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

Magnus Wiktorsson
Centre for Mathematical Sciences
Lund University, Sweden

Lecture 1
Introduction
January 21, 2020

Outline

People and literature

The following people are involved in the course:

Function	Name	Room	E-mail
Lecturer	Magnus Wiktorsson	MH:130	magnusw@maths.lth.se
Assistants	Samuel Wiqvist	MH:326	samuel@maths.lth.se
	Maria Juhlin	MH:138A	juhlin@maths.lth.se
Secretary	Susann Nordqvist	MH:221	susann.nordqvist@matstat.lu.se

The following material will be used:

- Slides. Will be available online immediately after each lecture.
- Geof H. Givens and Jennifer A. Hoeting *Computational Statistics* Second Edition (2012). Ebook available at: http://ludwig.lub.lu.se/login?url=http://onlinelibrary.wiley.com/book/10.1002/9781118555552 Book homepage: http://www.stat.colostate.edu/computationalstatistics/

Course schedule and homepage

The course schedule is as follows:

	Weekday	Time	Room
Lecture I	Tuesday	13.15–15	E:C
Computer session I	Wednesday	8.15–10	E:Neptunus, E:Pluto
Computer session II	Wednesday	15.15–17	E:Neptunus, E:Pluto
(Week 2 only)	Tuesday	15.15-17	E:Neptunus, E:Pluto
Lecture II	Thursday	15.15–17	E:C
Office hours	Friday	15.30–16.30	MH:130, MH:326 and MH:138A

The computer sessions and office hours start in Week 2.

Information and Matlab files will be available at the homepage:

`https://canvas.education.lu.se/courses/3630`.

Examination

The examination comprises

- three larger projects handed out during Weeks 2, 4, and 6. Each project requires the submission of a report. The projects, which are solved in pairs, concern
  1. simulation and Monte Carlo integration,
  2. sequential Monte Carlo methods, and
  3. Markov chain Monte Carlo methods and Bayesian inference.
- an oral exam.

The final mark will be computed according to the formula

$$\text{Final mark} = \left\lceil \frac{\text{Median project mark} + \text{Mark at oral exam}}{2} \right\rceil.$$

Course contents

Part I: Monte Carlo integration

- Simulation and Monte Carlo integration (Weeks 1–2)
- Sequential Monte Carlo methods (3–4)
- Markov chain Monte Carlo (MCMC) methods (4–5)

Part II: Applications to inference

- Applications of MCMC to Bayesian statistics (5–6)
- Bootstrap (6–7)
- Permutation tests (7)

Principal aim

The main problem of the course is to compute some expectation

$$\tau \overset{\text{def.}}{=} \mathbb{E}(\phi(X)) = \int_{\mathsf{A}} \phi(x) f(x)\, dx,$$

where

- X is a random variable taking values in $\mathsf{A} \subseteq \mathbb{R}^d$ (where $d \in \mathbb{N}$ may be very large),
- $f : \mathsf{A} \to \mathbb{R}_+$ is the probability density of X (referred to as the target density), and
- $\phi : \mathsf{A} \to \mathbb{R}$ is a function (referred to as the objective function) such that the above expectation is finite.

As we will see in the following, this framework covers a large set of fundamental problems in statistics, numerical analysis, and other scientific disciplines.

The Monte Carlo (MC) method in a nutshell (Ch. 6.1)

Let $X_1, X_2, \ldots, X_N$ be independent random variables with density f. Then, by the law of large numbers, as N tends to infinity,

$$\tau_N \overset{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \phi(X_i) \to \tau = \mathbb{E}(\phi(X)). \quad \text{(a.s.)}$$

Inspired by this result, we formulate the following basic MC sampler (Stanisław Ulam, John von Neumann, and Nicholas Metropolis; the Los Alamos Scientific Laboratory; 40's):

for $i = 1 \to N$ **do**
  draw $X_i \sim f$
end for
set $\tau_N \leftarrow \sum_{i=1}^{N} \phi(X_i)/N$
return τ_N

"The first thoughts and attempts I made to practice [the Monte Carlo method] *were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaires. The question was what are the chances that a Canfield solitaire*[1] *laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than 'abstract thinking' might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations. Later* [in 1946], *I described the idea to John von Neumann, and we began to plan actual calculations."*

–Stanisław Ulam

[1] See link on course home page for rules of the game.

