CMPE 58N - Lecture 1.
Monte Carlo methods

Introduction, Course structure, Motivating Examples, Applications

Department of Computer Engineering,

Boğaziçi University, Istanbul, Turkey

Instructor: A. Taylan Cemgil

Spring 2009

Goals of this Course

- Provide a basic understanding of underlying principles of Monte Carlo computation
- Orientation in the literature
- Focus on computational techniques rather than technical details,
  - ... the focus is not on proofs
  - ... but there will be some maths
    - Probability Theory
    - Statistics
    - Calculus and Linear Algebra
- Sharpening your intuition

Topics

- Markov Chain Monte Carlo
- Sequential Monte Carlo
- Probability theory
  - General background
  - Applications

Main study materials

- Handouts, Papers
- Jun S. Liu, Monte Carlo Strategies in Scientific Computing, 2001, Springer.
- Adam M. Johansen and Ludger Evers (edited by Nick Whiteley), Monte Carlo Methods, Lecture notes, University of Bristol

  `http://www.maths.bris.ac.uk/~manpw/teaching/notes.pdf`

- Information Theory, Inference, and Learning Algorithms
  David MacKay, Cambridge University Press – fourth printing (March 2005)

  `http://www.inference.phy.cam.ac.uk/mackay/itprnn/book.html`

General background about probability theory

- Geoffrey Grimmet and David Stirzaker, Probability and Random Processes, (3rd Ed), Oxford, 2006
  - Companion book containing 1000 exercises and solutions
- Grinstead and Snell, Introduction to probability available freely online!

  `http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html`

Main Book on Monte Carlo techniques

- Jun S. Liu,
  Monte Carlo Strategies for Scientific computing,
  Springer 2004
  - Short book
  - Covers almost everything we will mention on MCMC and SMC + more
  - Rather dense and Is not very easy to read

Other Books on Monte Carlo techniques

- Gilks, Richardson, Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman Hall, 1996
- Doucet, de Freitas, Gordon, *Sequential Monte Carlo Methods in Practice*, Springer, 2001

Tutorials and overviews (check course web page)

- Andrieu, de Freitas, Doucet, Jordan. *An Introduction to MCMC for Machine Learning*, 2001
- Andrieu. *Monte Carlo Methods for Absolute beginners*, 2004
- Doucet, Godsill, Andrieu. ”On Sequential Monte Carlo Sampling Methods for Bayesian Filtering”, Statistics and Computing, vol. 10, no. 3, pp. 197-208, 2000

Course Structure

- Web page

  http://www.cmpe.boun.edu.tr/courses/cmpe58N/2009spring/
- Required Work
  - Weekly Assignments (Reading, Programming, Analytic Derivations)
  - A project proposal and outline
  - Final Project: Implementation and Report
- Testing
  - 1 Midterm (in class), 1 Final (take home)
- Grading
  - Relative weights
    - % 25 Midterm
    - % 25 Take home final exam
    - % 50 Assignments, Quizzes and Final Project

Possible Topics

- In one application area (including but not limited to)
  - Scientific data analysis (DNA, Bioinformatics, Medicine, Seismology)
  - Robotics, Navigation, Self Localisation
  - Signal, Speech, Audio, Music Processing
  - Computer Vision (Object tracking)
  - Information Retrieval, Data mining, Text processing, Natural Language Processing
  - Sports, Finance, User Behaviour, Cognitive Science e.t.c.
- Reading a paper and writing a tutorial-like summary in own words and self designed examples
- Implementation and comparative study of inference algorithms on synthetic data

Remarks

- If you have already chosen a research topic
  - Use the project of this course to implement and write up a component of your work!
- If you have **not** chosen research/thesis topic but roughly have something in mind or simply don't know yet
  - Come and talk to me to clarify a topic/technique
  - Study/learn a few inference techniques more in depth
    - Never underestimate the insight gained from a well designed toy example
  - Investigate the feasibility/suitability of Monte Carlo techniques for your purposes

Remarks

- Ideally, a good report could be presented with some extensions at a national or international conference
  - Some well-known methods were master theses once,
  - Occasions when a forth year project report was published (and cited later!)
- Use TeX or LaTeX.
  - If you are serious with research in computer science, statistics or engineering but are using other ways of document preparation, it is very likely that you are wasting some of your valuable time.

