HIGH PERFORMANCE COMPUTING for SCIENCE & ENGINEERING (HPCSE) I

HOMEWORK 04: Monte-Carlo and MPI

Noah Baumann (noabauma@student.ethz.ch)

Computational **S**cience and **E**ngineering **Lab**
ETH Zürich

09.11.2022

Content

- HW04
- Theory recap: Curse of dimensionality, Monte Carlo Integration and Sampling methods
- Introduction to MPI

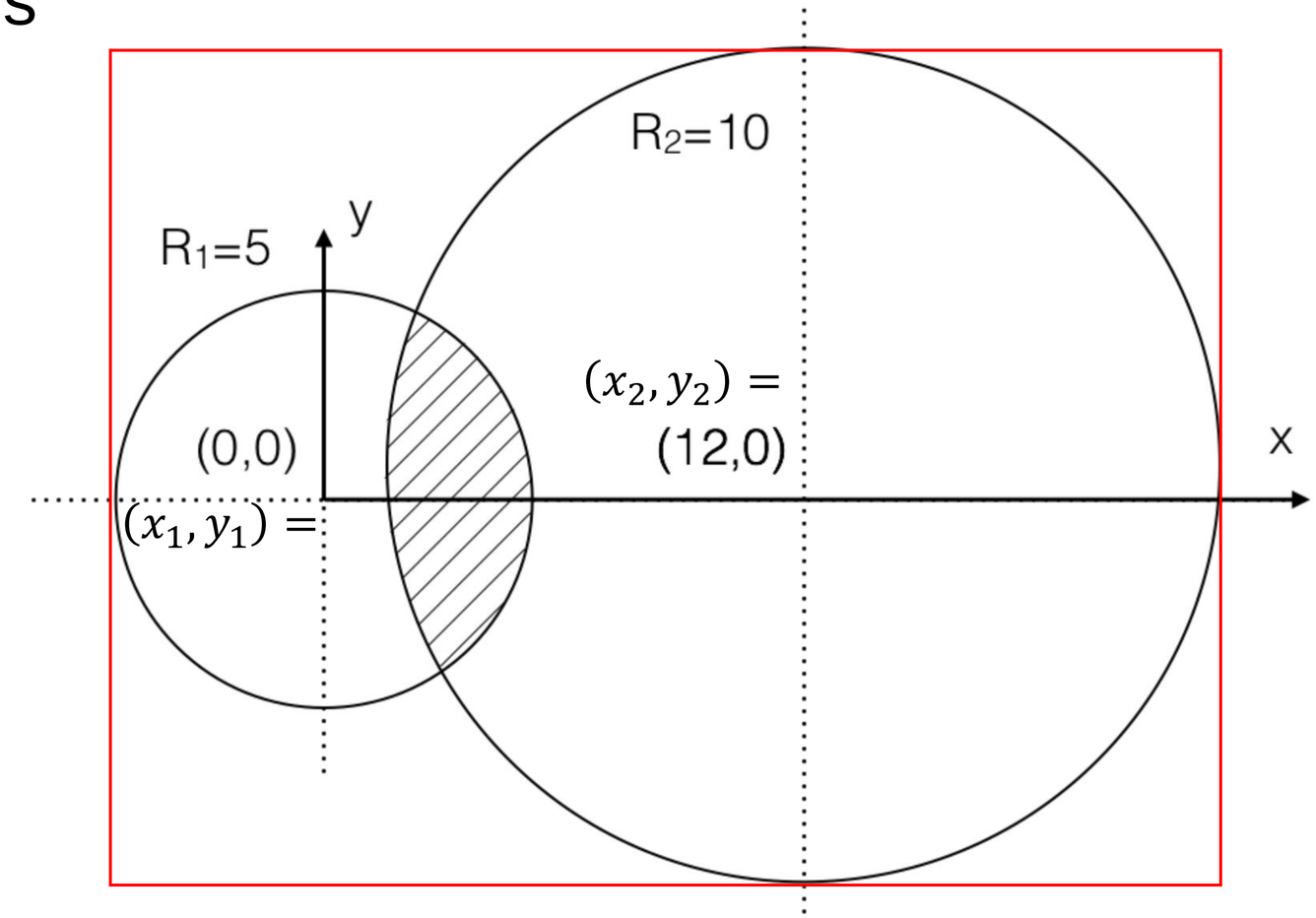
HOMEWORK 4

Q1) Probabilities, Expectation, Variance

- Find the distribution
- Calculating the probability of a given distribution
- Calculate the Expectation and Variance

