Monte Carlo Methods (I)

俞 声

清华大学统计学研究中心

Monte Carlo methods

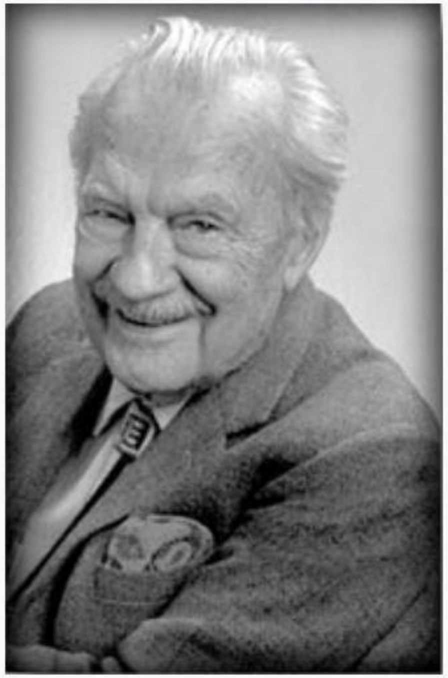
- A broad class of computational algorithms that rely on repeated random sampling to obtain numerical results.
- Use randomness to solve problems that might be deterministic in principle, but would be difficult to solve by other approaches.
- Commonly used in three classes of problems:
  - Integration
  - Generating samples from a probability distribution
  - Optimization

Who is Monte Carlo?

