

PHY604 Lecture 21

November 16, 2023

Review: Transformation method for changing distributions

- We have a source of random numbers z drawn from distribution $q(z)$
 - Probability of generating a number between z and $z+dz$ is $q(z)dz$
- Now we choose a function $x = x(z)$ whose distribution $p(x)$ is the one we want

• We know that:
$$p(x)dx = q(z)dz$$

- If our random numbers are drawn from a uniform distribution $[0,1)$, $q(z)=1$ from 0 to 1, zero elsewhere

• Then:

$$\int_{-\infty}^{x(z)} p(x')dx' = \int_0^z dz' = z$$

- We need to do the integral on the left and then solve for $x(z)$
 - Not always possible

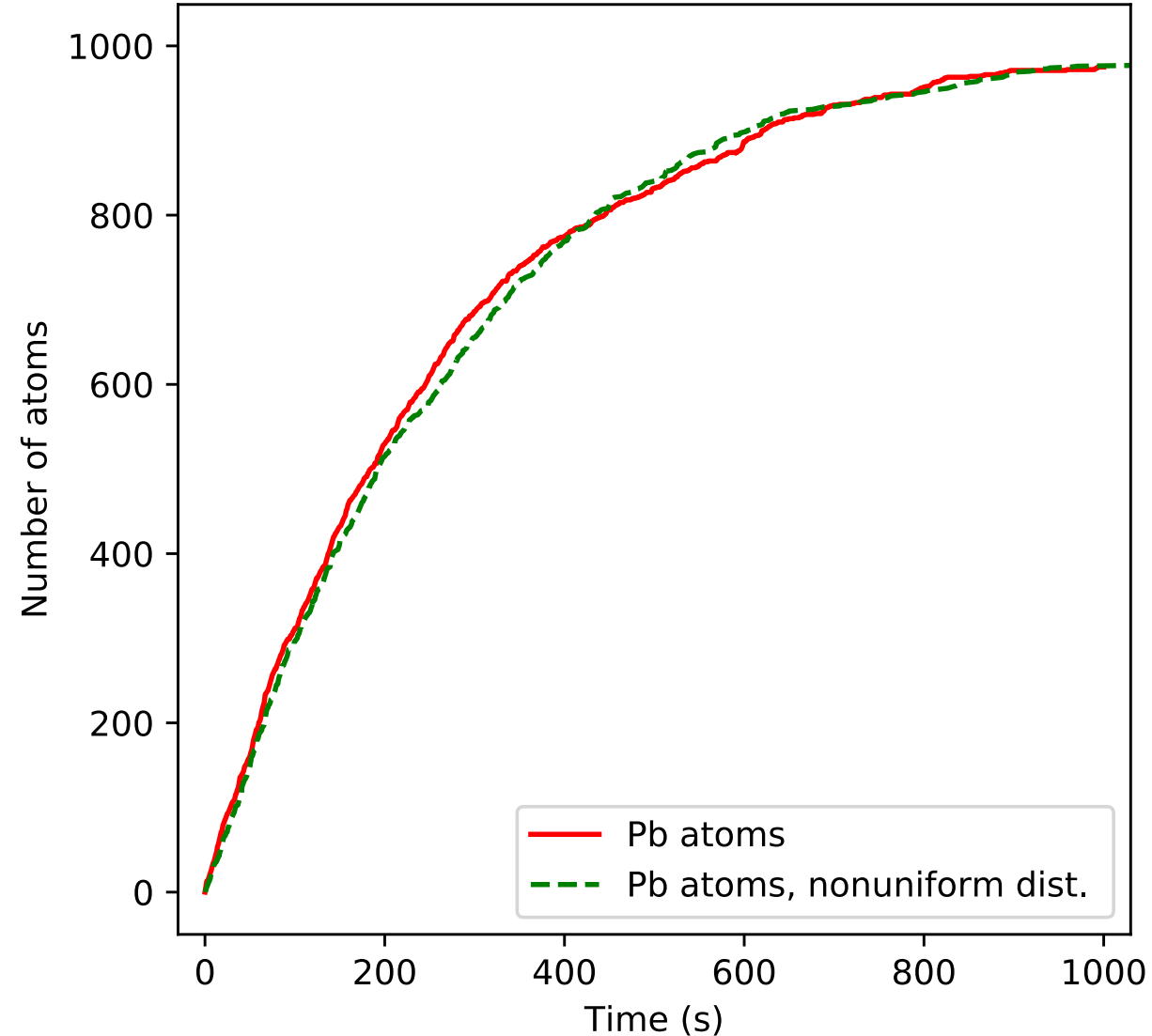
Review: Nonuniform distribution for decay example

- We can write the probability distribution for the decay example as

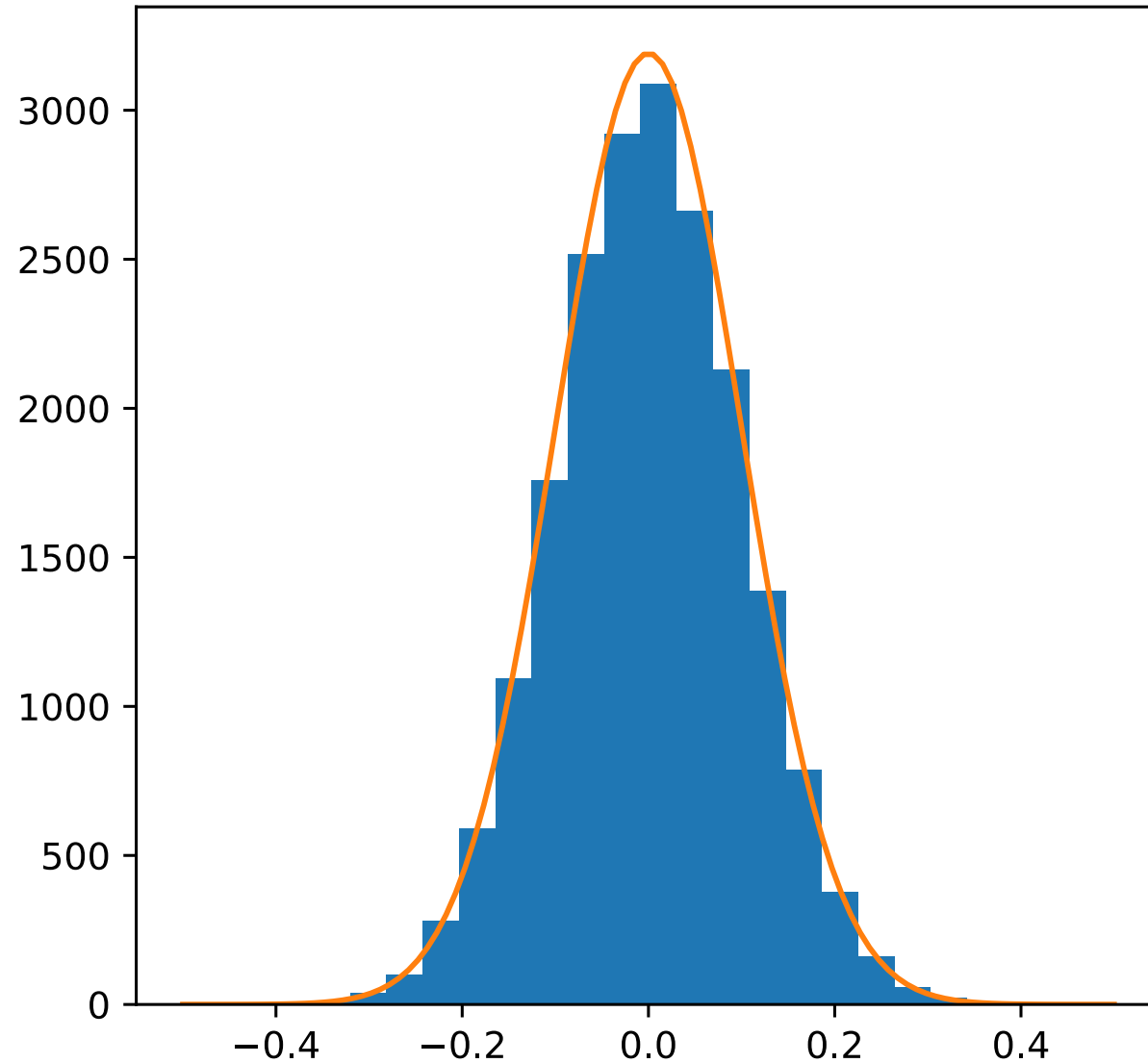
$$P(t)dt = 2^{-t/\tau} \frac{\ln 2}{\tau} dt = e^{-t \ln 2/\tau} \frac{\ln 2}{\tau}$$