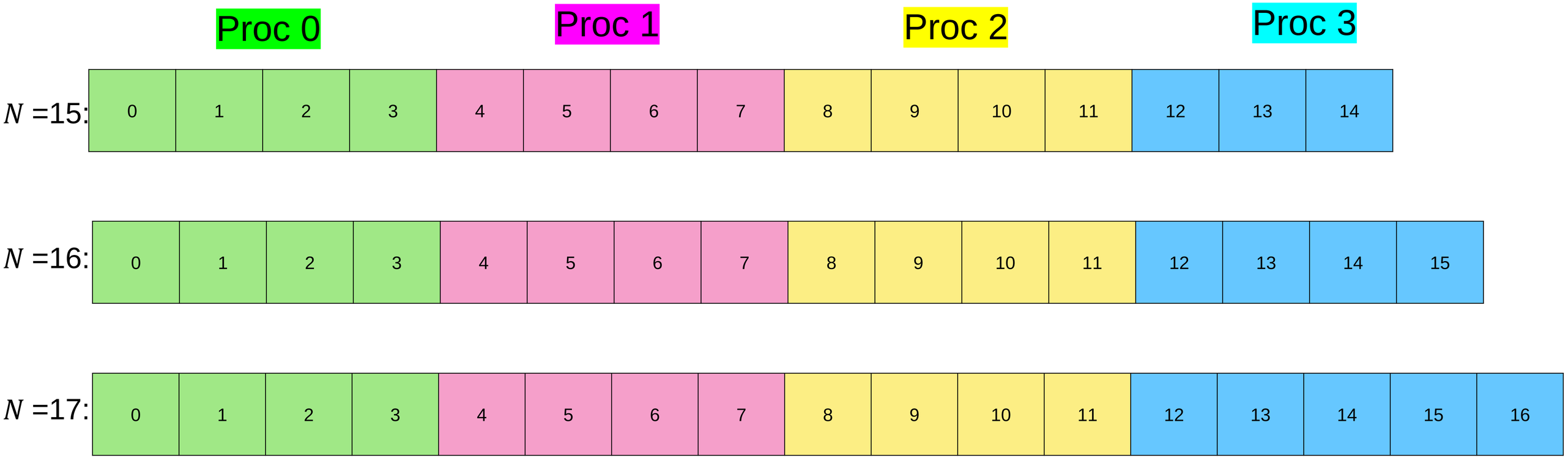
Q2) MC integration with MPI

- R_1, R_2 & x_2 are non-constant variables. Consider this when computing the random samples
- y_1, x_1 & y_2 are constant



