Monte Carlo Methods

Nando de Freitas

University of British Columbia

Overview

- Introduction
  - What is Monte Carlo?
  - History
  - Rejection sampling
  - Importance sampling
- Sequential Monte Carlo
- Markov chain Monte Carlo

Why Monte Carlo?

- Integration

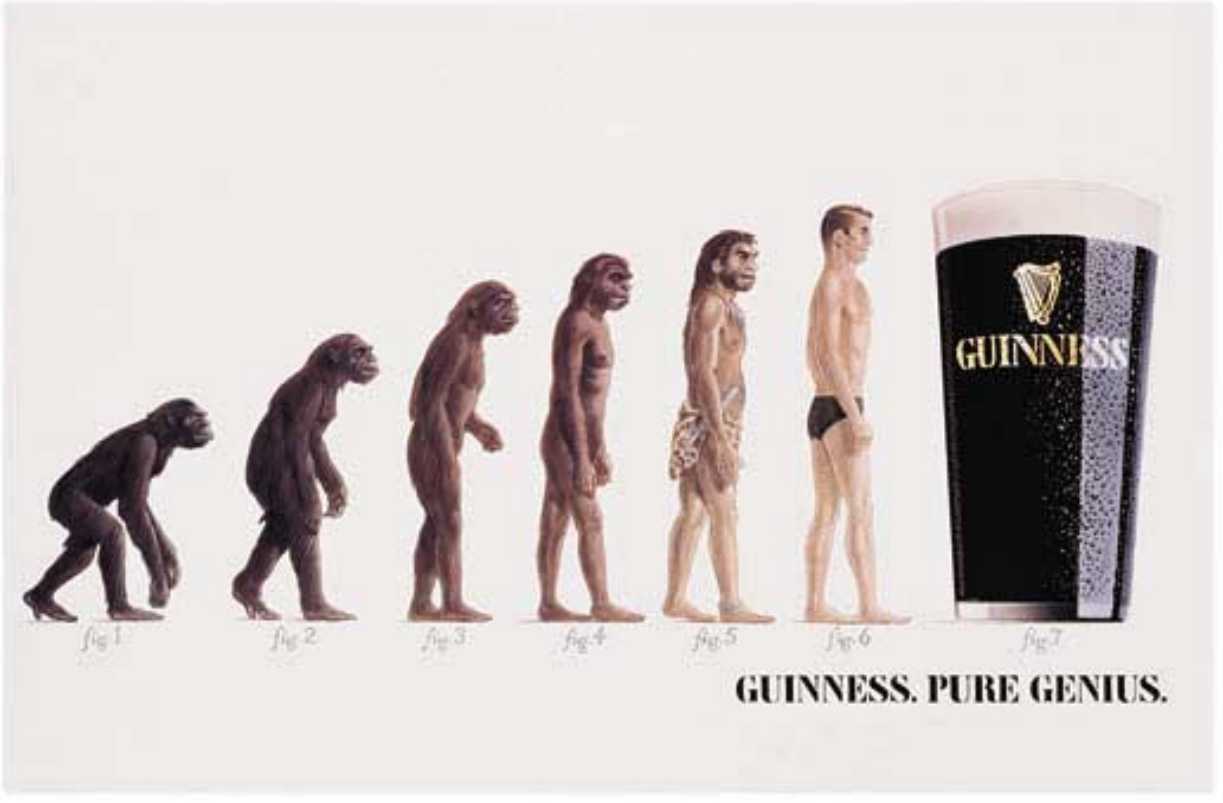
Why Monte Carlo?

