigma);
trunk_td_rnd = ...        % use e.g. HA1, Problem 1, to simulate
                          % the conditional transition density;
% SIS:

part = a*ones(N,1); % initialization of all particles in a
w = ones(N,1);
for k = 1:(n - 1), % main loop
    part_mut = trunk_td_rnd(part);
    w = w.*int(part);
    part = part_mut;
end
c = mean(w); % estimated probability
```

REA: Importance weight distribution

☹ Serious drawback of SIS: the importance weights degenerate!...



What's next?

Weight degeneration is a universal problem with the SIS method and is due to the fact that the particle weights are generated through subsequent multiplications.

This drawback prevented—during several decades—the SIS method from being practically useful.

Next week we will discuss an elegant solution to this problem: SIS with resampling (SISR).