- So:

$$x = -\frac{\tau}{\ln 2} \ln(1 - z)$$



Example: Random numbers from Gaussian distribution



Today's lecture: Monte Carlo integration and simulation

- Monte Carlo integration
- Monte Carlo simulation

Monte Carlo integration

- Let's come back to the Rutherford scattering example
- One way to look at: Our stochastic solution was in good agreement with the exact one
- Another way to look at it: Using a random process, we obtained an approximate solution to the integral:

$$\frac{1}{\sigma^2} \int_0^b \exp\left(-\frac{r'^2}{2\sigma^2}\right) r' dr'$$

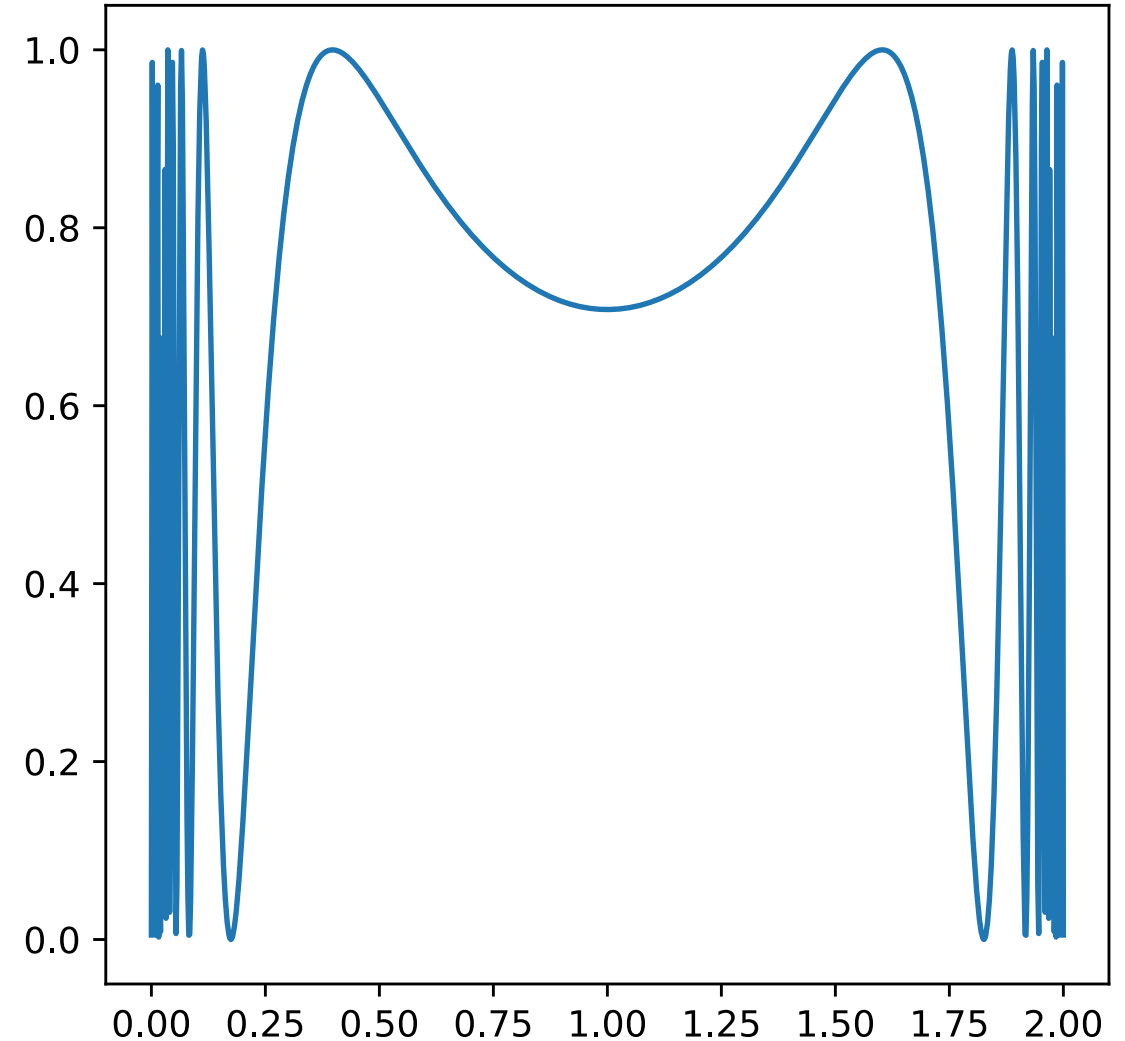
- **Monte Carlo integration:** Approximate the value of an integral (which has an exact solution) with random calculations

Example: Challenging integral with exact solution in principle

- Consider the function:

$$I = \int_0^2 \sin^2 \left[\frac{1}{x(2-x)} \right] dx$$

- Finite over the range, must be less than $2 \times 1 = 2$
- Oscillates infinitely fast at the edges so very challenging for numerical integration

