Computer Intensive Methods in Mathematical Statistics

Johan Westerborn

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

Lecture 1
Introduction
22 March 2017

Outline

1. Some course information

2. Overview of the Monte Carlo method
   - Preliminaries
   - Principal aim
   - Interlude: Two fundamental results in statistics
   - The basic Monte Carlo sampler

3. A few examples

4. What's next?

Some course information
○○○○○

Overview of the Monte Carlo method
○○○○○○○○○○○○○○○○○○○○

A few examples
○○○○○○○○○○○○○○

What's next?
○○

Outline

Course material

- The following material will be used:
    - Slides. Will be available online shortly after each lecture.
    - Sköld, M. (2005). *Computer Intensive Statistical Methods*. These notes can be downloaded through the course home page.
- Some recommended additional reference material:
    - Robert. C. and Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd ed.).
    - Doucet, A., de Freitas, N., Gordon, N. (eds.) (2001). *Sequential Monte Carlo methods in Practice*.
    - Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*.

People and administration

- The course will be taught by
    - Johan Westerborn (lectures, exercises),
    - Henrik Hult (examiner).
- All information concerning course registration, exam application, etc., is available through

    https://www.kth.se/sci/institutioner/math/utb/studentexp

    (important also for PhD students, who are not able to register via "Mina sidor").

Course schedule and homepage

- The course schedule is irregular, but there are in general three meetings per week:

  exercise class (E) – lecture (L) – lecture.

- The exercise classes are launched first in Week 2; thus, there are (three lectures) during the first course week.

- Information and MATLAB files will be available through

  http://www.math.kth.se/matstat/gru/sf2955/.

Some course information
○○○●○

Overview of the Monte Carlo method
○○○○○○○○○○○○○○○○○○○○

A few examples
○○○○○○○○○○○○○○○

What's next?
○○

Examination

- The examination comprises
    - two larger projects (3.0 credits, grade scale: P, F) handed out during Weeks 3 and 6, respectively. Each project requires the submission of a report. The projects, which are solved in pairs, treat
        1. sequential Monte Carlo methods, and
        2. Markov chain Monte Carlo methods and Bayesian inference.

        The projects are graded through a mandatory peer-review procedure, where each group reviews the reports of 2 other groups.
    - A written exam (4.5 credits, grade scale: A, B, C, D, E, FX, F) taking place on 30 May 2017, 14–19.

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 (7–8),
  - The EM algorithm (8),
  - Permutation tests (8).

Some course information
○○○○○

Overview of the Monte Carlo method
○○○○○○○○○○○○○○○○○○○○

A few examples
○○○○○○○○○○○○○○

What's next?
○○

Outline

1 Some course information

2 Overview of the Monte Carlo method
- Preliminaries
- Principal aim
- Interlude: Two fundamental results in statistics
- The basic Monte Carlo sampler

3 A few examples

4 What's next?

Some course information   Overview of the Monte Carlo method   A few examples   What's next?

Preliminaries

Outline

1. Some course information

2. Overview of the Monte Carlo method
   - **Preliminaries**
   - Principal aim
   - Interlude: Two fundamental results in statistics
   - The basic Monte Carlo sampler

3. A few examples

4. What's next?

Some course information | Overview of the Monte Carlo method | A few examples | What's next?

Preliminaries

Some fundamental concepts of probability

- A probability space contains
  - a sample space Ω, which is the set of possible outcomes $\omega \in \Omega$. Subsets $A \subseteq \Omega$ of the sample space are called events.
  - a probability measure $\mathbb{P}$ assigning a value in $[0, 1]$ to each event such that
    - (i) $\mathbb{P}(\Omega) = 1$,
    - (ii) $\mathbb{P}(\cup_{i \in I} A_i) = \sum_{i \in I} \mathbb{P}(A_i)$ for any (countable) collection $(A_i)_{i \in I}$ of pairwise disjoint events.
- Given a probability space $(\Omega, \mathbb{P})$, a random variable X is a function $\Omega \to X$, where X is called the state space of X (typically $\mathbb{R}^d$).

Some course information | Overview of the Monte Carlo method | A few examples | What's next?

Preliminaries

Some fundamental concepts of probability (cont.)

- Given a random variable X on some probability space $(\Omega, \mathbb{P})$, the function $X \ni x \mapsto F(x) = \mathbb{P}(X \leq x)$ is called distribution function of X.