The curse of dimensionality

Most numerical integration methods are of order $\mathcal{O}(N^{-c/d})$, where N is the number of function evaluations needed to approximate the integral and $c > 0$ is a constant—cf. the trapezoidal method ($c = 2$) or the Simpson method ($c = 4$).Thus, for some $C \geq 1$,

$$\epsilon_N \overset{\text{def.}}{=} |\tau - \tau_N| \leq CN^{-c/d}.$$

In order to guarantee that $\epsilon_N \leq \delta$, N should satisfy

$$CN^{-c/d} \leq \delta \Leftrightarrow N \geq \left(\frac{C}{\delta}\right)^{d/c}$$

This means that for a fixed error the number of function evaluations grows exponentially with the dimension d of A.

Rate of convergence of MC

For the MC method, the error is random. However, the central limit theorem implies, under the assumption that $\mathbb{V}(\phi(X)) < \infty$,

$$\sqrt{N}\left(\tau_N - \tau\right) \xrightarrow{\text{d.}} \mathcal{N}(0, \mathbb{V}(\phi(X))).$$

This means that for large N's,

$$\mathbb{V}\left(\sqrt{N}\left(\tau_N - \tau\right)\right) = N\mathbb{V}\left(\tau_N - \tau\right) = \mathbb{V}(\phi(X)),$$

implying that

$$\mathbb{D}\left(\tau_N - \tau\right) \overset{\text{def.}}{=} \sqrt{\mathbb{V}\left(\tau_N - \tau\right)} = \sqrt{\frac{\mathbb{V}(\phi(X))}{N}} = \frac{\mathbb{D}(\phi(X))}{\sqrt{N}}.$$

Thus, the MC convergence rate $\mathcal{O}(N^{-1/2})$ is independent of d!

Example: Integration

The problem of computing an integral of form

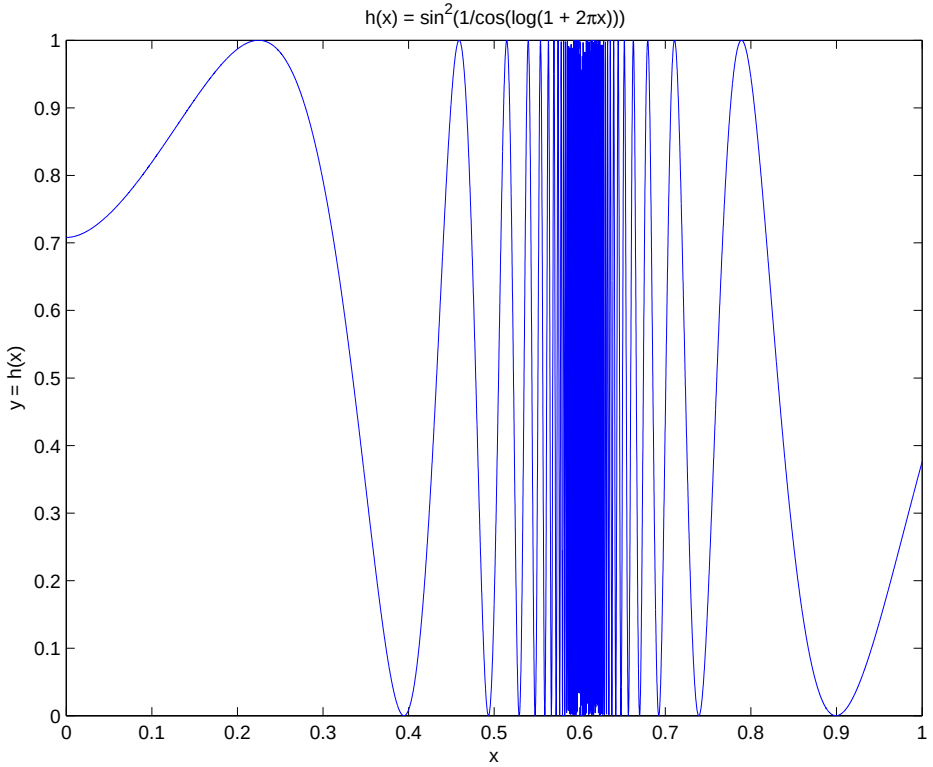
$$\int_{(0,1)^d} h(x)\,dx$$

can be cast into our framework by letting

$$\begin{cases} \mathsf{A} \leftarrow (0,1)^d \\ \phi \leftarrow h \\ f \leftarrow \mathbb{1}_{(0,1)^d}(= \mathsf{unif}(0,1)^d). \end{cases}$$