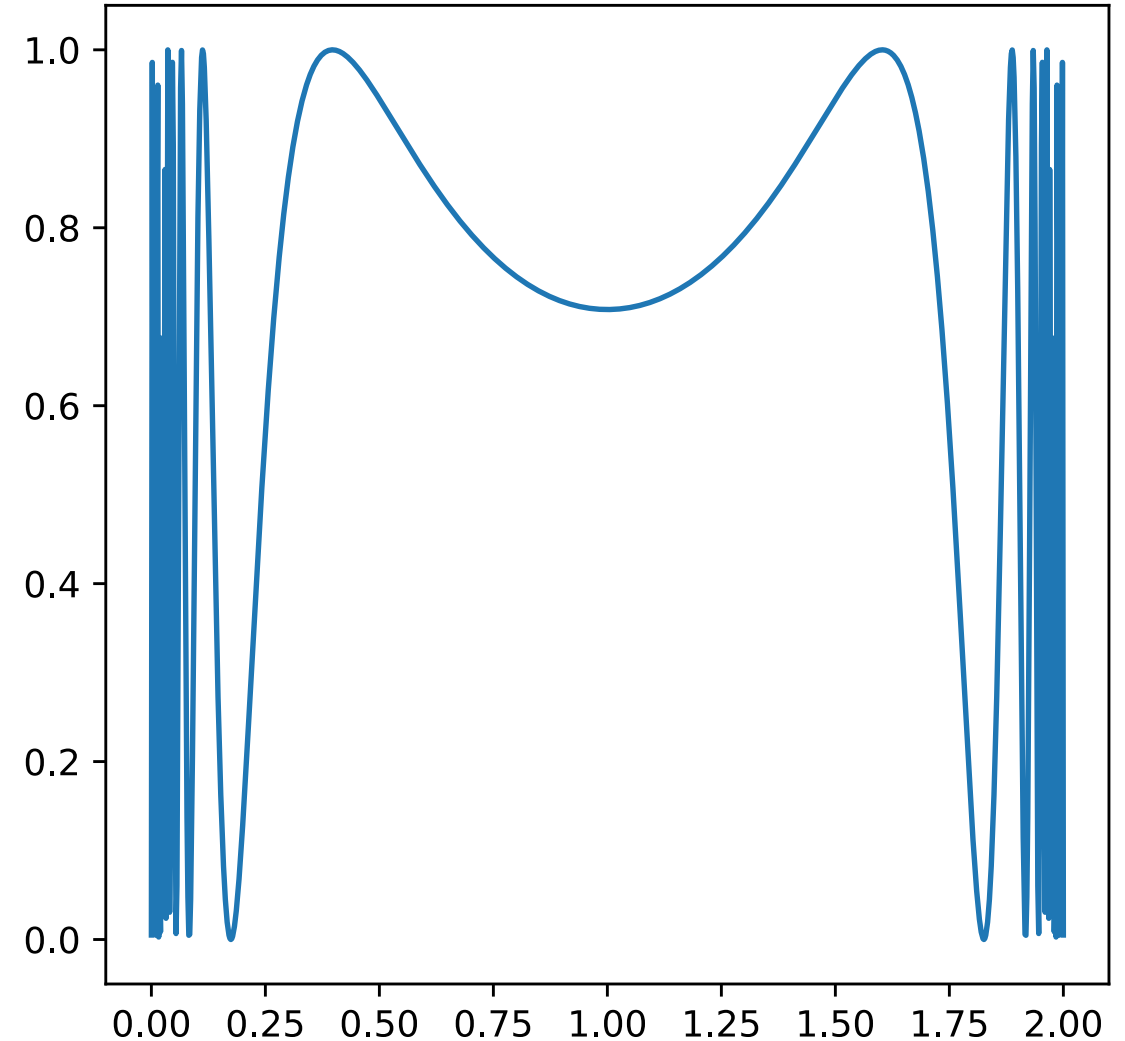


Monte Carlo Integration with random sampling

$$I = \int_0^2 \sin^2 \left[\frac{1}{x(2-x)} \right] dx$$

- Choose N random samples in the bounding rectangle with area $A=2$
- Check which lie under the curve
- Probability that point lies under the curve is $p = I/A$
- Fraction of points under the curve k/N should be approximately p
- So:

$$I \simeq \frac{kA}{N}$$



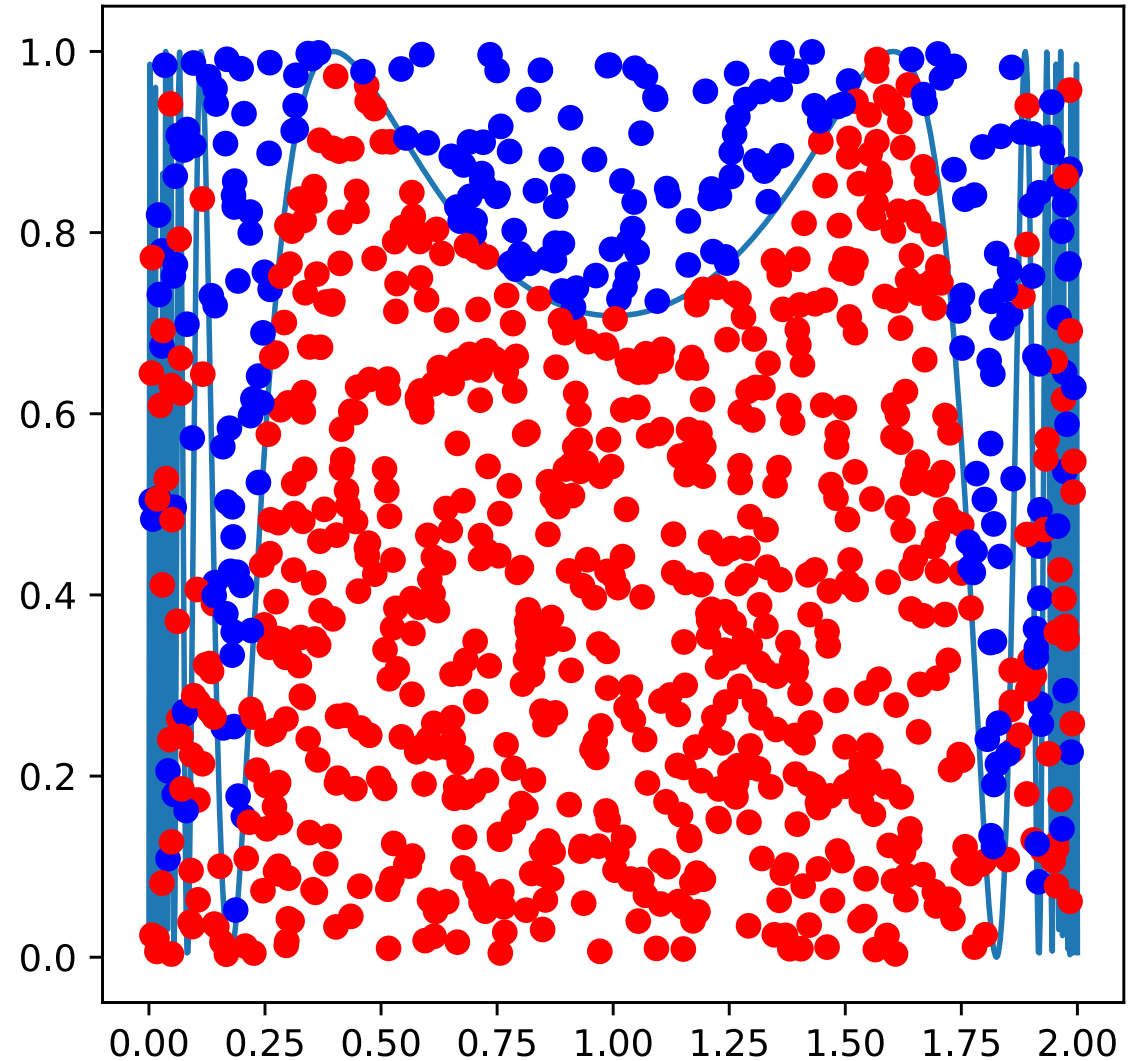
Errors in Monte Carlo method

- Normally gives worse results than, e.g., Simpson's or trapezoid rule for simple integrals
- Probability random point falls below the curve is p , above is $1-p$
- Probability that k points fall below the curve is

$$p^k (1 - p)^{N-k}$$

- There are $\binom{N}{k}$ ways to choose k points, so the probability to get k points under the curve is

$$P(k) = \binom{N}{k} p^k (1 - p)^{N-k}$$



Errors in Monte Carlo method

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k}$$

- This is a binomial distribution, which has variance:

$$\text{var}k \equiv \langle k^2 \rangle - \langle k \rangle^2 = Np(1-p) = N \frac{I}{A} \left(1 - \frac{I}{A} \right)$$

- And standard deviation is: $\sqrt{\text{var}k}$

- So, the error on the integral I is:

$$I_{\text{error}} = \sqrt{\text{var}k} \frac{A}{N} = \frac{\sqrt{I(A-I)}}{\sqrt{N}} \propto \frac{1}{\sqrt{N}}$$

Compare MC errors to quadrature rules

$$I_{\text{error}} = \sqrt{\text{var}k} \frac{A}{N} = \frac{\sqrt{I(A-I)}}{\sqrt{N}} \propto \frac{1}{\sqrt{N}}$$

- Errors for MC integration decrease like $N^{-1/2}$
- For the trapezoid rule, error was on the order of Δx^2 , where Δx is the width of the integration slice:

$$\Delta x = \frac{b-a}{N}$$

- So, error decreases like N^{-2} much better than MC!
- For Simpson's rule, it decreases like N^{-4}
- **Monte Carlo methods should be used only when other methods break down!**

Can we do better? Mean value method

- Consider general integration problem: $I = \int_a^b f(x)dx$

- Average value of f in the range between b and a is:

$$\langle f \rangle \equiv \frac{1}{b-a} \int_a^b f(x)dx = \frac{I}{b-a}$$

- So, we can get the integral by finding the average of f :

$$I = (b-a)\langle f \rangle$$

- We can estimate the average by measuring $f(x)$ at N points chosen at random between a and b

- Then:

$$I \simeq \frac{(b-a)}{N} \sum_{i=1}^N f(x_i)$$

Errors of the mean value method

- Can estimate the error using the general theorem: **The variance on the sum of N independent random numbers is the sum of the variances of the individual numbers**
 - Holds no matter what the distribution is

- So:

$$\text{var } f \equiv \langle f^2 \rangle - \langle f \rangle^2$$

- Where:

$$\langle f \rangle = \frac{1}{N} \sum_{i=1}^N f(x_i), \quad \langle f^2 \rangle = \frac{1}{N} \sum_{i=1}^N [f(x_i)]^2$$

- And:

$$I_{\text{error}} = \frac{b-a}{N} \sqrt{N \text{var } f} = (b-a) \frac{\sqrt{\text{var } f}}{\sqrt{N}}$$

Still $N^{-1/2}$, but prefactor turns out to be smaller



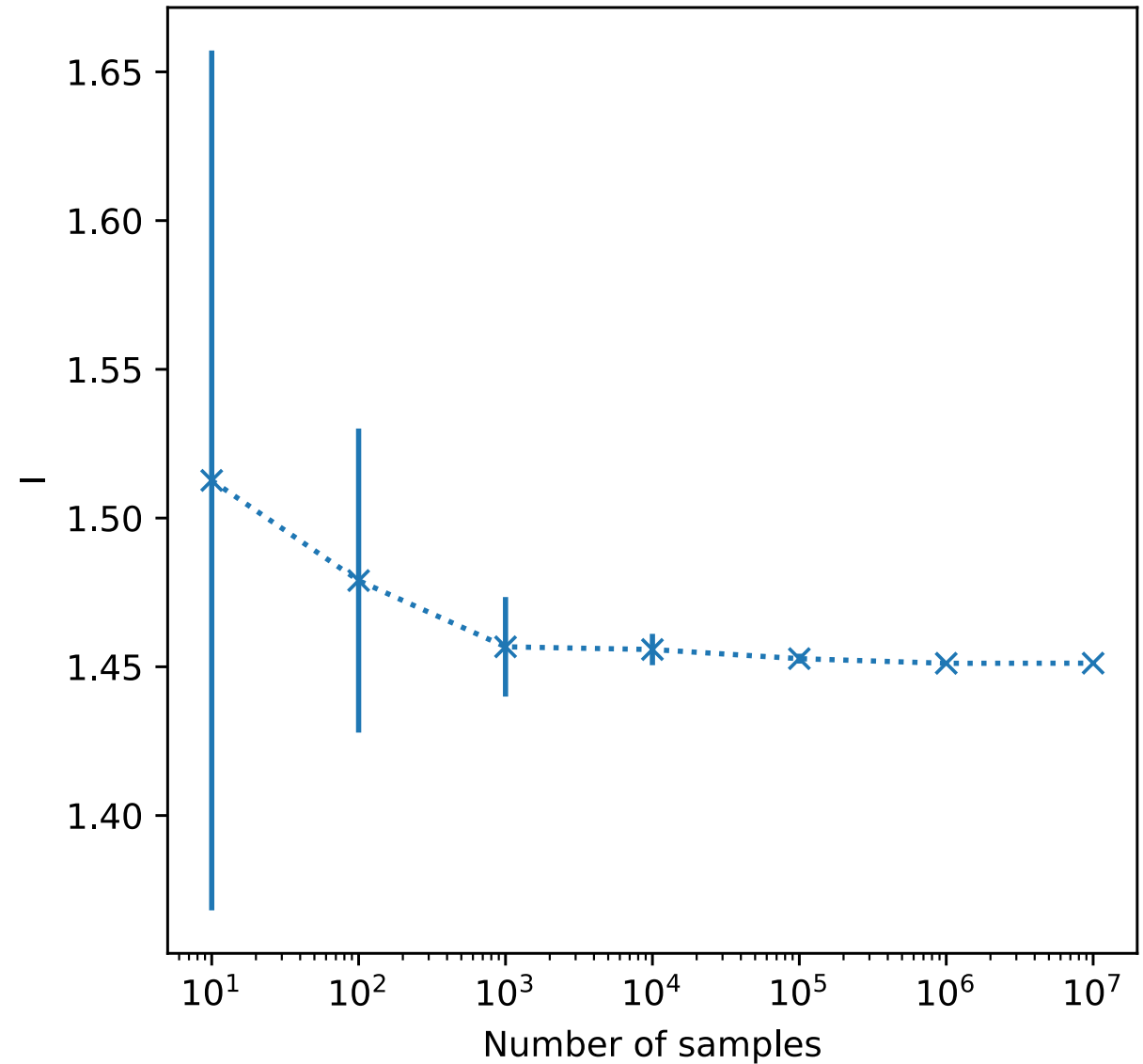
Mean value method

- Equation:

$$I = \int_0^2 \sin^2 \left[\frac{1}{x(2-x)} \right] dx$$

- Errors:

$$I_{\text{error}} = (b - a) \frac{\sqrt{\text{var } f}}{\sqrt{N}}$$



When to use Monte Carlo integration?

