CMPE 58N - Lecture 1.
Monte Carlo methods

Introduction to Monte Carlo method, Motivating Examples, Law of Large Numbers, Central Limit Theorem

Department of Computer Engineering,

Boğaziçi University, Istanbul, Turkey

Instructor: A. Taylan Cemgil

Outline

- Introduction to Monte Carlo method,
- Motivating Examples,
- Law of Large Numbers,
- Central Limit Theorem
- Example

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.

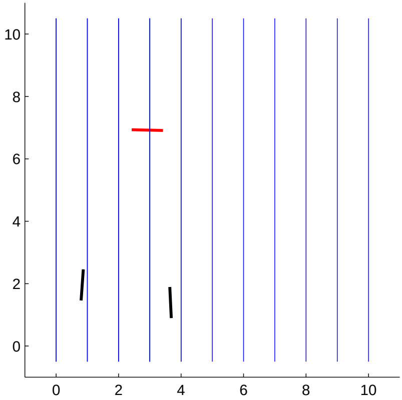
1947 John von Neuman and Stanislaw Ulam propose a computer simulation to solve the problem of neutron diffusion in fissionable material.

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

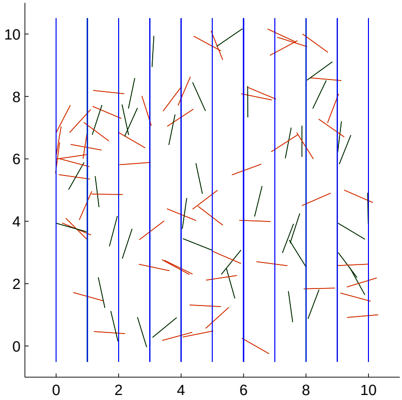
1984 Geman & Geman publish their paper on the Gibbs sampler

… continuously growing interest with increases in computational power

Buffon's needle



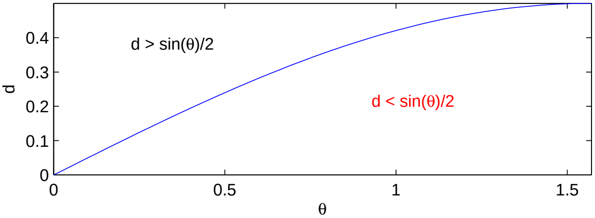
Buffon’s needle



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

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

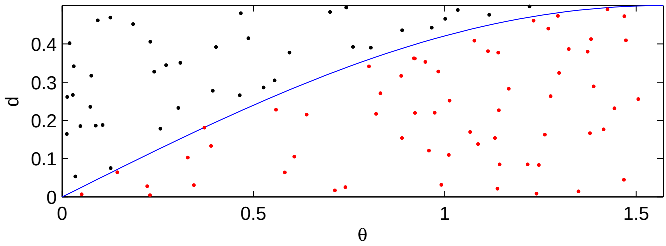
- The area under the *sin* is

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

-

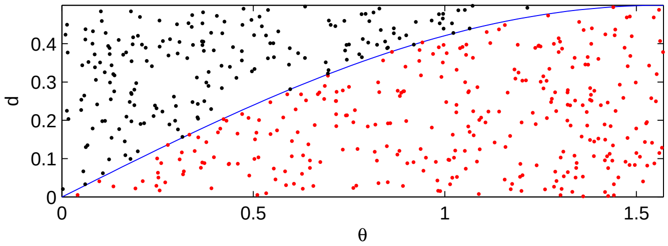
$$\Pr\{d < \sin(\theta)/2\} \quad = \quad \frac{1/2}{\pi/4} = \frac{2}{\pi}$$

Buffon's needle



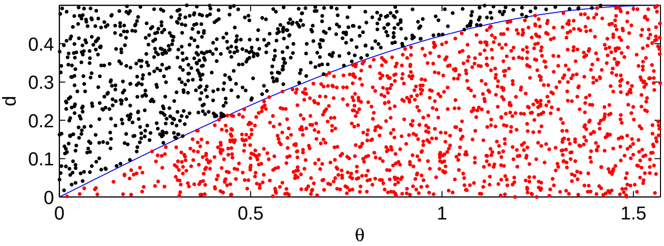
$$\pi \approx 3.2787$$

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

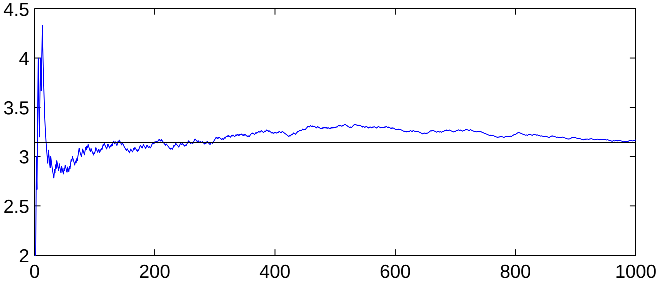
$$[x] = \begin{cases} 1 & x \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

Buffon's needle

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

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

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

$X_1, \ldots, X_n, \ldots$ 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

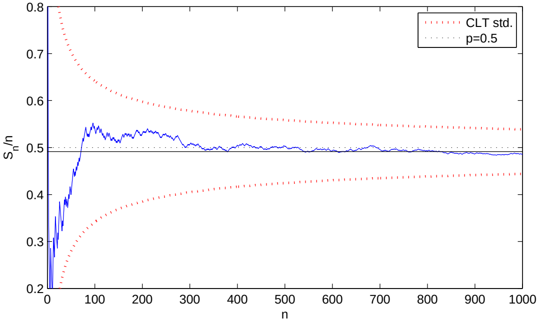
- Accuracy is given by the Central Limit Theorem

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

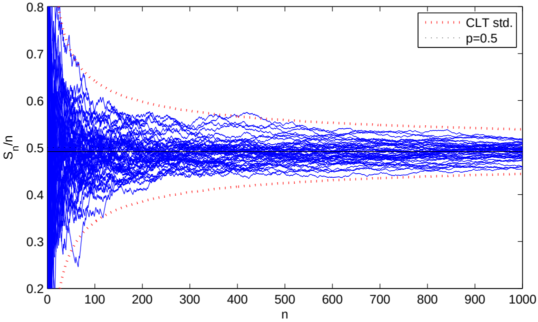
- Approximately

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

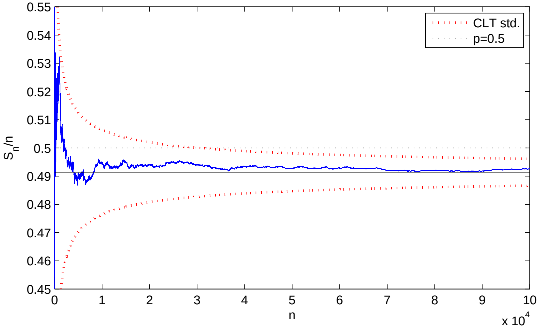
Chevalier de Méré



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.

Summary

- Law of large numbers: Consistency.
- CLT: Provides information about the rate of convergence
- If we can draw N independent and identically distributed samples from a distribution $p(x)$, we can estimate expectations $E_p\left(\varphi(x)\right)$ with an error $O(N^{1/2})$, **independent** of the dimensionality of x.