- In the case where $X = \mathbb{R}^d$, a function $f : X \to \mathbb{R}_+$ such that for all $x = (x_1, \ldots, x_d)$,

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(z) \, dz,$$

is called the density of X. Consequently, $f(x) = \partial_{x_1} \cdots \partial_{x_d} F(x)$ for all x.

- Two random variables X and Y are independent if

$$f(x, y) = f(x)f(y), \quad \text{for all } x, y.$$

Some course information | Overview of the Monte Carlo method | A few examples | What's next?
○○○○○ | ○○○●○○○○○○○○○○○○○○○ | ○○○○○○○○○○○○○○ | ○○

Preliminaries

Some fundamental concepts of probability (cont.)

- The integral

$$\mathbb{E}(X) = \int_X x\, f(x)\, dx$$

is called the expectation of X.

- If ϕ is some function $X \to \mathbb{R}$ and X is some random variable, then $\phi(X)$ is again a random variable with expectation

$$\mathbb{E}(\phi(X)) = \int_X \phi(x) f(x)\, dx.$$

Some course information   Overview of the Monte Carlo method   A few examples   What's next?
00000   0000●0000000000000   00000000000000   00

Principal aim

Outline

1 Some course information

2 Overview of the Monte Carlo method
- Preliminaries
- **Principal aim**
- Interlude: Two fundamental results in statistics
- The basic Monte Carlo sampler

3 A few examples

4 What's next?

| Some course information | Overview of the Monte Carlo method | A few examples | What's next? |
| ooooo | ooooo●oooooooooooooo | ooooooooooooooo | oo |

Principal aim

Principal aim of the course

- Most problems treated in this course are related to the computation of some expectation

$$\tau \stackrel{\text{def}}{=} \mathbb{E}(\phi(X)) = \int_X \phi(x)f(x)\,dx,$$

where

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

- Covers a large set of fundamental problems in statistics, numerical analysis, and other scientific disciplines!

Some course information   Overview of the Monte Carlo method   A few examples   What's next?

Interlude: Two fundamental results in statistics

Outline

1. Some course information

2. Overview of the Monte Carlo method
   - Preliminaries
   - Principal aim
   - Interlude: Two fundamental results in statistics
   - The basic Monte Carlo sampler

3. A few examples

4. What's next?

Some course information   Overview of the Monte Carlo method   A few examples   What's next?
○○○○○   ○○○○○○○●○○○○○○○○○○○○   ○○○○○○○○○○○○○○○○   ○○

Interlude: Two fundamental results in statistics

The law of large numbers

- The central limit theorem (CLT) and the law of large numbers (LLN) are the most important results in probability and statistics.
- The LLN can be stated as follows.

Theorem

Let $X^1, X^2, X^3, \ldots$ be independent and identically distributed (iid) random variables with mean μ and set $S_N = \sum_{i=1}^{N} X^i$. Then

$$\lim_{N \to \infty} \frac{1}{N} S_N = \mu \quad \text{(with probability one).}$$

- As we will see, the CLT describes the error between S_N/N and μ for large N and provides the rate of convergence.

Some course information | Overview of the Monte Carlo method | A few examples | What's next?

Interlude: Two fundamental results in statistics

Convergence in distribution: definition

- Let X and $X^1, X^2, X^3, \ldots$ be random variables.
- Denote by $F_N(x) = \mathbb{P}(X^N \leq x)$ and $F(x) = \mathbb{P}(X \leq x)$ the distribution functions of X^N and X, respectively.
- Let C_F be the set of continuity points of F.
- The CLT describes convergence in distribution, which is defined as follows.

Definition

(X^N) is said to converge in distribution to X if for all $x \in C_F$,

$$\lim_{N \to \infty} F_N(x) = F(x).$$

Some course information | Overview of the Monte Carlo method | A few examples | What's next?

Interlude: Two fundamental results in statistics

Convergence in distribution: remarks

- In words, convergence in distribution means that the probability that X_N falls in a given range is approximately equal to the probability that X falls in that range, provided N is sufficiently large.

- Notation: $X_N \xrightarrow{\text{d.}} X$.

