Overview of Slides 03

Monte Carlo integration – uniform sampling - recap

Monte Carlo integration with uniform sampling

$$I = \int_\Omega f(\vec{r})dV \quad \approx \quad \frac{V}{N}\sum_{i=1}^{N} f_i \quad \pm \quad \frac{V}{\sqrt{N}}\sqrt{\frac{1}{N}\sum_{i=1}^{N} f_i^2 - \left[\frac{1}{N}\sum_{i=1}^{N} f_i\right]^2}$$

with $f_i = f(\xi_i)$; ξ_i are uniformly sampled from Ω.

- Error scales as $1/\sqrt{N}$
- Error proportional to $\sqrt{\text{variance}}$ of function values
- Superior to fixed-grid and Gaussian quadrature in high dimensions
- Simple to deal with complicated boundaries

Monte Carlo integration – weights, importance sampling

Consider integral $$\boxed{I = \int_\Omega f(\vec{r})dV = \int_\Omega p(\vec{r})g(\vec{r})dV}$$

$p(\vec{r}) \rightarrow$ a prob. dist. from which we know how to draw random numbers.

Should be normalized: $\int_\Omega p(\vec{r})dV = 1.$

If $I_1 = \int_\Omega w(\vec{r})g(\vec{r})dV$ and $w(\vec{r})$ is not normalized, use appropriate factor:

$$I_1 = cI = c\int_\Omega p(\vec{r})g(\vec{r})dV \qquad \text{with} \qquad p(\vec{r}) = \frac{w(\vec{r})}{c} = \frac{w(\vec{r})}{\int_\Omega w(\vec{r})}$$

Monte Carlo integration – weights, importance sampling

If region is sampled according to distribution $p(\vec{r})$, expectation value of an observable $A(\vec{r})$ is

$$\langle A \rangle = \int_\Omega p(\vec{r}) A(\vec{r}) dV \qquad \begin{cases} \text{Normalized } p(x): \\ \text{no volume factor} \end{cases}$$

Thus

Monte Carlo integration with weighted sampling

$$I = \int_\Omega p(\vec{r}) g(\vec{r}) dV \quad \approx \quad \frac{1}{N} \sum_{i=1}^{N} g_i$$

with $g_i = g(\xi_i)$; ξ_i are sampled with probability $p(x)$.

Error given by $\sqrt{\text{variance}}$ of g, not of integrand $f = pg$.

Monte Carlo integration – weights, importance sampling

Monte Carlo integration with weighted sampling

$$I = \int_\Omega p(\vec{r}) g(\vec{r}) dV \quad \approx \quad \frac{1}{N} \sum_{i=1}^{N} g_i \quad \pm \quad \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{i=1}^{N} g_i^2 - \left[\frac{1}{N} \sum_{i=1}^{N} g_i \right]^2}$$

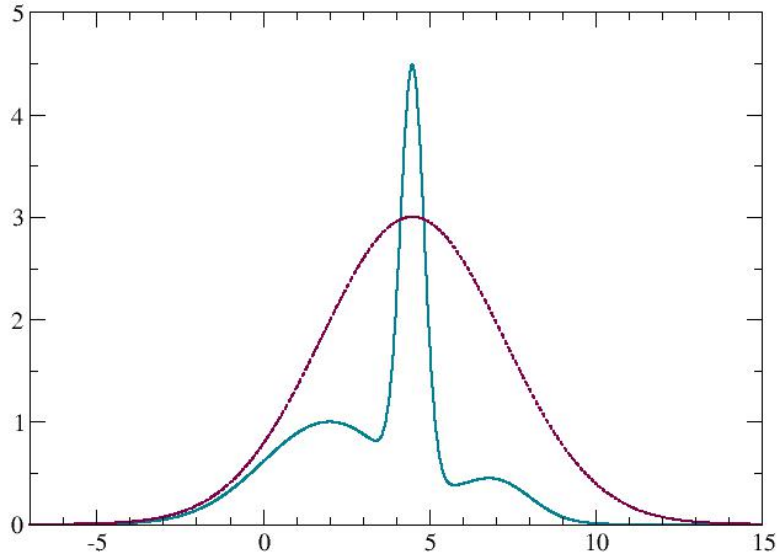
with $g_i = g(\xi_i)$; ξ_i are sampled with probability $p(x)$.

- Error still scales as $1/\sqrt{N}$
- When $g(\vec{r})$ has less variance than full integrand $f(\vec{r}) = p(\vec{r}) g(\vec{r}) \rightarrow$ better than uniform sampling — focuses on important regions of Ω.

Importance sampling

If integrand is sharply peaked in some region, and small in others:

- Need high accuracy where f is big and varying
- Dont waste time where it is close to zero



Sample more points where f is greater. How?

Choose x with probability $p(x)$ so that $f(x)/p(x) \approx$ constant.

$$I = \int f(x)dx = \int p(x)\left[\frac{f(x)}{p(x)}\right]dx \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}$$

x_i's sampled from probability distribution $p(x)$.

Importance sampling — Example 1

$$I = \int_0^1 e^{-x} g(x) dx$$

Very tempting to use `numpy.random.exponential()`.