Q2) MC integration with MPI

- Distribute N MC samples as balanced as possible
- Consider having 4 processors



Q3) MPI Bug Hunt

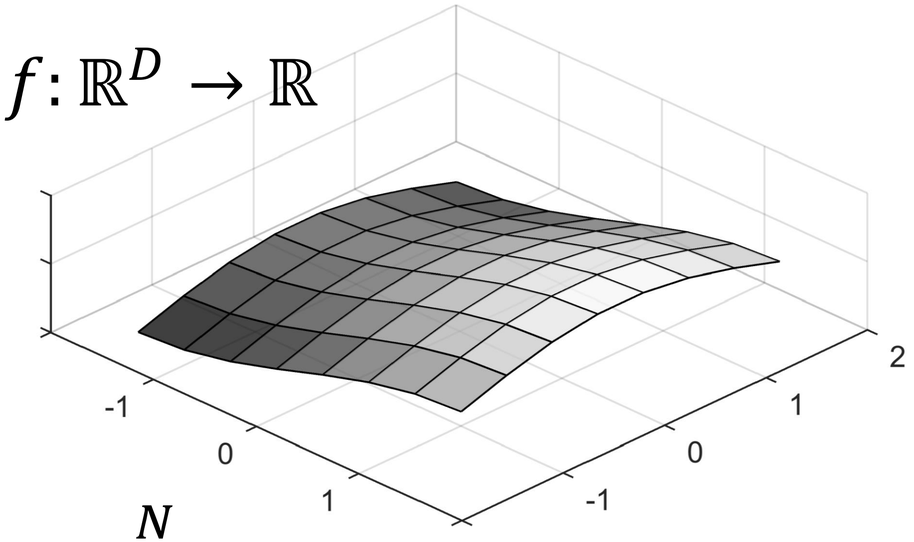
- a) & b) Find the bugs and fix them! Give a solution
- c) State the outputs with different number of ranks

Lecture Recap

Quadrature in Multiple Dimensions

Assume we want to integrate a function in D-dimensions $f: \mathbb{R}^D \to \mathbb{R}$

$$I = \int_{a_1}^{b_1} \cdots \int_{a_D}^{b_D} f(x_1, \ldots, x_D) \mathrm{d}x_1 \cdots \mathrm{d}x_D$$



Let us use Quadrature in each dimension $\int_{a_d}^{b_d} f(x_d) \mathrm{d}x_d \approx \sum_{i_d=1}^{N} w_{i_d} f(x_{i_d})$ with N gridpoints.

This gives

$$I \approx \sum_{i_1=1}^{N} \cdots \sum_{i_N=1}^{N} w_{i_1} \cdots w_{i_N} f(x_{i_1}, \ldots, x_{i_D})$$

Curse of Dimensionality

We found that we can approximate a D-dimensional integral with N points per dimension as

$$I \approx \sum_{i_1=1}^{N} \cdots \sum_{i_D=1}^{N} w_{i_1} \cdots w_{i_N} f(x_{i_1}, \ldots, x_{i_D})$$

We realise for $i_1, \ldots, i_D = 1, \ldots, N$ in total this requires $M = N^D$ function evaluations.

The order of accuracy λ is given in terms of the grid spacing $h = \frac{b-a}{N-1}$,thus we have

$$I - I_Q = \mathcal{O}(h^\lambda) = \mathcal{O}(N^{-\lambda}) = \mathcal{O}(M^{-\lambda/D})$$

Thus the accuracy per sample depends on the dimension of the problem.

Connection to Probability Theory

Warmup questions:

- Assume a uniform distribution $\mathcal{U}([0,1])$ which is defined in an interval $[0,1]$. The variance of a random variable sampled from this probability distribution is

  A. 1
  B. 1/3
  C. 1/2
  D. 1/4
  E. 1/6
  F. 1/12

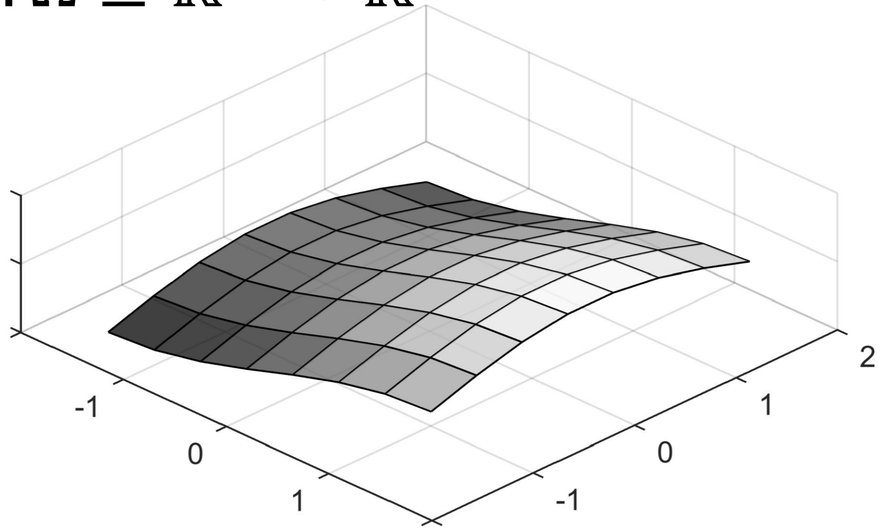
- Which of the following statements is/are True?