Remarks

- Any programming language or other system for computation and visualisation
  - Matlab (preferred)
  - Octave
  - Java,
  - C/C++, BLAS, ATLAS, GNU Scientific Library

Monte Carlo Methods

- Represent the solution of a problem as a parameter of a hypothetical population,
- use a pseudo-random sequence of numbers to construct a sample of a population, from which statistical estimates of the parameter can be obtained
- Stochastic Simulation or Sampling methods

History of Monte Carlo methods

1733 Buffon's needle problem.

1812 Laplace suggests using Buffon's needle experiment to estimate π.

1946 ENIAC (Electronic Numerical Integrator And Computer) built.

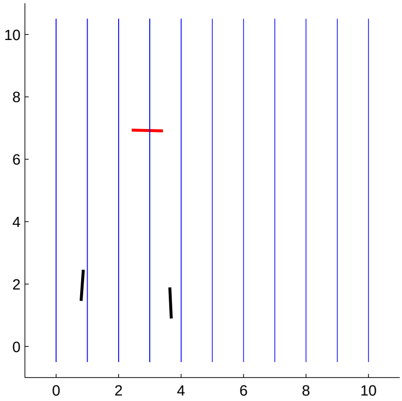
1947 John von Neuman and Stanislaw Ulam propose a computer simulation to solve the problem of neutron diffusion in fissionable material.

1949 Metropolis and Ulam publish their results in the Journal of the American Statistical Association.

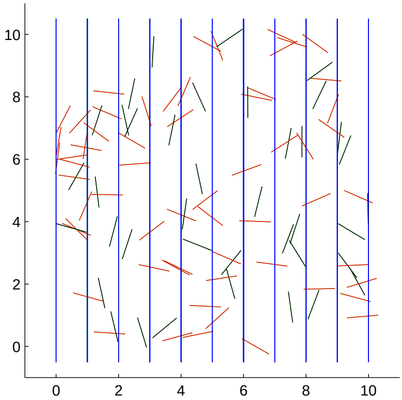
1984 Geman & Geman publish their paper on the Gibbs sampler

... the rest is history

Buffon's needle



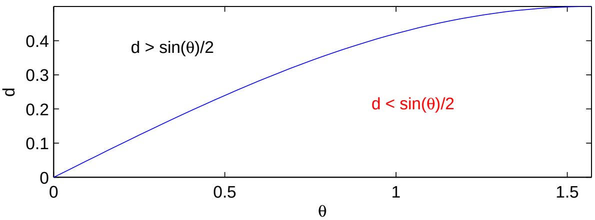
Buffon’s needle



Buffon's needle

- d : Distance from the middle of the needle to the nearest line
- θ : Acute angle between the parallel lines and the needle
- A needle touches a line iff

$$\frac{d}{\sin\theta} < \frac{1}{2}$$



Buffon's needle

- The area of the rectangle is

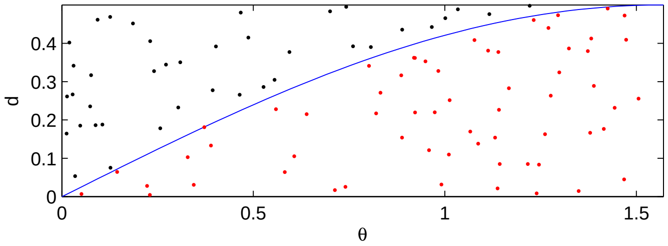
$$S = \frac{1}{2}\frac{\pi}{2}$$

- The area under the *sin* is

$$\int_0^{\pi/2} \sin(\theta)/2 \quad = \quad \frac{1}{2}$$

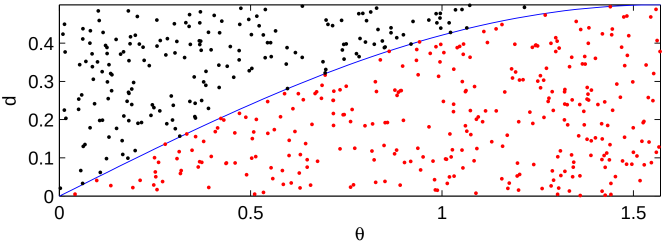
- $$\Pr\{d < \sin(\theta)/2\} \quad = \quad \frac{1/2}{\pi/4} = \frac{2}{\pi}$$