Turn e^{-x} into a probability distribution:

$$I = c \int_0^1 \left[\frac{e^{-x}}{c} \right] g(x) dx = c \int_0^1 p(x) g(x) dx \quad \text{with } c = \int_0^1 e^{-x} dx = \frac{e-1}{e}$$

How to sample from $p(x) \propto e^{-x}$ with support $(0, 1)$?

$\rightarrow$ sample from exponential on $(0, \infty)$; reject values > 1.

$$I \approx c \times \frac{1}{N} \sum_{i=1}^{N} g(\xi_i) \qquad \text{with } \xi_i\text{'s sampled from } p(x)$$

Importance sampling — Example 2: variational approx.

Variational principle:

apprxomiation (upper bound) to ground state energy of $\hat{H}$ given by

$$E_v = \frac{\langle \psi_v | \hat{H} | \psi_v \rangle}{\langle \psi_v | \psi_v \rangle} \qquad \begin{cases} \psi_v \to \text{variational wavefunction (trial wavefn)} \\ E_v \text{ to be minimized over class of trial wavefn's} \end{cases}$$

Before minimizing, have to calculate E_v first

$\to$ can be a serious computational problem. E.g.,

$$E_v = \frac{1}{\langle \psi_v | \psi_v \rangle} \int d\vec{x}\, \psi_v^*(\vec{x})\, \hat{H}\, \psi_v(\vec{x}) \qquad \begin{cases} \vec{x} \text{ could be, e.g.,} \\ 3N\text{-dimensional for } N \text{ particles} \end{cases}$$

Rewrite as

$$E_v = \int d\vec{x}\; \frac{|\psi_v(\vec{x})|^2}{\langle \psi_v | \psi_v \rangle}\; \frac{\hat{H}\, \psi_v(\vec{x})}{\psi_v(\vec{x})}$$

First factor is a probability dist. Sample from this using Markov chain MC.

Importance sampling — Example 2: variational approx.

Variational Quantum Monte Carlo

$$E_v = \int d\vec{x} \; \frac{\left|\psi_v(\vec{x})\right|^2}{\langle \psi_v | \psi_v \rangle} \; \frac{\hat{H}\,\psi_v(\vec{x})}{\psi_v(\vec{x})}$$

First factor is a probability dist. Sample from this using Markov chain MC.

One of several variants of Quantum Monte Carlo (QMC) techniques used in quantum many-body physics

Monte Carlo for classical statistical physics

Monte Carlo widely used to sample statistical distributions

e.g., thermodynamic properties of polymers, magnetic systems, proteins, ...

Sample the Boltzmann distribution $\frac{1}{Z}e^{-\beta E(X)}$, where
X is system configuration, a many-dimensional variable

Classical statistical physics

Microscopically

Large number of individual degrees of freedom $\{p_i, q_i\}, H(p, q)$
Classical evolution is deterministic and reversible

Macroscopically

A few thermodynamical control variables: T, μ, V
and bulk thermodynamical properties: $\varepsilon, P, S, \ldots$
Evolution is statistical and irreversible

Statistical physics: bridges micro $\leftrightarrow$ macro

Work with ensembles of microscopic configurations
Probability of configuration $X = \{p_i, q_i\}$ is

$$P(X) \propto e^{-\beta H(X)} \equiv e^{-H(X)/k_B T}$$

$\longrightarrow$ 'definition' of temperature

Why Monte Carlo?

Only a few systems can be 'solved' analytically
— simulation is an essential tool

Task

Generate configurations with probability weight $P(X) \propto e^{-\beta H(X)}$

Each value of X represents a state or configuration of the system.

Ideal gas: $H = \sum_i \frac{p_i^2}{2m} \implies P(X) =$ product of gaussians — easy
With interactions: not so easy

Common recipe

Construct a series of configurations via a stochastic process

Stochastic processes

This is itself a topic in statistical physics / mathematics / statistics

Unlike deterministic processes, not modelled by usual pde's

Main topic: probability of being in certain states at time t:

- What is $P(x)_t \equiv W_1(x,t)$ — prob of state x at time t?
- What is $P(x_{t_1}, y_{t_2}) \equiv W_2(x,t_1,y,t_2)$ — prob of x at t_1 and y at t_2?
- What is $P(x_{t_1}, y_{t_2}, z_{t_3}) \equiv W_3(x,t_1,y,t_2,z,t_3)$ etc?
- What is $P(y_{t_2}|x_{t_1})$ — prob of getting from x at t_1 to y at t_2?

Stochastic processes

Examples

- Brownian motion
- Diffusion and drift
- Population dynamics
- Stock markets
- Many-particle systems
  e.g., Ising model

Types

- Poisson process
- Markov process
- Bernoulli process
- Wiener process
- Lévy process
- Martingale
- Discrete-time random walks
- .....