  A. A probability density function is always non-negative.
  B. The standard deviation of a random variable is always smaller than the expectation value.
  C. Continuous probability density functions integrate up to one.
  D. The cumulative distribution function F is strictly monotonically increasing (i.e. $F(a) > F(b)$ if a > b)

Connection to Probability Theory

Assume we want to integrate a function in D-dimensions $f: \Omega \subseteq \mathbb{R}^D \rightarrow \mathbb{R}$

$$I = \int_\Omega f(\mathbf{x}) \mathrm{d}\mathbf{x} = |\Omega| \int_\Omega f(\mathbf{x}) \frac{1}{|\Omega|} \mathrm{d}\mathbf{x}$$



Recall that the density for an uniform distribution is

$$p_{\mathcal{U}}(\mathbf{x}) = \frac{1}{|\Omega|}$$

Thus we can see our integral as an expectation value

$$I = |\Omega| \int_\Omega f(\mathbf{x}) p_{\mathcal{U}}(\mathbf{x}) \mathrm{d}\mathbf{x} = |\Omega| \mathbb{E}_{x \sim p_{\mathcal{U}}}[f(\mathbf{X})] \approx |\Omega| \frac{1}{M} \sum_{i=1}^{M} f(\mathbf{x}_i)$$

For a set of independent and identically distributed samples $\{f(\mathbf{x}^{(1)}), \ldots, f(\mathbf{x}^{(M)})\}$ with mean $\mathbb{E}[f(\mathbf{x})] = I$ and $\mathrm{var}[f(\mathbf{x})] = \sigma^2 < \infty$, then we have the Central Limit Theorem

$$\sqrt{M} \left(\frac{1}{M} \sum_{i=1}^{M} f(x_1^{(i)}, \ldots, x_D^{(i)}) - I \right) \longrightarrow \mathcal{N}(0, \sigma^2)$$

Monte Carlo Integration

Use samples from uniform distribution $\mathbf{x}^{(i)} = (x_1^{(i)}, \dots, x_D^{(i)}) \sim \mathcal{U}([a_1, b_1] \times \cdots \times [a_D, b_D])$ with $i = 1, \dots, M$ to approximate

$$I \approx I_M = \frac{|\Omega|}{M} \sum_{i=1}^{M} f(x_1^{(i)}, \dots, x_D^{(i)})$$

Which is an estimate of the value of the integral (itself a random quantity).

Define the error of the Monte Carlo integration as the standard deviation of this estimate

$$\epsilon = \sqrt{\mathrm{var}[I_M]}$$

We find (by the CLT or a direct calculation as done in the script) that this can be written as

$$\epsilon_M = \sqrt{\frac{\mathrm{var}[f]}{M}}$$

Thus we find an error which scales as $\mathcal{O}(M^{-1/2})$ and is independent of the dimension!

"Naive" Monte Carlo Integration

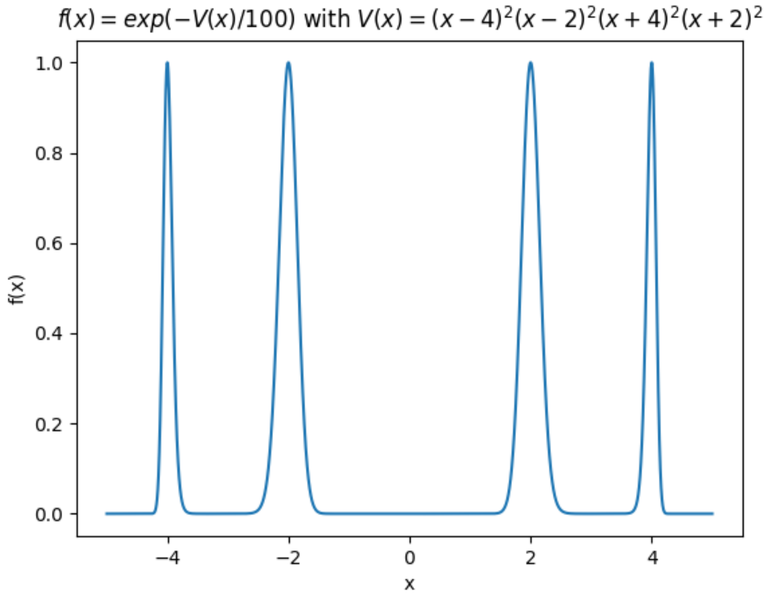
Assume we want to integrate a function in D-dimensions $f: \Omega \subseteq \mathbb{R}^D \to \mathbb{R}$