- Integration
- Optimisation

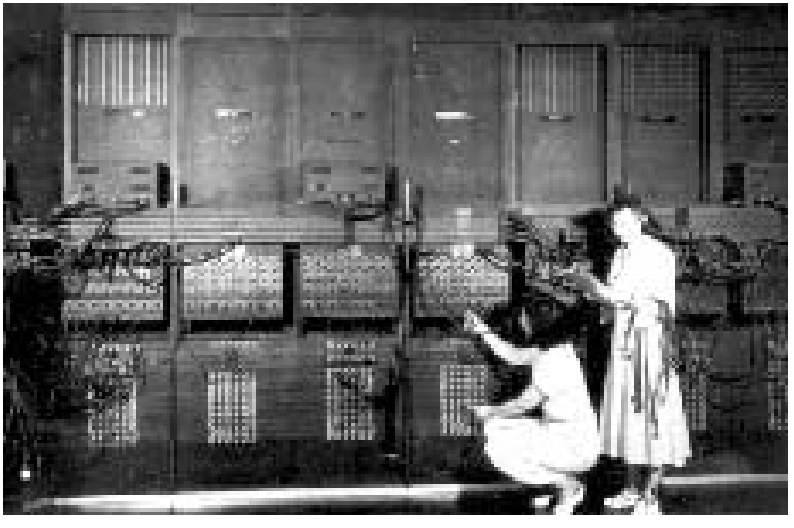
Why Monte Carlo?

- Integration
- Optimisation
- Simulation

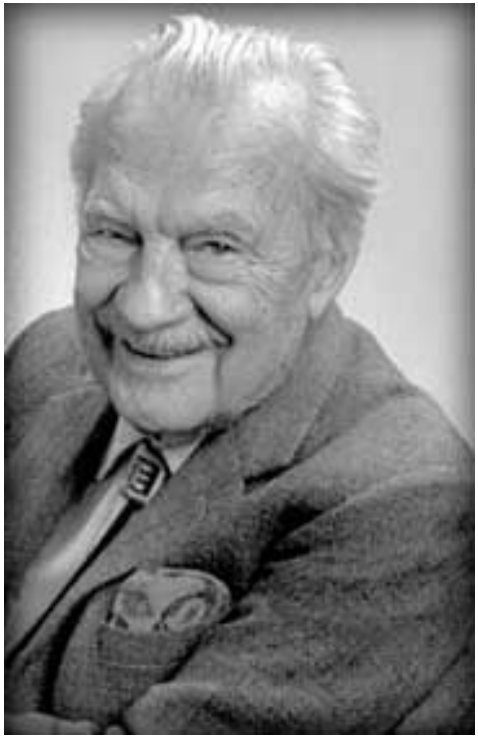
William Gosset (aka Student)



The Bomb and the ENIAC



Metropolis, Ulam and von Neumann



Enrico Fermi and the FERMIAC

History of Modern Monte Carlo

- 1949 Metropolis and Ulam publish the first paper.
- 1953 The Metropolis algorithm.
- 1970 The Metropolis-Hastings generalisation.
- 1984 The Gibbs sampler becomes popular.
- 1990 Statisticians learn about it.
- Renaissance.

A Few Applications

- Sophisticated statistical modelling.
- Queries on the web.
- Tracking.
- Econometrics.
- Probabilistic graphical models.
- Control and Communications.
- Computer graphics.

Approximating probabilities

. .

Tracking

. .

Realistic graphics

. .

Integration and Probabilistic Inference

1. *Normalisation*:

$$p(x|y) = \frac{p(y|x)p(x)}{\int_X p(y|x')p(x')dx'}$$

Integration and Probabilistic Inference

1. *Normalisation*:

$$p(x|y) = \frac{p(y|x)p(x)}{\int_X p(y|x')p(x')dx'}$$

2. *Marginalisation*:

$$p(x|y) = \int_Z p(x, z|y)dz$$

Integration and Probabilistic Inference

1. *Normalisation*:

$$p(x|y) = \frac{p(y|x)p(x)}{\int_X p(y|x')p(x')dx'}$$

2. *Marginalisation*:

$$p(x|y) = \int_Z p(x, z|y)dz$$

3. *Expectation*:

$$\mathsf{E}_{p(x|y)}(f(x)) = \int_X f(x)p(x|y)dx$$

Statistical Physics

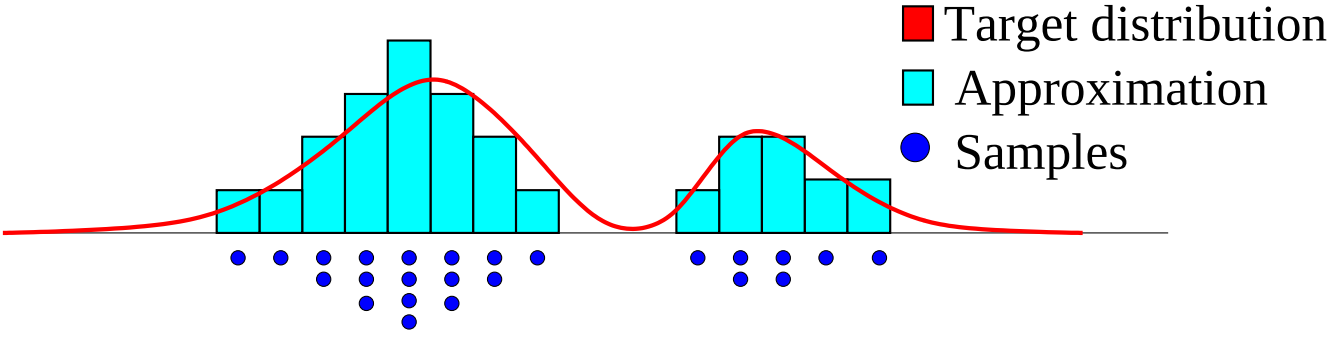
Here, one needs to compute the partition function Z of a system with states s and Hamiltonian $E(s)$

$$Z = \sum_s \exp\left[-\frac{E(s)}{kT}\right],$$

where k is Boltzmann's constant and T denotes the temperature of the system. Summing over the large number of possible configurations is prohibitively expensive.

Monte Carlo Integration

If we have samples $\{x^{(i)}\}_{i=1}^{N}$ distributed according to $p(x|y)$, then



$$\int f(x)p(x|\mathbf{x})dx \quad \text{is approximated with} \quad \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})$$

Monte Carlo Optimisation

- Simple global optimisation.

$$\widehat{x} = \underset{x^{(i)}; i=1,\ldots,N}{\arg\max} \; p\left(x^{(i)}\right)$$

Monte Carlo Optimisation

- Simple global optimisation.

$$\widehat{x} = \underset{x^{(i)}; i=1,\ldots,N}{\arg\max} \; p\left(x^{(i)}\right)$$

- Simulated annealing.

The catch

- We only know how to sample from standard distributions, *e.g.* uniform, multinomial, Gaussian and Gamma distributions.

Rejection Sampling

We can sample from a distribution $p(x)$, which is known up to a proportionality constant, by sampling from another easy-to-sample proposal distribution $q(x)$ that satisfies $p(x) \leq Mq(x)$, $M < \infty$, as follows:

Set $i = 1$

Repeat until $i = N$

1. Sample $x^{(i)} \sim q(x)$ and $u \sim \mathrm{U}_{(0,1)}$.
2. If $u < \frac{p(x^{(i)})}{Mq(x^{(i)})}$ then accept $x^{(i)}$ and increment the counter i by 1. Otherwise, reject.