Buffon's needle



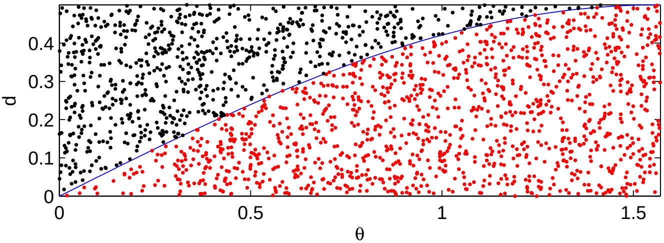
$$\pi \approx 3.2787$$

Buffon's needle



$$\pi \approx 3.1949$$

Buffon's needle



$$\pi \approx 3.1596$$

Indicator function

$$\mathbb{I}\{x\} = \begin{cases} 1 & x \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

Alternative notation: Iverson convention

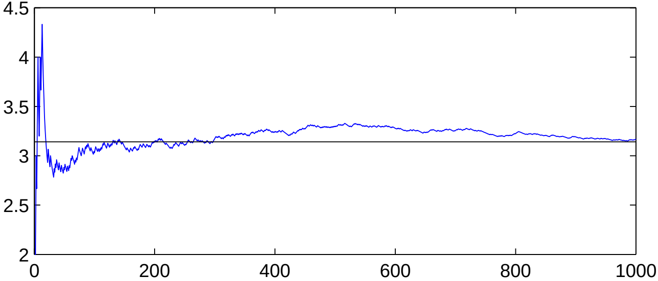
$$[x] = \begin{cases} 1 & x \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

Buffon's needle

- Draw $(d^{(n)}, \theta^{(n)}) \sim U_S$ and estimate π via

$$
\begin{aligned}
\pi &= \frac{2}{\Pr\{d < \sin(\theta)/2\}} \approx \frac{2\text{\# of all dots}}{\text{\# of red dots}} \\
&= \frac{2N}{\sum_{n=1}^{N} \mathbb{I}\{d^{(n)} < \sin(\theta^{(n)})/2\}}
\end{aligned}
$$

Buffon's needle



Speed of convergence

- Monte Carlo integration: error behaves as $n^{-1/2}$.
- Numerical integration of a one-dimensional function by Riemann sums: error behaves as n^{-1}.
- For one-dimensional problems Riemann is better; however deteriorates with increasing dimension: curse of dimensionality.
- Order of convergence of Monte Carlo integration is **independent of the dimension of the problem**. $\rightsquigarrow$ Monte Carlo methods can be a good choice for high-dimensional integrals.

Convergence of random variables

(Liu, Appendix A.1.4.)

$$y_n \sim p_n(y_n) \qquad F_n(y_n) = \int_{-\infty}^{y_n} p_n(\tau) d\tau$$

1 Convergence in distribution

$$\lim_{n\to\infty} F_n(y_n) = F(y)$$

2 Convergence in probability

$$\lim_{n\to\infty} \Pr(|y_n - y| > \epsilon) = 0$$

3 Convergence almost surely

$$\Pr(\lim_{n\to\infty} |y_n - y| = 0) = 1$$

- $3 \Rightarrow 2 \Rightarrow 1$

Convergence of Random variables

- Convergence of random variables is a delicate subject
- Important to get a deeper understanding
- Not get intimidated while reading the literature; remember the definitions and different modes of convergence
- See, e.g., Grimmet and Stirzaker, Ch. 7

Law of Large Numbers

$X_1, \dots, X_n, \dots$ are i.i.d.