- The condition $x \in C_F$ is essential. Indeed, let $X_N = 1/N$ and $X = 0$ be deterministic; then

$$F_N(0) = \mathbb{P}(X_N \leq 0) = 0 \nrightarrow F(0) = \mathbb{P}(X \leq 0) = 1.$$

Some course information | Overview of the Monte Carlo method | A few examples | What's next?

Interlude: Two fundamental results in statistics

The central limit theorem

- The CLT can be stated as follows.

Theorem

Let $X^1, X^2, X^3, \ldots$ be iid with mean μ and variance σ^2 and set $S_N = \sum_{i=1}^{N} X^i$. Then

$$\sqrt{N} \left(\frac{1}{N} S_N - \mu \right) \xrightarrow{d.} Z,$$

where Z is normally distributed with zero mean and variance σ^2.

- This means that for large N,

$$S_N \overset{\text{appr.}}{\sim} N(\mu N, \sigma^2 N).$$

| Some course information | Overview of the Monte Carlo method | A few examples | What's next? |
| 00000 | 00000000000000000000 | 0000000000000 | 00 |

Interlude: Two fundamental results in statistics

Rolling dice

- As an example, we roll repeatedly a fair die and call the outcomes $X^1, X^2, X^3, \ldots$
- Recall that $\mu = \mathbb{E}(X^i) = 3.5$ and $\sigma^2 = \mathbb{V}(X^i) \approx 2.9$.
- After each roll we note the sum $S_N = \sum_{i=1}^{N} X^i$.
- MATLAB simulation:

```
X = randi(6,1,N);
S = cumsum(X);
means = S./(1:N);
```

Some course information | Overview of the Monte Carlo method | A few examples | What's next?

Interlude: Two fundamental results in statistics

Rolling dice: outcome
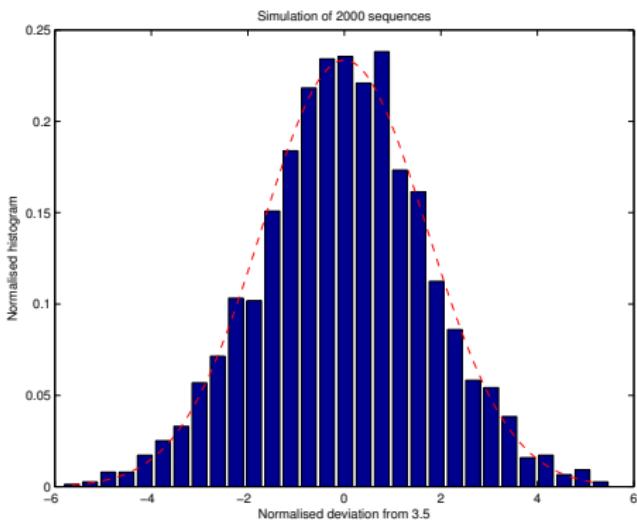
Figure: Normalised histogram (relative frequences) of 2,000 replicates of $\sqrt{2,000}(S_{2,000}/2,000 - 3.5)$ along with $N(0, 2.9)$.

Some course information   Overview of the Monte Carlo method   A few examples   What's next?
○○○○○   ○○○○○○○○○○○○○●○○○○○○   ○○○○○○○○○○○○○○   ○○

The basic Monte Carlo sampler

Outline

Some course information | Overview of the Monte Carlo method | A few examples | What's next?

The basic Monte Carlo sampler

The Monte Carlo (MC) method in a nutshell (Ch. 4)

- Let (X^i) be independent random variables with density f. Then, by the LLN, as $N \to \infty$,

$$\tau_N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \phi(X^i) \to \tau = \mathbb{E}(\phi(X)) = \int_X \phi(x) f(x) \, dx.$$

- Using this, the basic MC sampler can be formulated as:

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

Some course information	Overview of the Monte Carlo method	A few examples	What's next?
○○○○○	○○○○○○○○○○○○○○●○○○○	○○○○○○○○○○○○○	○○

The basic Monte Carlo sampler

(a) J. v. Neumann   (b) S. Ulam   (c) N. Metropolis

Figure: The inventors of Monte Carlo simulation (the Los Alamos Scientific Laboratory; 40's)

Some course information	Overview of the Monte Carlo method	A few examples	What's next?
○○○○○	○○○○○○○○○○○○○○○●○○○	○○○○○○○○○○○○○○	○○

The basic Monte Carlo sampler

"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 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

Some course information | Overview of the Monte Carlo method | A few examples | What's next?