Properties of stochastic processes

x, t can be discrete or continuous

$$\int W_1(x,t)dx = 1\,; \qquad W_n \geq 0\,\forall n$$

$$\int W_2(x,t_1,y,t_2)dx = W_1(y,t_2)\,; \quad \int W_2(x,t_1,y,t_2)dy = W_1(x,t_2)$$

x can be either a single number (discrete or continuous),
or a collection of numbers representing a configuration.

Properties of stochastic processes

x, t can be discrete or continuous

$$\int W_1(x,t)dx = 1\,; \qquad W_n \geq 0 \,\forall n$$

$$\int W_2(x,t_1,y,t_2)dx = W_1(y,t_2)\,; \quad \int W_2(x,t_1,y,t_2)dy = W_1(x,t_2)$$

Stationary processes

$$W_1(x,t) = W_1(x) = P(x) \quad \text{independent of } t$$
$$W_2(x_1,t_1,x_t,t_2) = W_2(x_1,x_2;t_2-t_1)$$

Ergodic processes

If ensemble mean equals time average then the process is ergodic

The process will 'visit' all possible states of the system

- every state must be accessible from every other state
- no periodicity

Autocorrelations

Measure the 'memory' of the process:

$$R_\phi(t_1, t_2) \equiv \langle \Delta\phi(t_1)\Delta\phi(t_2)\rangle = \Big\langle \big(\phi(t_1) - \langle\phi(t_1)\rangle\big)\big(\phi(t_2) - \langle\phi(t_2)\rangle\big)\Big\rangle$$
$$= \int \big(\phi(x_1, t_1) - \langle\phi(t_1)\rangle\big)\big(\phi(x_2, t_2) - \langle\phi(t_2)\rangle\big) W_2(x_1, t_1, x_2, t_2)\, dx_1\, dx_2$$

Here ϕ is a stochastic variable.

For stationary process: $\langle\phi(t_1)\rangle = \langle\phi(t_2)\rangle = \text{constant}$

$$R_\phi(t_2 - t_1) = R_\phi(\tau) = \big\langle \Delta\phi(0)\Delta\phi(\tau)\big\rangle$$

Typically $R_\phi(\tau) \sim e^{-\tau/\tau_\phi}$, where τ_ϕ is the autocorrelation time.

Markov processes

Simple(st) stochastic process: uncorrelated random process

$$W_2(x_1, t_1, x_2, t_2) = W_1(x_1, t_1)W_2(x_2, t_2) \quad \Longrightarrow \quad R(\tau) = c\delta(\tau)$$

All information contained in W_1's.

Next simplest: all information is in W_2 — Markov process

A Markov process is a random process in which the future is independent of the past, given the present.

Markov processes

Define transition probabilities
(for either discrete time and continuous time)

$$P(x_2 t_2 | x_1 t_1) \equiv T(x_1, t_1 \to x_2, t_2) \qquad \int P(x_2 t_2 | x_1 t_1) dx_2 = 1$$

$$W_2(x_1, t_1, x_2, t_2) = W_1(x_1, t_1) P(x_2 t_2 | x_1 t_1)$$

All information about dynamics of the process is in T.

Where we go next is independent of how we got there.

We focus on discrete time:

$$P(x_n t_n | x_{n-1} t_{n-1} \ldots x_1 t_1) = P(x_n t_n | x_{n-1} t_{n-1})$$

Only the most recent time counts!

Markov processes

Examples of Markov processes

- random walk
- population dynamics

Examples of non-Markov processes

- self-avoiding random walk
- stock markets

Markov chains

With discrete time steps: Markov chain
— need only look at transition probabilities from one time step to next

$$T(Xt_n \to Yt_{n+1}) \equiv T(X \to Y)$$

The Master Equation

Look at probability of being in state X at time t, $P(X,t)$

$$P(X,t_{n+1}) = \sum_Y P(Y,t_n)\,T(Y \to X)$$

$$P(X,t_n) = \sum_Y P(X,t_n)\,T(X \to Y)$$

Master equation, discrete time

$$P(X,t_{n+1}) - P(X,t_n) = \sum_Y \Big[P(Y,t_n)\,T(Y \to X) - P(X,t_n)\,T(X \to Y)\Big]$$

Continuous time:

$$\frac{\partial P(x,t)}{\partial t} = \int \Big[P(y,t)t(y \to x) - P(x,t)t(x \to y)\Big]\,dy$$

Detailed balance

For a stationary process $P(X, t_{n+1}) = P(X, t_n) = P(X)$

$\Longrightarrow$ rhs of Master Equation is 0.

Sufficient condition

$$P(X)T(X \to Y) = P(Y)T(Y \to X) \qquad \text{Detailed balance}$$

If detailed balance is satisfied: Markov chain will follow irreversible process towards stationary distribution (towards equilibrium)

Markov chain Monte Carlo

Use such a process to create configurations X
with desired distribution $P(X)$

Design update rule/algorithm (transition probability T)
satisfying detailed balance for desired distribution $P(X)$

Summary

Monte Carlo integration

- Via uniform or weighted sampling.
- Importance sampling can drastically improve performance

Markov processes

- A Markov process is a stochastic process with no memory
- Described by transition probabilities
- All Markov processes obey the master equation
- Detailed balance is a sufficient condition for a stationary process