Multi-dimensional integrals

- If we have an integral over many dimensions (> 4), grid sizes get very large, scale as N^d
- Monte Carlo integration can give reasonable results with many fewer points
- Straightforward to generalize methods discussed to more dimensions
 - E.g., mean value method

$$I \simeq \frac{V}{N} \sum_{i=1}^N f(\mathbf{r}_i)$$

Example: Volume of hypersphere

- Consider a hypersphere of unit radius in all dimensions:

$$f(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- Let's use the mean value method to compute the integral of a 10-dimensional hypersphere
 - Trapezoid rule with 100 samples per dimension: 10^{20} grid points!

- We can compare to the exact solution:

$$V_d(r) = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2} + 1\right)} r^d$$

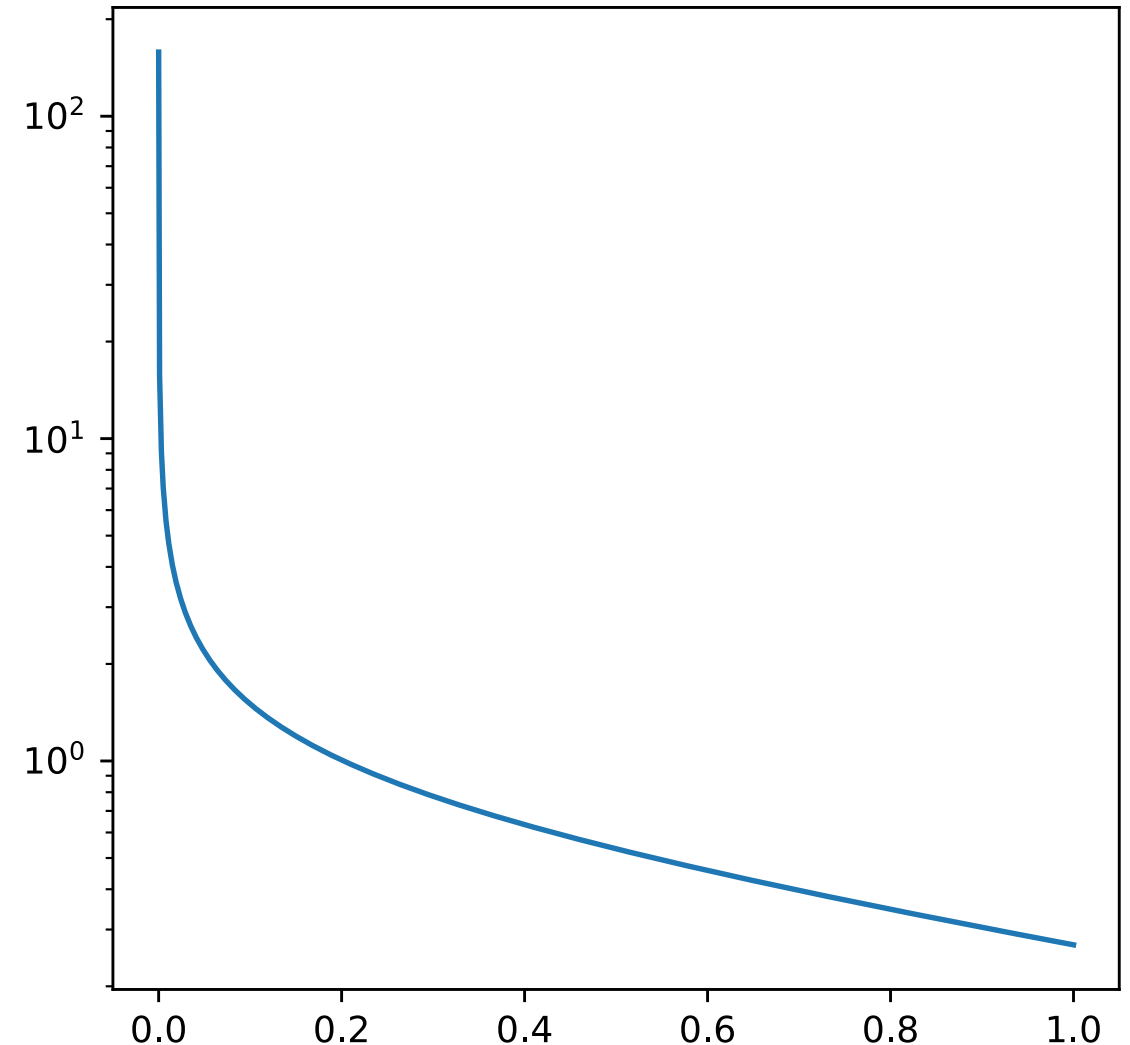
Monte Carlo integration with divergences

- Monte Carlo integration fails for some pathological functions, e.g., those that contain divergences

- Consider:

$$I = \int_0^1 \frac{x^{-1/2}}{e^x + 1} dx$$

- Function diverges at $x=0$, but integral is finite
- E.g., for mean value method, will occasionally get a very large contribution
 - Estimate varies widely between runs



Importance sampling

- Can get around these issues by drawing points nonuniformly
- For a general function $g(x)$ can define a **weighted average**:

$$\langle g \rangle_w = \frac{\int_a^b w(x)g(x)dx}{\int_a^b w(x)dx}$$

- $w(x)$ is a weighting function

- If we want to solve a general 1D integral: $I = \int_a^b f(x)dx$

- We set $g(x)=f(x)/w(x)$:

$$\left\langle \frac{f(x)}{w(x)} \right\rangle_w = \frac{\int_a^b f(x)dx}{\int_a^b w(x)dx} = \frac{I}{\int_a^b w(x)dx}$$

Importance sampling, 1D integral

- Thus, we have:

$$I = \left\langle \frac{f(x)}{w(x)} \right\rangle_w \int_a^b w(x) dx$$

- Equivalent to the mean value method, but from a weighted average
- How do we calculate the weighted average?
- Define probability density function as normalized $w(x)$