The basic Monte Carlo sampler

The curse of dimensionality

- Recall that we are solving a possibly high dimensional integration problem.

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

- More specifically, this means that there is a constant $\tilde{c} < \infty$ such that for all N,

$$\varepsilon_N \stackrel{\text{def}}{=} |\tau - \tau_N| \leq \tilde{c} N^{-c/d}.$$

The basic Monte Carlo sampler

The curse of dimensionality

- Consequently, in order to guarantee that $\varepsilon_N \leq \delta$ for some small $\delta > 0$, the number N should satisfy

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

- When δ is small, $\tilde{c}/\delta > 1$. This means that for a given threshold the number of function evaluations grows exponentially with the dimension d of X.

- This prevents numerical integration from working efficiently in high dimensions.

Some course information | Overview of the Monte Carlo method | A few examples | What's next?

The basic Monte Carlo sampler

Rate of convergence of MC

- For the MC method, the error is random. However, the CLT implies, under the assumption that $\sigma^2(\phi) \overset{\text{def}}{=} \mathbb{V}(\phi(X)) < \infty$,

$$\sqrt{N}(\tau_N - \tau) \overset{\text{d.}}{\longrightarrow} \mathsf{N}(0, \sigma^2(\phi)).$$

- In particular,

$$\mathbb{V}(\tau_N - \tau) = \mathbb{V}(\tau_N) = \mathbb{V}\left(\frac{1}{N}\sum_{i=1}^{N}\phi(X^i)\right) = \frac{1}{N}\sigma^2(\phi),$$

implying that

$$\mathbb{D}(\tau_N - \tau) \overset{\text{def}}{=} \sqrt{\mathbb{V}(\tau_N - \tau)} = \frac{1}{\sqrt{N}}\sigma(\phi).$$

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

Outline

1. Some course information

2. Overview of the Monte Carlo method
   - Preliminaries
   - Principal aim
   - Interlude: Two fundamental results in statistics
   - The basic Monte Carlo sampler

3. A few examples

4. What's next?

Example: Integration

- The problem of computing any integral of form

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

can be perfectly cast into our framework by letting

$$\begin{cases} X \leftarrow (0,1)^d, \\ \phi \leftarrow h, \\ f \leftarrow \mathbb{1}_{(0,1)^d} \quad (= U(0,1)^d). \end{cases}$$

Example: Integration (cont.)

- For an example in the case $d = 1$, let
  $h(x) = \sin^2(1/\cos(\log(1 + 2\pi x)))$, $x \in (0, 1)$:

$h(x) = \sin^2(1/\cos(\log(1 + 2\pi x)))$

Some course information
○○○○○

Overview of the Monte Carlo method
○○○○○○○○○○○○○○○○○○○○

A few examples
○○●○○○○○○○○○○○○○

What's next?
○○
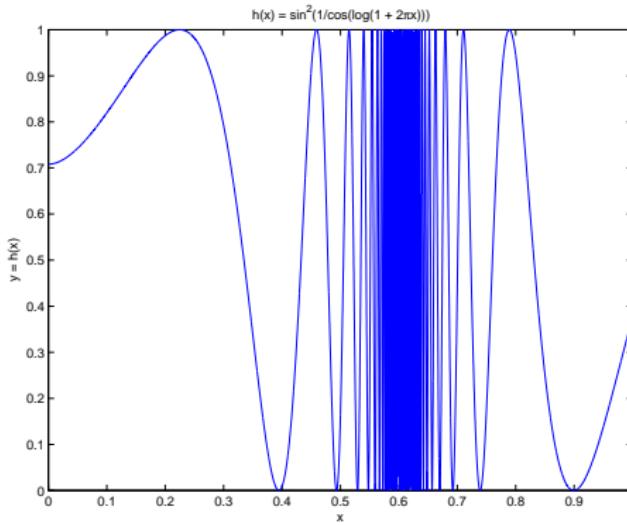
Example: Integration (cont.)

