Computational Statistics & Machine Learning

Lecture 1

The Monte Carlo Method

Mark Girolami

mag92@cam.ac.uk

Department of Engineering

University of Cambridge

December 19, 2022

Overview

- Introduction and History.
- Evaluating integrals numerically.
- Evaluation of statistical expectations.
- The Central Limit Theorem.
- Rates of convergence for the Monte Carlo method.

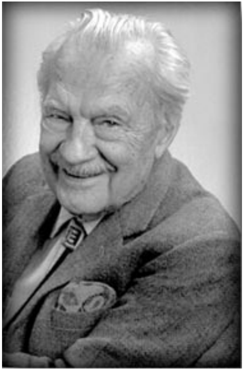
Introduction

- Monte Carlo - an area of the principality of Monaco

- Famous for the casino

Introduction

- Monte Carlo - code-name suggested by Nick Metropolis for Los Alamos research project using randomisation to estimate nuclear diffusion processes for nuclear weapons



Introduction

- The Monte Carlo methods are a broad suite of tools of stochastic simulation and computation
- Monte Carlo methods are ubiquitous with applications in physics, chemistry, biology, engineering, computer science, finance, social sciences, medicine, weather forecasting, to name a few
- Google DeepMind Alpha-Go employs Monte Carlo Tree Search
- Foundational tool in all of Machine Learning

Evaluating Integrals Numerically

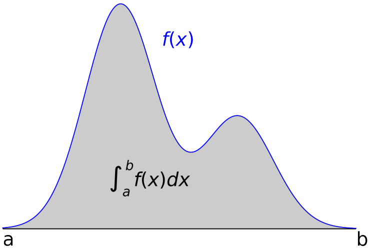
- Consider the one-dimensional Riemann integral

$$\int_0^1 \exp(-x^3)dx$$

- There is no analytic way to evaluate this integral so one resorts to deterministic numerical approaches
- Midpoint Rule $\int_a^b f(x)dx \approx (b-a)f\left(\frac{a+b}{2}\right)$
- Trapezoidal Rule $\int_a^b f(x)dx \approx (b-a)\left(\frac{f(a)+f(b)}{2}\right)$
- Higher order Quadrature rules provide greater accuracy (how do we define this?)
- Generalising to greater than one dimension is not straightforward (why?)

Evaluating Integrals Numerically

- The Monte Carlo approach employing stochastic computation sidesteps the issues with higher-dimensional domains of integration as we shall see later
- Returning to our integral $\int_a^b f(x)dx$, one approach to evaluation is to estimate the fraction of area below the graph, bounded below its maximum value



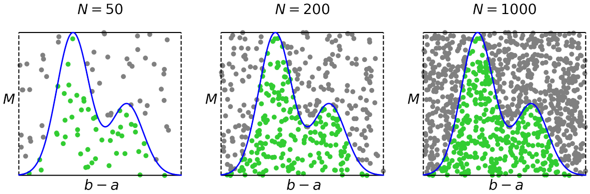
Evaluating Integrals Numerically

- Assume $f(x)$ has a maximum value, say M, within the domain of integration $[a, b]$, then the enclosing rectangle has area $(b - a) \times M$
- We require the fraction of points in the rectangle that lie on or under the graph $f(x)$ to obtain the value of the integral
- We can estimate this fraction as follows
- Pick uniformly at random a point x between a and b and compute $f(x)$
- Pick uniformly at random a point v in between 0 and 1 and use this fraction to scale the maximum M, that is $y = M \times v$
- If y is equal to or less than $f(x)$ then we accept it otherwise we reject the choice
- By repeating this say N times, and counting the total number of accepts, acc, the ratio $\frac{acc}{N}$ approximates the fraction of rectangle area that lies under the graph

Evaluating Integrals Numerically

- Then our stochastic approximation is

$$\int_a^b f(x)dx \approx (b-a) \times M \times \frac{acc}{N}$$



Evaluating Integrals Numerically

- This is what is called a form of Rejection Sampling
- We would hope that there is some form of convergence to the actual value of the integral
- How fast will this convergence be is a critical factor in application
- If $f(x)$ has a sharp narrow peak, or peaks, in the domain of integration there will be many rejections, requiring a large value of N for a good approximation
- The rate of convergence would then seem to be related to the characteristics of the integrand