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

- So

$$\langle g \rangle_w = \int_a^b p(x) g(x) dx$$

Importance sampling, 1D integral

- Now let's sample N random points in the interval with the distribution $p(x)$. Then:

$$\sum_{i=1}^N g(x_i) \simeq \int_a^b N p(x) g(x) dx$$

- So:

$$\langle g \rangle_w = \int_a^b p(x) g(x) dx \simeq \frac{1}{N} \sum_{i=1}^N g(x_i)$$

- Where x_i are chosen from the distribution:

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

Importance sampling, 1D integral

- Putting everything together:

$$I \simeq \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{w(x_i)} \int_a^b w(x) dx$$

- Generalization of mean value method, which is where $w(x)=1$
- $w(x)$ can be any function that we choose
 - Can be chosen to remove pathologies in the integrand
- However, now we need to draw from a nonuniform distribution

Error on importance sampling method

- Error is given by:

$$I_{\text{error}} = \frac{\sqrt{\text{var}_w(f/w)}}{\sqrt{N}} \int_a^b w(x) dx$$

- Where:

$$\text{var}_w g = \langle g^2 \rangle_w - \langle g \rangle_w^2$$

- Still goes like $N^{-1/2}$

Importance sampling for pathological function

- Let's return to the integral: $I = \int_0^1 \frac{x^{-1/2}}{e^x + 1} dx$

- Choose: $w(x) = x^{-1/2}$

- Then: $f(x)/w(x) = (e^x + 1)^{-1}$

- Finite and well-behaved over the range

- Probability distribution is:

$$p(x) = \frac{x^{-1/2}}{\int_0^1 x^{-1/2} dx} = \frac{1}{2\sqrt{x}}$$

- So, using the transformation method:

$$\int_0^x \frac{1}{2\sqrt{x'}} dx' = \sqrt{x} = z \quad \implies \quad x = z^2$$

Importance sampling for pathological function

- So finally, we need to sample:

$$I \simeq \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{w(x_i)} \int_a^b w(x) dx = \frac{1}{N} \sum_{i=1}^N \frac{1}{e^{x_i} + 1} \int_0^1 \frac{1}{\sqrt{x}} dx = \frac{1}{N} \sum_{i=1}^N \frac{2}{e^{x_i} + 1}$$

- With the distribution $x = z^2$

Today's lecture: Monte Carlo integration and simulation

- Monte Carlo integration
- Monte Carlo simulation

Monte Carlo simulation

- Any computer simulation that uses random numbers to simulate physical process
- We saw a few examples already: radioactive decay and Rutherford scattering
- Used in every branch of physics
 - Particularly important in **statistical mechanics** and many-body physics

Monte Carlo simulation in stat mech

- Fundamental problem in statistical mechanics: Calculate expectation value of quantity of interest in thermal equilibrium
- Don't know the exact state of the system, only probability of occupying state i with energy E_i

$$P(E_i) = \frac{e^{-\beta E_i}}{Z}, \quad Z = \sum_i e^{-\beta E_i}$$

- Then average value of observable X :

$$\langle X \rangle = \sum_i X_i P(E_i)$$

States with large numbers

$$\langle X \rangle = \sum_i X_i P(E_i)$$


- Calculating this sum exactly can only be done in a few specific systems (e.g., harmonic oscillator)
- Numerically challenging: states are order Avogadro's number in size
- E.g., one mole of gas with two states: total number of states is $2^{10^{23}}$
- Instead, use Monte Carlo approach to evaluate the sum

Monte Carlo approach to expectation values

- We could choose N terms in the sum at random to add up:

$$\langle X \rangle \simeq \frac{\sum_{k=1}^N X_k P(E_k)}{\sum_{k=1}^N P(E_k)}$$

Needed to normalize
the weighted average if
not summing over *all*
states



- This would not work well! Boltzmann probability is exponentially small for states $E_i \gg k_B T$
- Usually, most of the states are high energy, only a few contribute significantly
- **Need to use importance sampling!**

Importance sampling for thermal average

- Choose nonuniform distribution to focus on this small set
- Define weighted average over states:

$$\langle g \rangle_w \simeq \frac{\sum_i w_i g_i}{\sum_i w_i}$$

- We choose: $g_i = X_i P(E_i) / w_i$

- So:

$$\left\langle \frac{X_i P(E_i)}{w_i} \right\rangle_w = \frac{\sum_i w_i X_i P(E_i) / w_i}{\sum_i w_i} = \frac{\sum_i X_i P(E_i)}{\sum_i w_i} = \frac{\langle X \rangle}{\sum_i w_i}$$

- Or:

$$\langle X \rangle = \left\langle \frac{X_i P(E_i)}{w_i} \right\rangle_w \sum_i w_i$$

Importance sampling for thermal average

$$\langle X \rangle = \left\langle \frac{X_i P(E_i)}{w_i} \right\rangle_w \sum_i w_i$$

- Evaluate by selecting N states randomly with nonuniform distribution:

$$\langle X \rangle \simeq \frac{1}{N} \sum_{k=1}^N \frac{X_k P(E_k)}{w_k} \sum_i w_i$$

↑
Summed over N samples

← Summed over all states

- Still need to choose w_i to bias us towards high-probability samples
 - Also, so that sum over all states i can be evaluated analytically

Weights for importance sampling

- Simple choice: $w_i = P(E_i)$
- Sums to 1 over all by definition
- Then we have:

$$\langle X \rangle \simeq \frac{1}{N} \sum_{k=1}^N X_k$$

- Thus, choose N states in proportion to their Boltzmann weights, and average X over them

Markov chain Monte Carlo

- Recall that:
$$P(E_i) = \frac{e^{-\beta E_i}}{Z}, \quad Z = \sum_i e^{-\beta E_i}$$
- Partition function requires a sum over all states that we are trying to avoid
- Can use a **Markov chain** to choose states with probability $P(E_i)$ without knowing the partition function:
 - Start with a state i
 - Generate a new state j by making a small change to i
 - Choice of new state is determined probabilistically by a set of **transition probabilities** T_{ij} that give probability for changing from state i to j
- If we chose T_{ij} correctly, probability of visiting any state on a step of the Markov chain is $P(E_i)$!

Transition probabilities in the MC

- We must end up in some state on every MC step, so:

$$\sum_j T_{ij} = 1$$

- Choose transition probabilities such that:

$$\frac{T_{ij}}{T_{ji}} = \frac{P(E_j)}{P(E_i)} = \frac{e^{-\beta E_j} / Z}{e^{-\beta E_i} / Z} = e^{-\beta(E_j - E_i)}$$

- I.e., choosing particular ratio of the probability to go from i to j , and j to i
- Partition function cancels out!