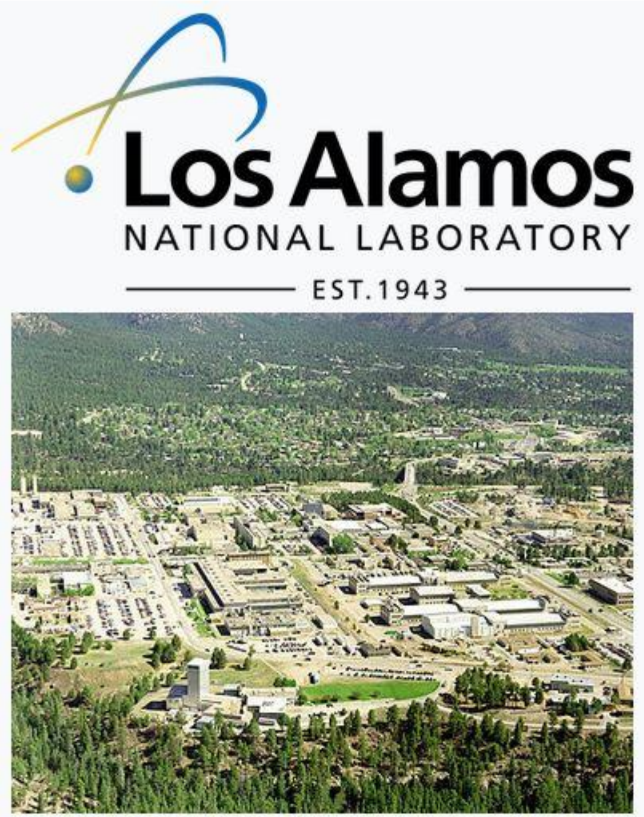
Stanisław Ulam

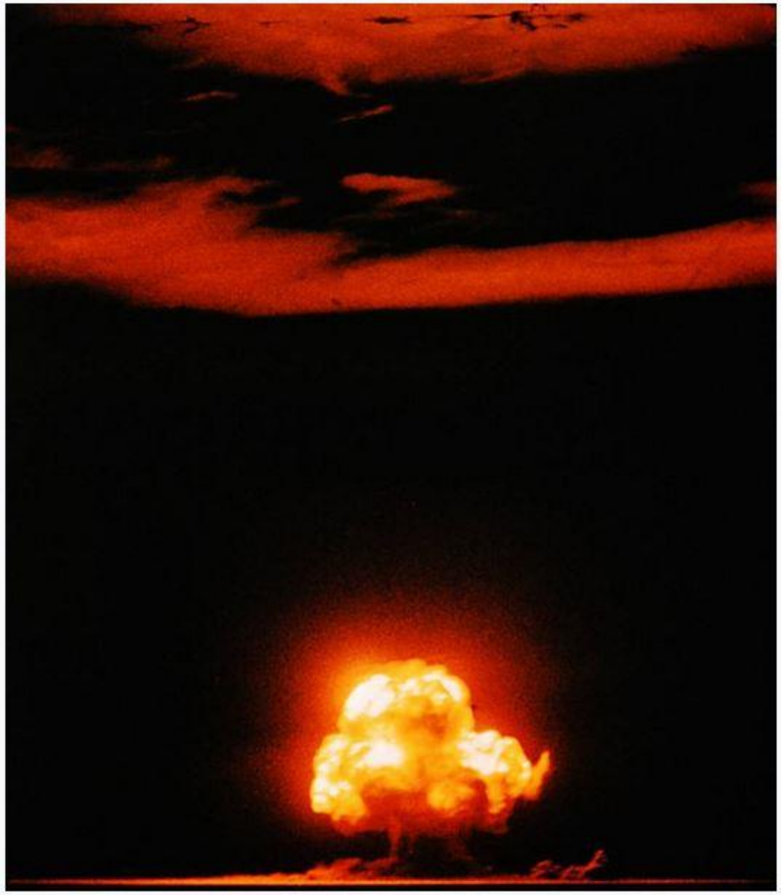
John von Neumann

Nicholas Metropolis



Who is Monte Carlo?





The Trinity test of the Manhattan Project was the first detonation of a nuclear weapon.

Who is Monte Carlo?

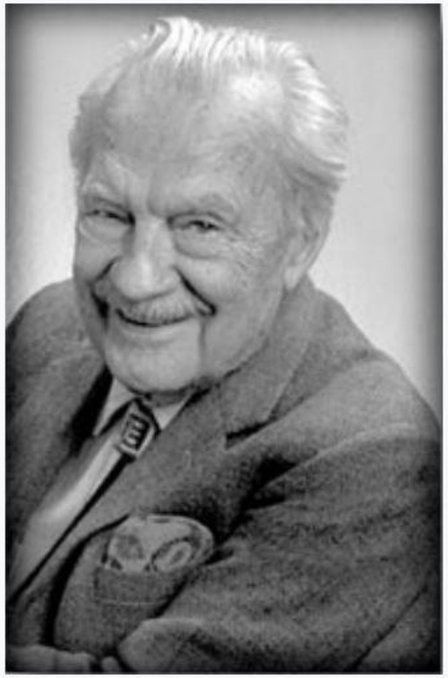
Stanisław Ulam

Proposed using random experiments

John von Neumann

Invented a way to generate pseudorandom numbers

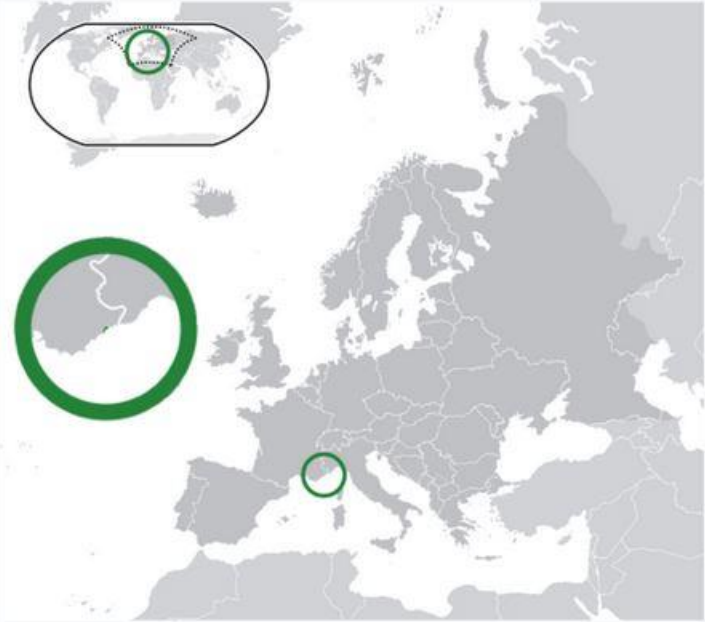
Nicholas Metropolis



Invented the Metropolis sampler and gave the project the code name “Monte Carlo”

Who is Monte Carlo?