- Weak Law: $\langle X_i \rangle = \mu$

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \quad \text{in probability}$$

- Strong Law: $\langle X_i \rangle = \mu$ and X_i with finite variance

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \quad \text{a. s.}$$

Central Limit Theorem

X_i are i.i.d. with mean μ and variance σ^2

- $$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

  $$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to \mathcal{N}(0, 1)$$

- We have approximately

  $$\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$$

Chevalier de Méré

- The famous letters between Pascal and Fermat (start of probability) mention a request for help from a French nobleman and gambler, Chevalier de Méré.
- Méré bets for:

  in four rolls of a die, at least one six would turn up
- Later he bets for:

  in 24 rolls of two dice, a pair of sixes would turn up.

  but he was not happy with the latter schema

Chevalier de Méré

- Setup a computer simulation for a single die

```
K = 4; % Number of dice throws
N = 1000; % Number of games
for trial=1:10,
    D = ceil(rand(N,K)*6);
    disp(sum(sum(D==6, 2) > 0)/N)
end
```

- Per game, Méré won

```
    0.4950, 0.4950, 0.5090, 0.5210, 0.5460
    0.5420, 0.5360, 0.5160, 0.5210, 0.5010
```

Chevalier de Méré

The analytical solution

$$\begin{aligned}\Pr\{\text{Méré wins}\} &= 1 - \Pr\{\text{Méré loses}\} \\ &= 1 - (5/6)^4 = 0.5177\end{aligned}$$

Chevalier de Méré

- Setup a computer simulation for a pair of dice

```
K = 24; % Number of dice throws
N = 1000; % Number of games
for trial=1:10,
    D = ceil(rand(N,K,2)*6);
    sum(sum(D(:,:,1)==6 & D(:,:,2)==6,2) > 0)/N
end
```

- Per game, Mere wins

```
0.502,  0.486, 0.497, 0.533, 0.521
0.474,  0.451, 0.508, 0.470, 0.481 ...
```

- Accurate results by simulation require a large number of experiments

Chevalier de Méré

The analytical solution

$$\Pr\{\text{Méré wins}\} \;=\; 1 - (35/36)^{24} = 0.4914$$

Therefore, 24 times is not a good bet. But with 25 (Pascal)

$$\Pr\{\text{Méré wins}\} \;=\; 1 - (35/36)^{25} = 0.5055$$

Chevalier de Méré

- What is the distribution of the estimate for N games ?
- V_n the outcome that Méré wins the n'th game

$$
\begin{aligned}
V_n &\sim \mathcal{BE}(V_n; p) \\
S_n &= \frac{V_1 + \cdots + V_n}{n}
\end{aligned}
$$

- Evoke the law of large numbers $\langle V_n \rangle = p$

$$
S_n \rightarrow p \qquad n \rightarrow \infty
$$

Chevalier de Méré

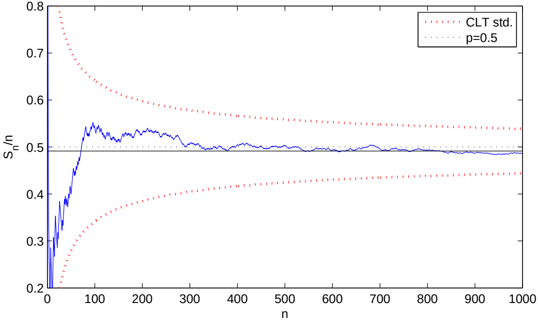
- Accuracy is given by the Central Limit Theorem

$$
\begin{aligned}
\langle V_n \rangle &= p \\
Var\{V_n\} &= p(1-p) \\
\sqrt{\frac{n}{p(1-p)}}(S_n - p) &\to \mathcal{N}(0,1)
\end{aligned}
$$