Transition probabilities in the MC

- If we have correct probability of being in a given state at one step, we will have the correct probability for all later steps

- To see this:

- Suppose we find a set of T_{ij} 's that satisfy the previous conditions
- Suppose the probability to be in state i on one particular step is $P(E_i)$
- Then, probability to be in state j on the next step is:

$$\sum_i T_{ij} P(E_i) = \sum_i T_{ji} P(E_j) = P(E_j) \sum_i T_{ji} = P(E_j)$$

- Once we get a Boltzmann distribution over states, we will keep it
 - Boltzmann distribution is a fixed point of the Markov chain
- **Can also prove that we will converge to Boltzmann distribution**
 - See, e.g., Appendix D of Newman

Metropolis-Hastings accept/reject

- Still have not worked out what elements of T_{ij} are
 - Actually, many possible choices
- Most common choice: Metropolis-Hastings algorithm:
 - Choose the change between i and j from specified set of possible changes
 - Can be, e.g., chosen at random, uniformly
 - Accept or reject the new state with acceptance probability:

$$P_a = \begin{cases} 1 & \text{if } E_j \leq E_i \\ e^{-\beta(E_j - E_i)} & \text{if } E_j > E_i \end{cases}$$

- I.e., definitely accept if energy is lowered (or equal); may still accept if energy is increased

Transition probabilities under Metropolis-Hastings

- Total probability to move from i to given j (if $E_j < E_i$)

$$T_{ij} = \frac{1}{M} e^{-\beta(E_j - E_i)}$$



Probability we choose j



Probability we accept

Transition probabilities under Metropolis-Hastings

- If $E_j > E_i$:

$$T_{ij} = \frac{1}{M} e^{-\beta(E_j - E_i)}, \quad T_{ji} = \frac{1}{M} \quad \Longrightarrow \quad \frac{T_{ij}}{T_{ji}} = e^{-\beta(E_j - E_i)}$$

- If $E_j \leq E_i$:

$$T_{ij} = \frac{1}{M}, \quad T_{ji} = \frac{1}{M} e^{-\beta(E_i - E_j)} \quad \Longrightarrow \quad \frac{T_{ij}}{T_{ji}} = e^{-\beta(E_j - E_i)}$$

- Thus, both consistent with :

$$\frac{T_{ij}}{T_{ji}} = \frac{P(E_j)}{P(E_i)} = \frac{e^{-\beta E_j} / Z}{e^{-\beta E_i} / Z} = e^{-\beta(E_j - E_i)}$$

Some comments about the Metropolis algorithm

- Note that many steps will not change the system
 - Still need to include in the sum
- The number of possible moves M , must be the same when going from i to j as j to i
- Moves must be chosen to get you to every state
 - Move set for which all states are accessible is called **ergodic**
- Will generally take some (unknown) time to equilibrate to Boltzmann distribution

Steps of Markov chain Monte Carlo:

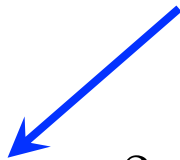
- 1. Choose random starting state
- 2. Choose a move uniformly at random from set of moves
- 3. Calculate the acceptance probability
- 4. Accept or reject the move
- 5. Measure X in current state, add to sum
- 6. Go back to step 2

Example: Ideal gas

- Consider the quantum states of a particle or atom of mass m in cubic box of length L
- Energy of one particle given by:

$$E(n_x, n_y, n_z) = \frac{\pi^2 \hbar^2}{2mL^2} (n_x^2 + n_y^2 + n_z^2)$$

Quantum numbers from 1 to infinity.



- Ideal gas: no interactions between particles
 - Energy is sum of individual particles:

$$E = \sum_{i=1}^N E(n_x^{(i)}, n_y^{(i)}, n_z^{(i)})$$

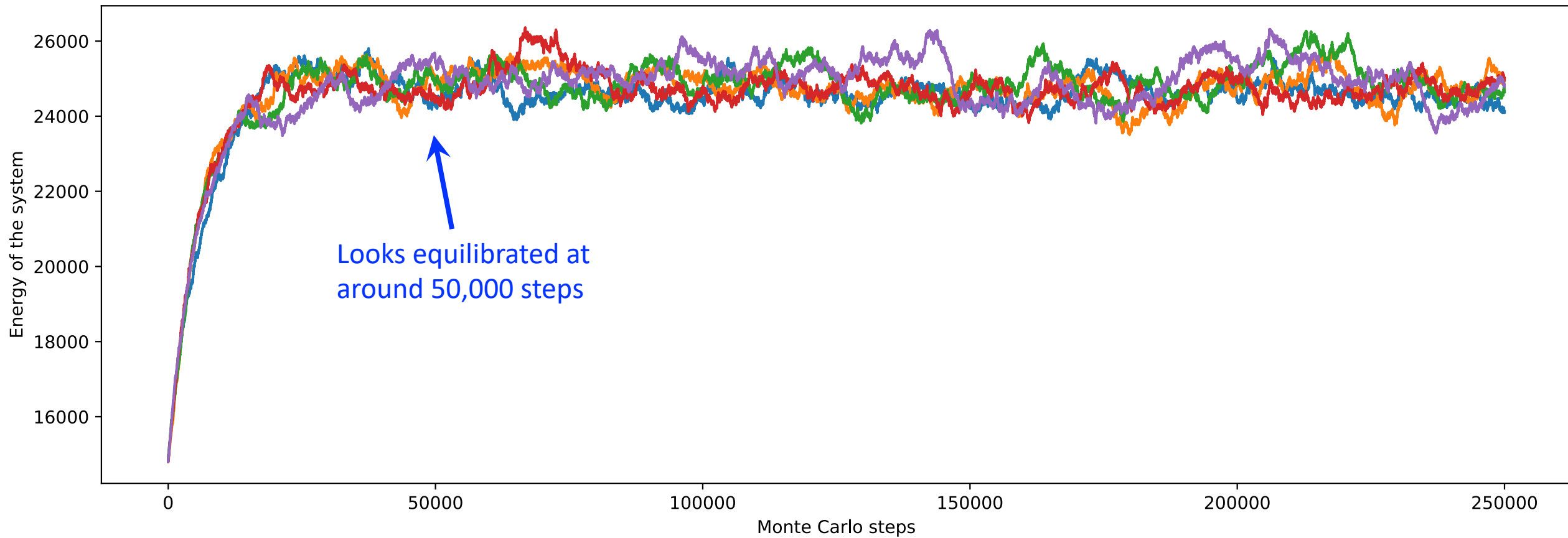
Move set for ideal gas

- Choose set of all moves of a single atom to one of the six “neighboring” states where n_x , n_y , or n_z differ by +/- 1
- Each Monte Carlo step, choose a random particle, chose a quantum number, change it by +/- 1
- Change in total energy just the change for single particle since there are no interactions
 - E.g., increase or decrease n_x of atom i by one:

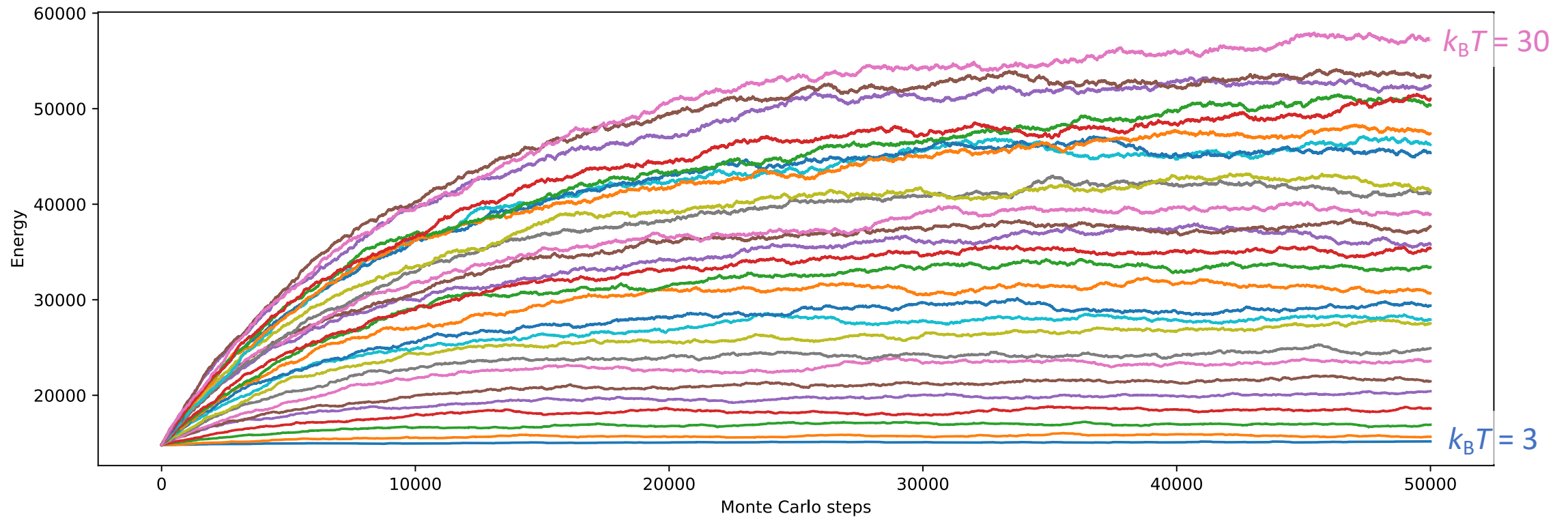
$$\begin{aligned}\Delta E &= \frac{\pi^2 \hbar^2}{2mL^2} [(n_x \pm 1)^2 + n_y^2 + n_z^2] - \frac{\pi^2 \hbar^2}{2mL^2} (n_x^2 + n_y^2 + n_z^2) \\ &= \frac{\pi^2 \hbar^2}{2mL^2} [(n_x \pm 1)^2 - n_x^2] = \frac{\pi^2 \hbar^2}{2mL^2} (\pm 2n_x + 1)\end{aligned}$$

- Note: Reject moves that try to make $n < 1$

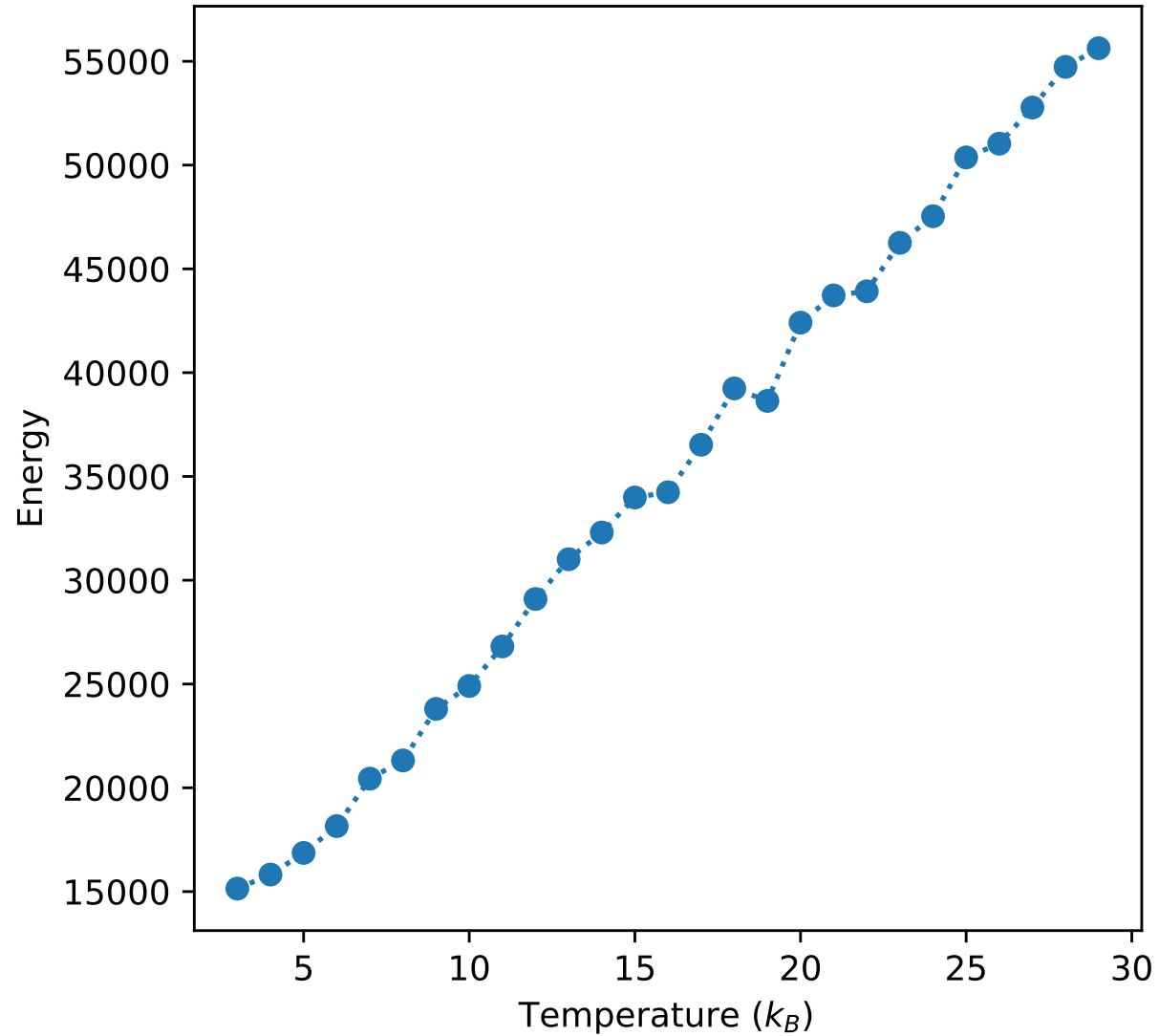
Monte Carlo simulation of ideal gas



Monte Carlo simulation of ideal gas: Dependence on T



Monte Carlo simulation of ideal gas: E vs. T



After class tasks

- Homework 5 is posted, due Nov. 28, 2021 (should be short)
- Final project ideas due Nov.20
- Readings:
 - Newman Sec. 10.3