■ In MATLAB:

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


MC convergence

Some course information
00000

Overview of the Monte Carlo method
0000000000000000000

A few examples
0000●00000000000

What's next?
00
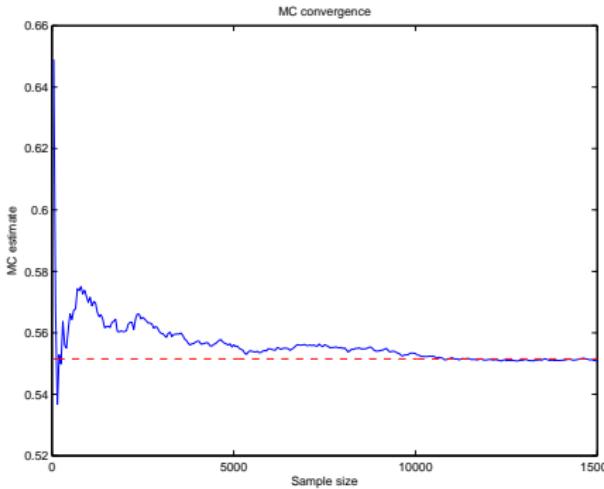
Example: Integration (cont.)

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

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

- Then we may choose some positive reference density g on Ω (e.g. the $N(\mathbf{0}, \mathbf{I}_d)$ 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

$$\begin{cases} X \leftarrow \Omega, \\ \phi \leftarrow h/g, \\ f \leftarrow g. \end{cases}$$

Some course information
00000

Overview of the Monte Carlo method
0000000000000000000

A few examples
0000●000000000

What's next?
00

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 $S \subseteq T$ and define a random variable X taking values in T with probabilities $p(x) = \mathbb{P}(X = x) > 0$, $x \in T$.
- Then we can write

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

which again fits into our MC integration framework with

$$\begin{cases} X \leftarrow T, \\ \phi \leftarrow \mathbb{1}_S/p, \\ f \leftarrow p. \end{cases}$$

Some course information
○○○○○

Overview of the Monte Carlo method
○○○○○○○○○○○○○○○○○○

A few examples
○○○○○●○○○○○○○○○

What's next?
○○

Example: Conditional distributions

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

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

of the conditional distribution 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 (MCMC, Weeks 4–6).

Example: Conditional distributions

- Computing conditional distributions play a critical role in models where X is latent and only partially observed through Y, e.g., in filtering of a signal/image from noisy observations.
- Moreover, determining conditional distributions is essential in
  - frequentist statistic and, in particular, maximum likelihood estimation in latent data models via the expectation-maximization (EM) algorithm (Week 8).
  - Bayesian statistics, where the variable X plays the role of an unknown parameter (usually denoted by θ) and $p(x \mid y)$ is the so-called posterior distribution of the parameter given observed data Y (Week 6).

Some course information
ooooo

Overview of the Monte Carlo method
oooooooooooooooooooo

A few examples
oooooooo●oooooo

What's next?
oo

Example: Meteorology/climate science

- Given precipitation measurements at weather stations, estimate the full precipitation field using some spatio-temporal model.
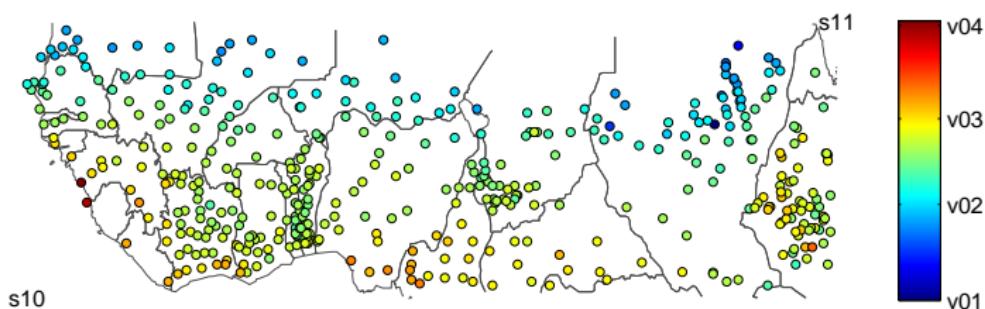
Example: Meteorology/climate science (cont.)