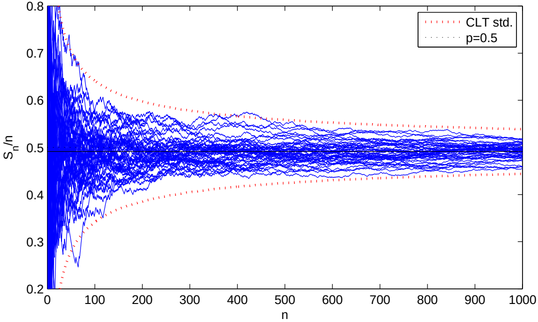
- Approximately

$$S_n \sim \mathcal{N}(p, p(1-p)/n)$$

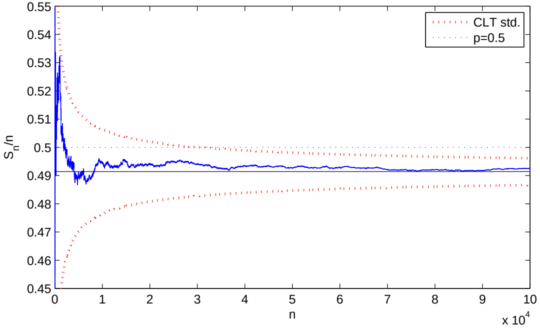
Chevalier de Méré



Chevalier de Méré (cont.)



Chevalier de Méré



- We need around 30000 games to say with about $\%99$ confidence that the game with 24 throws is truly unfavorable.

Applications of Monte Carlo

- Statistics, Bioinformatics
- Signal Processing, Machine learning
- Seismology, Acoustics
- Computer Science, Analysis of algorithms, Randomized algorithms
- Networks, System simulation
- Robotics, Tracking, Navigation
- Econometrics, Finance
- Operations Research, Combinatorics, Optimisation
- Physics, Chemistry, Computational Geometry
- Environmental sciences, monitoring

Bayesian Statistics

- Computation of analytically intractable high dimensional integrals
- Inference, Model selection

Probabilistic Inference

- **expectations** of functions under probability distributions: **Integration**

$$\langle f(x) \rangle = \int_{\mathcal{X}} dx p(x) f(x) \qquad \langle f(x) \rangle = \sum_{x \in \mathcal{X}} p(x) f(x)$$

- **modes** of functions under probability distributions: **Optimization**

$$x^* = \operatorname*{argmax}_{x \in \mathcal{X}} p(x) f(x)$$

- However, computation of multidimensional integrals is hard

Combinatorics

- Counting

Example : What is the probability that a solitaire laid out with 52 cards comes out successfully given all permutations have equal probability ?

$$|A| = \sum_{x \in \mathcal{X}} [x \in A] \qquad [x \in A] \equiv \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

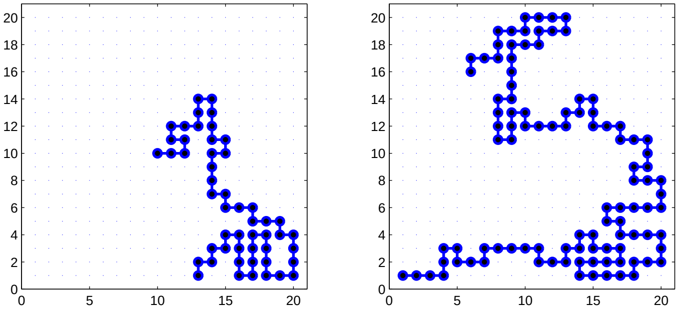
$$p(x \in A) = \frac{|A|}{|\mathcal{X}|} = \frac{?}{\approx 2^{225}}$$

Random Combinatorial Objects

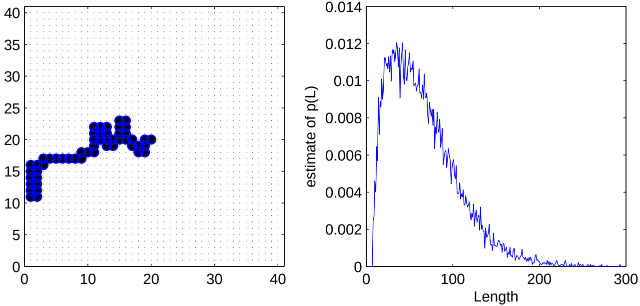
Generate uniformly from

- Self avoiding random walks on a $N \times N$ grid
- All spanning trees of a graph
- Binary trees with N nodes
- Directed Acyclic Graphs