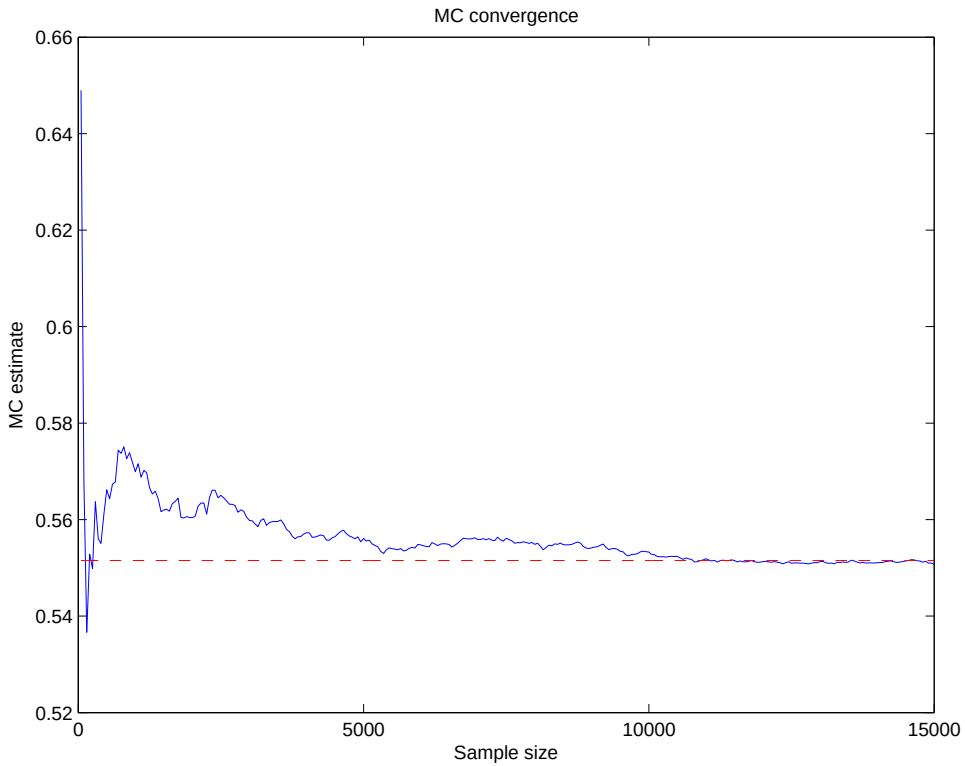
Example: Integration (cont.)

As an example for $d = 1$, let $h(x) = \sin^2(1/\cos(\log(1 + 2\pi x)))$:



Example: Integration (cont.)

```
h = @(x) (sin(1./cos(log(1 + 2*pi*x)))).^2;
U = rand(1,N);
tau = mean(h(U));
```



Example: Integration (cont.)

Now, let $\Omega \subseteq \mathbb{R}^d$ be arbitrary and consider the general case

$$\int_\Omega h(x)\,dx.$$

Then we may choose some positive reference density g on Ω (e.g. the $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ density if $\Omega = \mathbb{R}^d$) and write

$$\int_\Omega h(x)\,dx = \int_\Omega \frac{h(x)}{g(x)} g(x)\,dx,$$

which can again be cast into the MC framework by letting

$$\begin{cases} \mathsf{A} \leftarrow \Omega \\ \phi \leftarrow h/g \\ f \leftarrow g. \end{cases}$$

Example: Intractable likelihoods

A similar technique can be used for estimating intractable likelihood functions. Indeed, assume that we know a density $f_\theta(x)/c(\theta)$ up to $f_\theta(x)$ only. Then

$$\int \frac{f_\theta(x)}{c(\theta)}\,dx = 1 \quad \Leftrightarrow \quad c(\theta) = \int f_\theta(x)\,dx = \int \frac{f_\theta(x)}{g(x)} g(x)\,dx,$$

where g is again some density that is easy to simulate from.

Thus, an estimate of $c(\theta)$ can be formed by generating a sample $X_1, \ldots, X_N$ from g and setting

$$c_N(\theta) \overset{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \frac{f_\theta(X_i)}{g(X_i)}.$$

This (and the last part of the previous example) is a first instance of importance sampling (Week 2).