Casino de Monte-Carlo



Location of Monaco (green)
in Europe (green & dark grey)

Monte Carlo (literally "Mount Charles"), is an administrative area of the Principality of Monaco.

Who is Monte Carlo?

Panoramic view of Monaco from the Tête de Chien in 2017

Area of Monaco ~ 499 acres
Area of Tsinghua ~ 1000 acres

Monte Carlo integration

Integration $\Rightarrow$ *sample mean:*

Let $g(x)$ be a function and suppose that we want to compute $\int_a^b g(x)dx$.
Recall that if X is a random variable with density $f(x)$, then the mathematical expectation of the random variable $g(X)$ is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx\,.$$

If a random sample is available from the distribution of X, an unbiased estimator of $E[g(X)]$ is the sample mean.

A simple case

- Consider the problem of estimating $\theta = \int_0^1 g(x)dx$.

- If $X_1, \ldots, X_m$ is a random Uniform(0,1) sample, then
$$\hat{\theta} = \overline{g_m(X)} = \frac{1}{m}\sum_{i=1}^{m} g(X_i)$$
converges to $E[g(X)] = \theta$ with probability 1, by the *Strong Law of Large Numbers*.

- The simple Monte Carlo estimator of $\int_0^1 g(x)dx$ is $\overline{g_m(X)}$.

Exercise

Compute a Monte Carlo estimate of $\int_0^1 e^{-x}dx$ and compare the estimate with the exact value.

More general cases

To compute $\theta = \int_a^b g(x)dx$, we can replace the Uniform(0,1) density with Uniform(a,b):

$$\int_a^b g(x)dx = (b-a)\int_a^b g(x)\frac{1}{b-a}dx$$

The integral is therefore $(b-a)$ times the average value of $g(\cdot)$ over (a,b).

Exercise

Compute a Monte Carlo estimate of $\int_2^4 e^{-x} dx$ and compare the estimate with the exact value.

Exercise

Use the Monte Carlo approach to estimate the standard normal cdf

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

- the uniform distribution approach
- the normal distribution approach

Importance sampling

- The estimate $\frac{b-a}{m}\sum_{i=1}^{m} g(X_i)$ with uniformly distributed $\{X_i\}$ converges to $\int_a^b g(x)dx$ with probability 1 by the strong law of large numbers.

- One limitation of this method is that it does not apply to unbounded intervals.

- Another drawback is that it can be inefficient to draw samples uniformly across the interval if the function $g(x)$ is not very uniform.

- It seems reasonable to consider other densities than uniform. This leads us to a general method called *importance sampling*.

Importance sampling

- Suppose X is a random variable with density function $f(x)$, such that $f(x) > 0$ on the set $\{x : g(x) > 0\}$. Let Y be the random variable $g(X)/f(X)$. Then

$$\int g(x)dx = \int \frac{g(x)}{f(x)} f(x)dx = E[Y].$$

- Estimate $E[Y]$ by simple Monte Carlo integration. That is, compute the average

$$\frac{1}{m}\sum_{i=1}^{m} Y_i = \frac{1}{m}\sum_{i=1}^{m} \frac{g(X_i)}{f(X_i)},$$

where the random variables X_i are generated from the distribution with density $f(x)$. The density $f(x)$ is the called the *importance function*.

Importance sampling

- Question: what is the variance of the importance sampling estimate?

- Conclusion: choose $f(x)$ to make $g(x)/f(x)$ close to a constant.