Random Combinatorial Objects



Random Combinatorial Objects

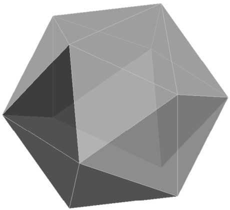


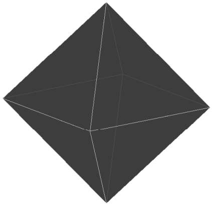
Geometry

- Given a simplex S in N dimensional space by

$$S = \{x : Ax \leq b, \quad x \in \mathbb{R}^N\}$$

find the Volume $|S|$



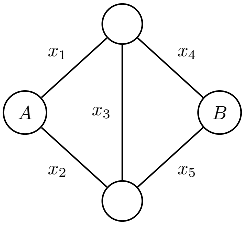


Network analysis, Rare Events

- Given a graph with random edge lengths

$$x_i \sim p(x_i)$$

Find the probability that the **shortest path from A to B** is larger than γ.



Random Number Generation

- Physical methods
  - throw dice, flip coins, shuffle playing cards, roulette wheel
  - thermal noise in Zener diodes or other analog circuits
  - Listen to atmospheric noise (www.Random.org)
  - Run a hash function against a frame of a video stream
  - ...
- A random number generator (deterministic computation) to obtain numbers that "look" random
  - Efficient
  - Repeatable (seeds) – good for debugging

Pseudo-random number generator

- Linear congruential generator

$$
\begin{aligned}
Z_i &= (aZ_{i-1} + b) \mod M \\
X_i &= Z_i / M
\end{aligned}
$$

main flaw: the crystalline nature

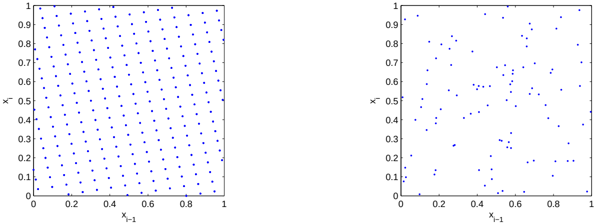


Figure: (left) $a = 81; c = 35; M = 256;$, (right) Matlab's `rand`

Remarks

A poor design which was very popular during 1970's
$a = 2^{16} + 3; c = 0; M = 2^{31}$

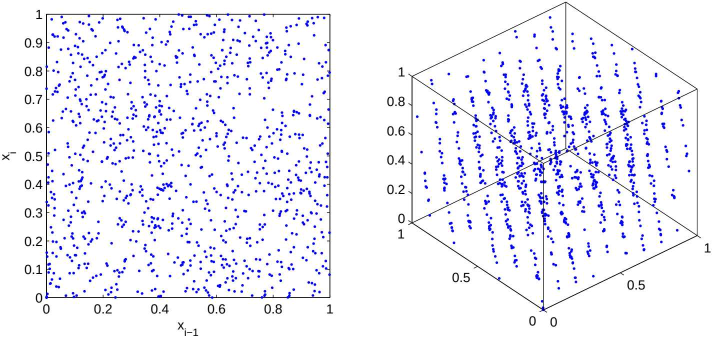


Figure: (left) (X_{i-1}, X_i), (right) (X_{i-2}, X_{i-1}, X_i)

Bernoulli Random Variables

- $x \in \{0, 1\}$

$$x \sim \mathcal{BE}(x; p) = p^x (1 - p)^{(1-x)},$$

- How to sample from a Bernoulli distribution on a computer given $p \in \mathbb{R}$ using samples from the uniform distribution $u \sim \mathcal{U}(u; 0, 1)$?

$$u \sim \mathcal{U}(u; 0, 1)$$
$$x = u < p$$

Note that this is an idealisation as we can not represent irrational numbers on a computer.

The Knuth-Yao algorithm

- How to sample exactly from a Bernoulli distribution $x \sim \mathcal{BE}(x; p)$, $x \in \{0, 1\}$ on a computer given $p \in \mathbb{R}$ using a random bit source $\omega \sim \mathcal{BE}(\omega; 1/2)$?

The Knuth-Yao algorithm

Represent p in binary $p = 0.p_1p_2p_3\dots$

$i \leftarrow 0$

Repeat

$i \leftarrow i + 1$

$\omega \sim \mathcal{BE}(\omega; 1/2)$

Until $\omega \neq p_i$

$x \leftarrow \omega < p_i$

See Devroye, Ch. 15