Example: Filtering, Bayesian statistics

Often the joint density $p(x, y)$ of a pair (X, Y) of random variables is easily obtained while the conditional density

$$p(x|y) = \frac{p(x, y)}{\int p(x, y)\, dx}$$

of X given Y is by far more complicated due to the normalizing integral. Again MC applies, especially in the shape of Markov Chain Monte Carlo methods (Weeks 4–6). Typical examples are

- filtering of a signal/image from noisy observations,
- Bayesian statistics, where the variable X plays the role of an unknown parameter (usually denoted by θ) and $p(x|y)$ is the so-called posterior distribution of the parameter given observed data Y.

Example: Pricing of contingent claims

Diffusion processes are processes related to Brownian motion. These are fundamental within mathematical finance modeling.

Figure: Evolution of the Nike, Inc. stock price S_t for $t \in (2003, 2011)$.

Example: Pricing of contingent claims (cont.)

Let $S \overset{\text{def.}}{=} (S_t)_{t \geq 0}$ be a price process. A contingent claim is a financial contract which stipulates that the holder of the contract will obtain $\mathcal{X}$ SEK at time T, where for some contract function Φ,

$$\mathcal{X} = \Phi(S_T).$$

Under certain assumptions, one may prove that the fair price F of the claim $\mathcal{X}$ at time $t \leq T$ is given by

$$F(s,t) = e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}} \left[\Phi(S_T) \mid S_t = s \right],$$

where $\mathbb{Q}$ indicates the "risk neutral dynamics" of S and r is the interest rate.

Thus, compute the price by (i) simulating $S_T | S_t = s$ repeatedly, (ii) compute the claim for each realization, and (iii) take the mean!

Example: Computing the size of a BIG set

Say that we want to compute the size of a finite but huge set S. Assume that $\mathsf{S} \subseteq \mathsf{T}$ and define a random variable X taking values in T with probabilities $p(x) = \mathbb{P}(X = x) > 0$, $x \in \mathsf{T}$. Then we can write

$$|\mathsf{S}| = \sum_{x \in \mathsf{T}} \mathbb{1}_{\mathsf{S}}(x) = \sum_{x \in \mathsf{T}} \frac{1}{p(x)} \mathbb{1}_{\mathsf{S}}(x) p(x) = \mathbb{E}(\mathbb{1}_{\mathsf{S}}(X)/p(X)),$$

which again fits into our MC integration framework with

$$\begin{cases} \mathsf{A} \leftarrow \mathsf{T} \\ \phi \leftarrow \mathbb{1}_{\mathsf{S}}/p \\ f \leftarrow p. \end{cases}$$

Sequential MC problems

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

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

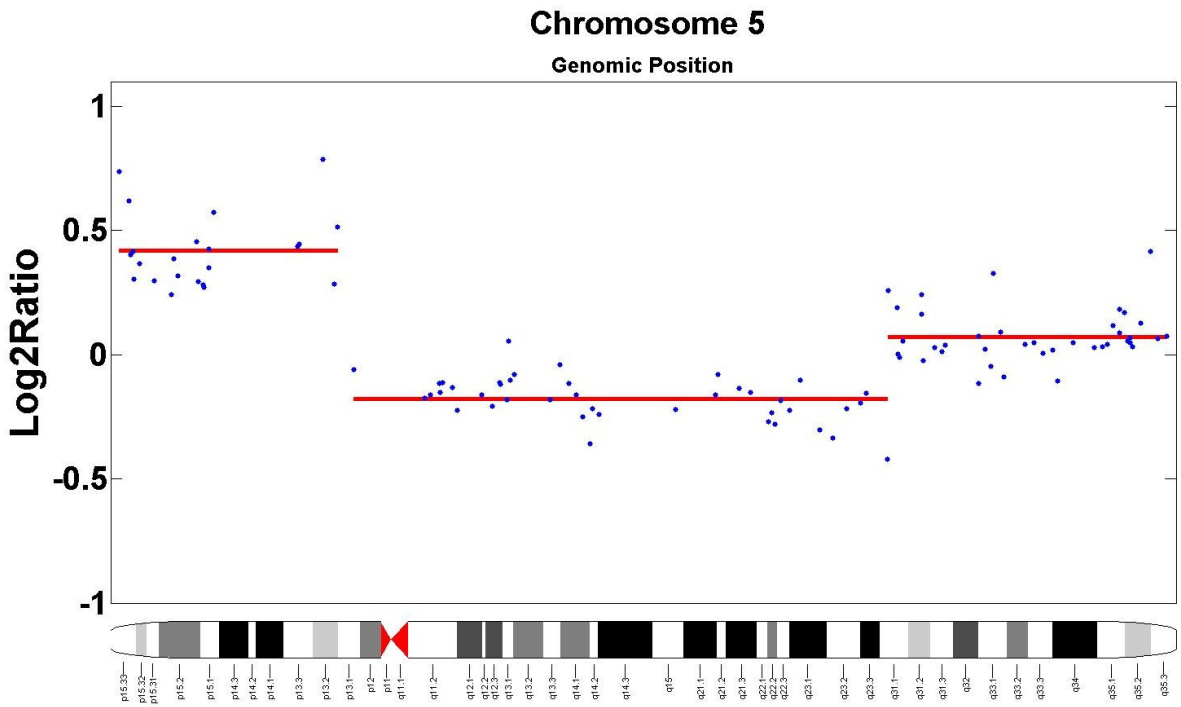
over spaces A_n of increasing dimension, where the densities $(f_n)_{n\geq 0}$ are known up to normalizing constants only, i.e., for every $n \geq 0$,