- The accuracy of the interpolated precipitation can be analyzed using Bayesian statistics MCMC methods.
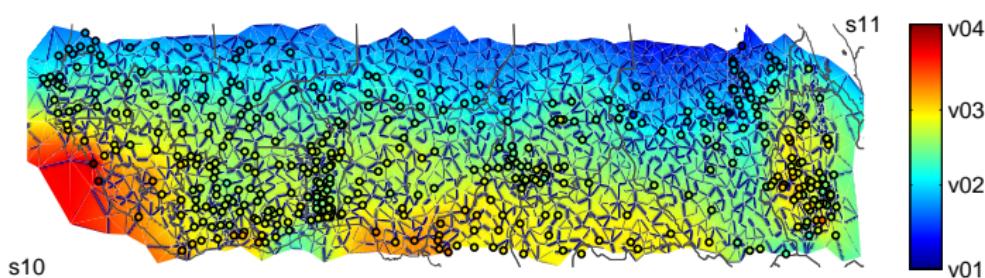
Example: Pricing of contingent claims

- **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 $S \stackrel{\text{def}}{=} (S_t)_{t \geq 0}$ 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

$$\mathcal{X} = \Phi(S_T) = \text{"contract function of } S_T\text{"}.$$

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

Sequential MC problems

- In the sequential Monte Carlo framework, we aim at 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 (which can be very large), 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: Filtering in genetics

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

Chromosome 11 - GM05296

Some course information
00000

Overview of the Monte Carlo method
00000000000000000

A few examples
0000000000000●0

What's next?
00
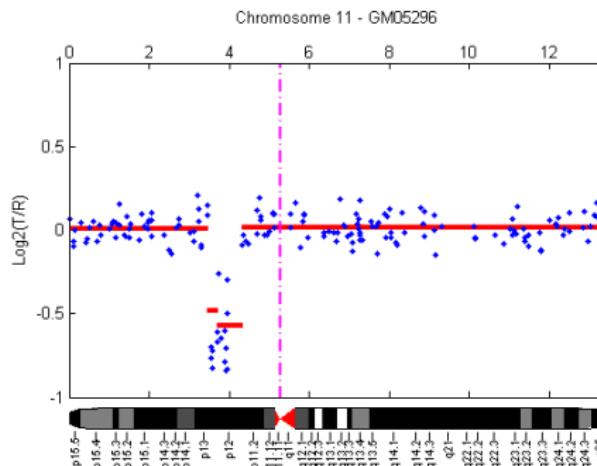
Example: Filtering in target tracking

- An observer obtains noisy observations of the bearing of a
  moving target (such as a submarine). In this HMM, the
  conditional distribution of the target given the observations
  can be estimated online using SMC methods (Weeks 1–2).
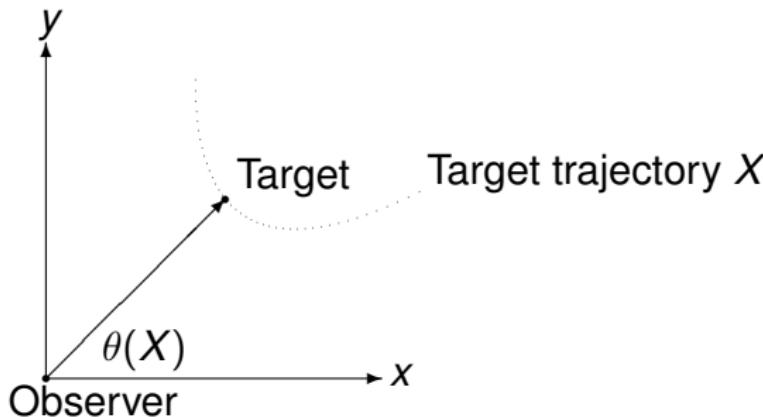
Further fields of application

- At present, MC methods are successfully applied within, among others,
    - mathematical finance and econometrics,
    - physics (nuclear physics, statistical physics, plasma physics, . . .) and astronomy,
    - signal processing,
    - biology (genetics, molecular biology, ecology, . . .),
    - climate sciences,
    - automatic control,
    - . . .

Outline

1 Some course information

2 Overview of the Monte Carlo method
- Preliminaries
- Principal aim
- Interlude: Two fundamental results in statistics
- The basic Monte Carlo sampler

3 A few examples

4 What's next?

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 (i.e. how large should N be)? What can be said about the error and the numerical stability of the algorithm?
    3. Can we exploit problem structure to speed up the computation?

Next lecture

- Next time we will deal with the first two questions and discuss
  - Pseudo-random number generation and
  - MC output analysis.

  See you!