Importance Sampling

$$
\begin{aligned}
I(f) &= \int f(x)p(x)dx \\
&= \int f(x)\frac{p(x)}{q(x)}q(x)dx \\
&= \int f(x)w(x)q(x)dx
\end{aligned}
$$

By simulating N i.i.d. samples $\{x^{(i)}\}_{i=1}^{N}$ according to $q(x)$ and evaluating $w(x^{(i)})$, we obtain

$$
\widehat{I}_N(f) = \sum_{i=1}^{N} f\left(x^{(i)}\right) w\left(x^{(i)}\right)
$$

Importance Sampling

The IS estimator is unbiased, but has variance

$$\mathrm{var}_{q(x)}\left(\widehat{I}_N(f)\right) = \mathsf{E}_{q(x)}\left(f^2(x)w^2(x)\right) - I^2(f)$$

This variance is minimised when

$$q^\star(x) = \frac{|f(x)|p(x)}{\int |f(x)|p(x)dx}$$

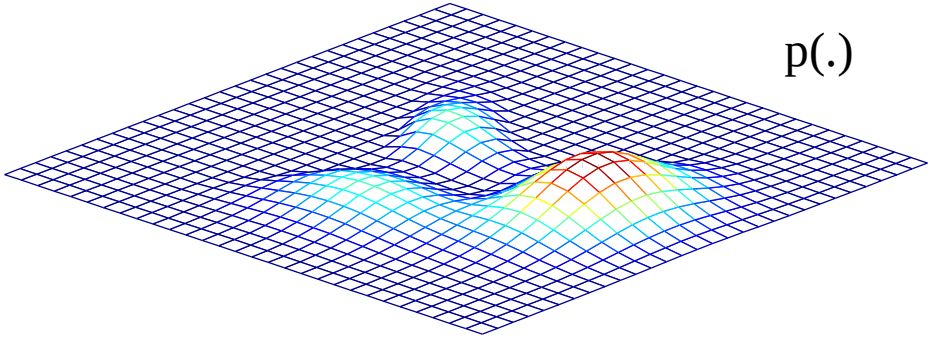
Adaptive Importance Sampling

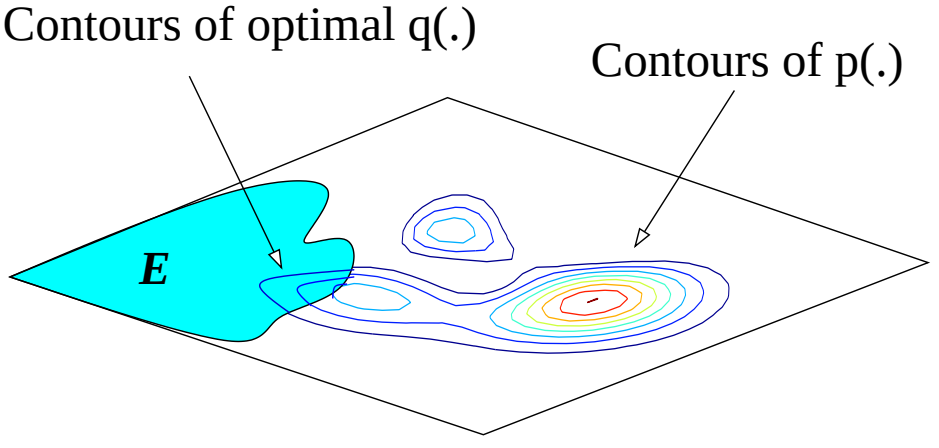
Introduce parametric proposals and adapt the parameters so as to minimise the variance

$$\theta_{t+1} = \theta_t - \alpha \frac{1}{N} \sum_{i=1}^{N} f^2(x^{(i)}) w(x^{(i)}, \theta_t) \frac{\partial w(x^{(i)}, \theta_t)}{\partial \theta_t}$$

where α is a learning rate and $x^{(i)} \sim q(x, \theta)$.

Importance Sampling and Rare Events





Importance Sampling and Rare Events

$$
\begin{aligned}
P(E) &= \int \mathtt{I}_E(x) p(x) dx \\
&= \int \mathtt{I}_E(x) w(x) q(x, \theta^\star) dx
\end{aligned}
$$

The theory of large deviations tells us how to map this problem to a constrained optimisation problem.