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

where $z_n(x_{0:n}) \geq 0$ and c_n is an unknown constant.

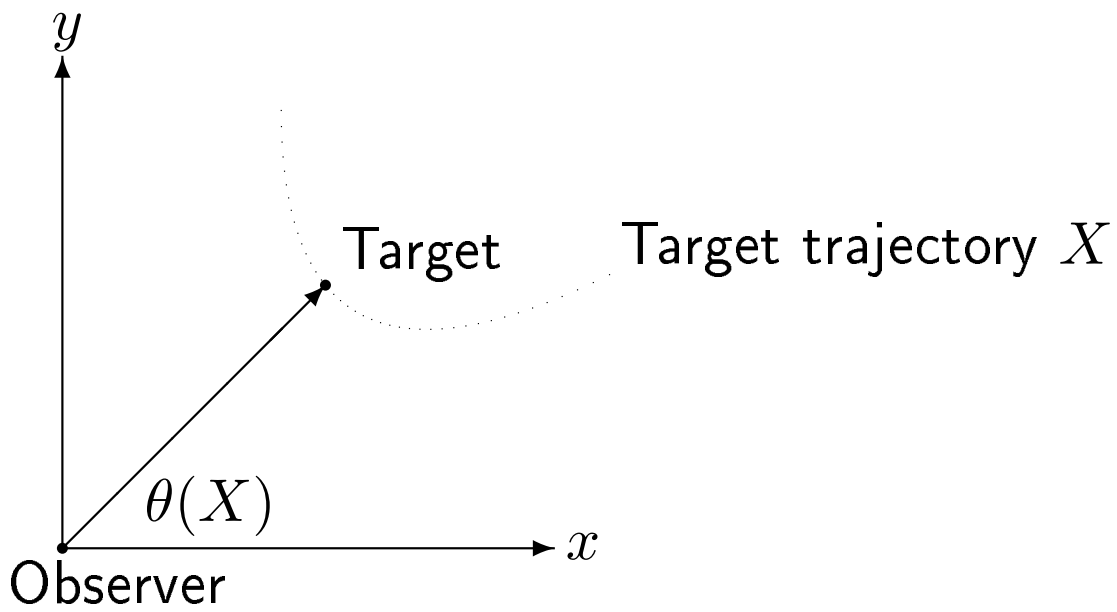
Example 3: Filtering in genetics

Cancer cells might have parts of chromosomes with different copy numbers. These numbers can be modelled efficiently using hidden Markov models (HMM), which are efficiently estimated using SMC (Weeks 3–4).



Example 3: Filtering in target tracking

An observer obtains noisy observations of the bearing of a moving target. In this HMM, the conditional distribution of the target given the observations can again be estimated online using SMC methods (Weeks 3–4).



What do we need to know?

OK, so what do we need to master for having practical use of the MC method?

Well, for instance, the following questions should be answered:

1: How do we generate the needed input random variables?

2: How many computer experiments should we do? What can be said about the error?

3: Can we exploit problem structure to speed up the computation?

Next lecture

Next time we will deal with the first two issues and discuss

- Pseudo-random number generation and
- MC output analysis.

See you!