Bayesian inference

- From a Bayesian perspective, in a statistical model both the observables and the *parameters are random*.

- The parameters θ has *prior* distribution $f_\theta(\theta)$.

- The distribution of X depends on θ. The likelihood of the observed data $x = \{x_1, \ldots, x_n\}$ given θ is $f(x_1, \ldots, x_n \mid \theta)$.

Bayesian inference

- Once data is observed, one can update the distribution of θ conditional on the information in the sample x. The *posterior* distribution of θ given x is

$$f_{\theta|x}(\theta \mid x) = \frac{f_{x|\theta}(x \mid \theta) f_\theta(\theta)}{f_x(x)} = \frac{f_{x|\theta}(x \mid \theta) f_\theta(\theta)}{\int f_{x|\theta}(x \mid \theta) f_\theta(\theta) d\theta}.$$

- Then a point estimate for $g(\theta)$ could be

$$E_{\theta|x}[g(\theta)] = \int g(\theta) f_{\theta|x}(\theta \mid x) d\theta = \frac{\int g(\theta) f_{x|\theta}(x \mid \theta) f_\theta(\theta) d\theta}{\int f_{x|\theta}(x \mid \theta) f_\theta(\theta) d\theta}.$$

MCMC integration

- Markov Chain Monte Carlo (MCMC) integration is a popular way to compute $E_{\theta|x}[g(\theta)]$.

- We have learned that we can approximate $E_{\theta|x}[g(\theta)] = \int g(\theta) f_{\theta|x}(\theta \mid x) d\theta$ by generating random samples $Y_1, \ldots, Y_m$ from the distribution $f_{\theta|x}(\theta \mid x)$, then
$$\bar{g} = \frac{1}{m} \sum_{i=1}^{m} g(Y_i)$$
is an estimate of $E_{\theta|x}[g(\theta)]$. This explains the second "MC" in "MCMC".

- The remaining problem is: $f_{\theta|x}(\theta \mid x) = \frac{f_{x|\theta}(x|\theta) f_\theta(\theta)}{\int f_{x|\theta}(x|\theta) f_\theta(\theta) d\theta}$ is hard to compute. How do we generate samples from this distribution?

Random sample generation

- Generating simple univariate random samples is easy:
  - Generate a uniform sample $u = \{u_1, u_2, \dots, u_n\}$ from Uniform(0,1);
  - Let $x = F^{-1}(u)$, where F is the cdf of the target distribution;
  - Then x is a random sample from F.

Rejection sampling

Another method is called rejection sampling, or the acceptance-rejection method, which doesn' t require calculating $F^{-1}(x)$.

Find a distribution with density g satisfying $f(t) \leq c \cdot g(t)$, for all t such that $f(t) > 0$.

1. Generate a random y from the distribution with density g.
2. Generate a random u from the Uniform(0, 1) distribution.
3. If $u < \frac{f(y)}{c \cdot g(y)}$, accept y and output $x = y$; otherwise reject y. Repeat 1-3.

Rejection sampling

- Given Y, the probability of acceptance is

$$P(\,accept \mid Y\,) = P\left(U < \frac{f(Y)}{c \cdot g(Y)} \,\middle|\, Y\right) = \frac{f(Y)}{c \cdot g(Y)}.$$

- The overall probability of acceptance is

$$P(accept) = \int P(\,accept \mid y\,)g(y)dy = \int \frac{f(y)}{c \cdot g(y)}g(y)dy = \frac{1}{c}.$$

Thus c should be small.

Rejection sampling

- To see that the accepted sample has density f, apply Bayes' Theorem.

$$pdf(x \mid accepted) = \frac{P(accepted \mid x)g(x)}{P(accepted)} = \frac{\left(\frac{f(x)}{c \cdot g(x)}\right) g(x)}{1/c} = f(x)$$

Exercise

- Generate random samples that follows the Beta(2, 2) distribution, whose density function is $f(x) = 6x(1 - x)$, $0 < x < 1$.

- On average, how many iterations will be required to generate 1000 samples?