Monte Carlo Integration and the Metropolis Algorithm

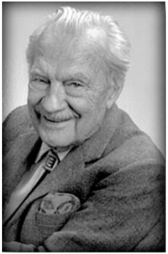
Sohan Vartak

Yale University

December 3, 2018

Goals

- Explain the advantages of Monte Carlo methods
- Motivate the ingredients that go into designing a Monte Carlo algorithm (ergodicity, detailed balance)
- Justify the widespread use of the Metropolis algorithm



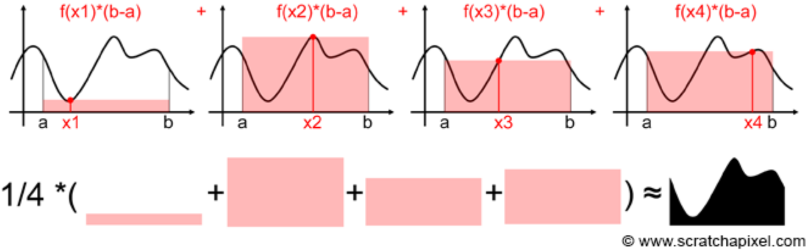
What is Monte Carlo Integration?

Overview of the Method

We want to approximate the integral $I = \int_D f(\mathbf{x})dx$. To do this, we sample points $\mathbf{x}_1, ..., \mathbf{x}_N \in D$ and approximate

$I \approx \frac{Vol(D)}{N} \sum_{i=1}^{N} f(\mathbf{x}_i)$

$\langle \frac{Vol(D)}{N} \sum_{i=1}^{N} f(\mathbf{x}_i) \rangle = \frac{Vol(D)}{N} \sum_{i=1}^{N} \langle f \rangle = \int_D f(\mathbf{x})dx$



- Law of large numbers ensures convergence to correct result (convergence rate is NOT specified...we'll return to this)

Why Use Monte Carlo Integration?

- MC integration is one of the best tools we have for high-dimensional integrals. We can see this by looking at how error scales with number of points N and dimension d
- Simpson's rule error scales like $N^{-4/d}$ where N is the number of integration points and d is the dimension of the integral (very bad for higher dimensions!!)
- Monte Carlo error always scales like $N^{-1/2}$ (independent of dimension!!)

A Brief Proof

We approximate $I = \int f(\mathbf{x})d\mathbf{x} \implies I \approx \frac{Vol(D)}{N}\sum_{i=1}^{N} f(\mathbf{x}_i)$. The variance of this object is then
$\sigma^2[\frac{Vol(D)}{N}\sum_{i=1}^{N} f(\mathbf{x}_i)] = \frac{Vol(D)^2}{N^2}\sum_{i=1}^{N} \sigma^2[f(\mathbf{x}_i)] = \frac{Vol(D)^2\sigma^2[f]}{N}$

Importance Sampling and Random Walks

- Rather than sampling uniformly from the region of integration, we want to favor dominant contributions to an integral
- This situation arises frequently when the integral can be interpreted as a weighted average over some probability distribution (notable example - literally any quantum mechanical observable: $\langle \mathcal{O} \rangle = \int d\Phi e^{-S[\Phi]} \mathcal{O}[\Phi] / \int d\Phi e^{-S[\Phi]}$)
- We can use random walks to implement importance sampling via Markov chains (a sequence of random variables in which probability of moving to next state depends only on current state)

Stationary Distributions

- Goal: we want to design Markov chains where the random variables are distributed according to our distribution of interest (the weight in our integral)
- Moreover, we want to reproduce this distribution regardless of which state our random walk starts in
- A stationary Markov chain is one whose transition probabilities are stationary in time. This implies, in particular, that each random variable in the chain is identically distributed

Existence/Uniqueness of Stationary Distributions

- Our Markov chains are irreducible (any state can be in principle reached from any other state)
- In this case, unique stationary distribution exists if and only if Markov chain is positive recurrent
- Unique LIMITING stationary distribution exists if and only if Markov chain is positive recurrent AND aperiodic (i.e. ergodic – math people hate this word!)

Detailed Balance

- A reversible Markov chain is one that satisfies the detailed balance criterion: $\pi(\mathbf{x}')P(\mathbf{x}' \to \mathbf{x}) = \pi(\mathbf{x})P(\mathbf{x} \to \mathbf{x}')$
- Notably, if a Markov chain obeys the detailed balance criterion for some distribution π, then π is the stationary distribution of that MC
- Detailed balance makes it easy to determine the limiting distribution, hence used as a convenient criterion for algorithms!

The Metropolis Algorithm

- Suppose we have a target distribution $\pi(x)$ and a proposal distribution $P(x'|x)$. Define our sampling algorithm as follows:
  - Sample x' from $P(x'|x_n)$
  - Set $x_{n+1} = x'$ with acceptance probability $min(1, \frac{\pi(x')P(x_n|x')}{\pi(x_n)P(x'|x_n)})$
  - If proposed step is rejected, set $x_{n+1} = x_n$
- It is straightforward to show that Metropolis satisfies detailed balance

Why is Metropolis So Popular?

- Only need to know a function proportional to the target distribution, not the exact probability distribution itself
- Freedom in choice of proposal distribution (can choose one based on convenience for specific problem)
- Metropolis acceptance probability function maximizes number of accepted steps!

Main Points

- Monte Carlo integration is useful for high-dimensional integrals (use something else for low-dimensional integrals)
- Ergodicity is crucial to ensure that a Markov chain converges to the correct distribution
- Detailed balance is NOT fundamental, but is extremely useful in a practical sense
- Metropolis was a pretty smart guy