Evaluating Integrals Numerically

- Numerically the Riemann integral is approximated by taking evenly (deterministically) spaced points in the domain of x_k i.e. $[a, b]$, if there are K points they are separated by $\Delta_k = \frac{b-a}{K}$
- The Riemann integral is then

$$\sum_k \Delta_k f(x_k) = \frac{b-a}{K} \sum_k f(x_k)$$

- This converges to $\int_a^b f(x)dx$ as $K \to \infty$ or $\Delta_k \to 0$

Evaluating Statistical Expectations

- Let us now rewrite the integral in an equivalent form such that

$$\int_a^b f(x)dx = \int_a^b f(x) \times \frac{p(x)}{p(x)} dx$$

- Where $p(x)$ is a probability density function defined on the domain of integration, so now x is treated as a random variable
- We can go further noting that

$$\int_a^b f(x)dx = \int_a^b f(x) \times \frac{p(x)}{p(x)} dx = \mathbb{E}\left\{\frac{f(x)}{p(x)}\right\}$$

- The integral is the statistical expectation under the probability density $p(x)$ of the weighted version of the integrand $\frac{f(x)}{p(x)}$

Evaluating Statistical Expectations

- If $p(x)$ is uniform on $[a, b]$ then $p(x) = \frac{1}{b-a}$ and

$$\int_a^b f(x)dx = (b-a) \times \mathbb{E}\{f(x)\}$$

- Noting that $\mathbb{E}\{f(x)\}$ is an expectation with respect to the uniform on $[a, b]$
- Then

$$\int_a^b f(x)dx = (b-a) \times \mathbb{E}\{f(x)\} \approx \frac{(b-a)}{N} \sum_n f(u_n)$$

- where each u_n is uniform between a and b
- Compare with (deterministic) Riemann integral $\sum_k \Delta_k f(x_k) = \frac{(b-a)}{K} \sum_k f(x_k)$

Central Limit Theorem for Monte Carlo

- In essence, approximations of expectations are obtained via Monte Carlo estimates
- Then

$$\mathbb{E}\{f(X)\} \approx \frac{1}{N}\sum_{n} f(x_n)$$

- where each x_n is distributed according to $p(x)$
- What can be said about this approximation ?
- The Law of Large Numbers tells us that the estimate will converge (in some sense) to the expectation
- The Monte Carlo estimate is unbiased - what does this mean?

Central Limit Theorem for Monte Carlo

- Denoting $S_N = \sum_n f_n$ with each $f_n = f(x_n)$ without loss of generality assume each f_n has a mean of zero and variance σ_f^2.
- The Moment-Generating Function (MGF - from second year Maths) for each $f(x)$, if it exists is $M_f(t) = \mathbb{E}\{\exp(tf(X))\}$
- MGF for $Z_N = \frac{S_N}{\sqrt{N\sigma_f^2}}$ is $M_{Z_N}(t) = \left(M_f\left(\frac{t}{\sqrt{N\sigma_f^2}}\right)\right)^N$
- Truncating Taylor expansion of MGF gives $M_{Z_N}(t) = \left(1 + \frac{t^2}{2N} + \epsilon_N\right)^N$
- As $N \to \infty$ then $M_{Z_N}(t) = \exp\left(\frac{t^2}{2}\right)$ (Check second year maths notes for details)
- This is the MGF of variable that has probability density $\mathcal{N}(0, 1)$

Central Limit Theorem for Monte Carlo

- What this says is that

$$\left(\frac{1}{N}\sum_n f(x_n) - \mathbb{E}\left\{f(X)\right\}\right) \times \frac{\sqrt{N}}{\sigma_f} \sim \mathcal{N}(0,1)$$

- In other words

$$\frac{1}{N}\sum_n f(x_n) \sim \mathcal{N}\left(\mathbb{E}\left\{f(X)\right\}, \frac{\sigma_f^2}{N}\right)$$

- As $N \to \infty$ the Monte Carlo estimate converges to a Normal distributed variable whose mean is the desired expectation, whose standard deviation shrinks at a rate $\frac{\sigma_f}{\sqrt{N}}$
- σ_f is a rate constant, the smoother the function the faster the convergence, we will explore this further later