1. Sample $\mathbf{x}^{(i)} = (x_1^{(i)}, \ldots, x_D^{(i)}) \sim \mathcal{U}(\Omega)$ for $i = 1, \ldots, M$
2. Evaluate function $f(\mathbf{x}^{(i)})$ for $i = 1, \ldots, M$
3. Compute Monte Carlo Estimate of the Integral $I \approx \frac{|\Omega|}{M} \sum_{i=1}^{M} f(\mathbf{x}^{(i)})$

But, what if the function is very peaked?

We will evaluate many $f(\mathbf{x}) = 0$ samples

$\rightarrow$ wasted effort!



Accept-Reject Sampling

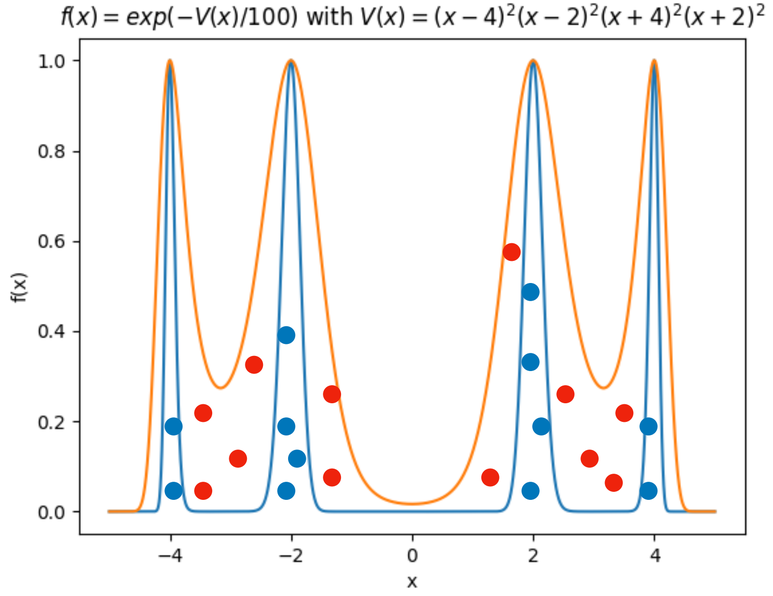
Assume we want to integrate a function in D-dimensions $f: \Omega \subseteq \mathbb{R}^D \rightarrow I \subseteq \mathbb{R}$

1. find simple distribution $h(x)$, define λ s.t. $f(x) \leq \lambda h(x)$
2. draw sample $\mathbf{x}^{(i)}$ from $h(\mathbf{x}^{(i)})$ for $i = 1, \ldots, M$
3. Evaluate function $f(\mathbf{x}^{(i)})$ for $i = 1, \ldots, M$
4. Accept, if $u^{(i)} < f(\mathbf{x}^{(i)}) / \lambda h(\mathbf{x}^{(i)})$, otherwise reject the sample $\quad (u^{(i)} \sim \mathcal{U}_{[0,1]})$
5. Compute Monte Carlo Estimate of the Integral $I \approx |\Omega_{\lambda h(x)}| \dfrac{\#accepted\ samples}{M}$

What if function is very peaked, high irregular of multimodal?

1) Use smart bounding distribution $\lambda h(x)$

2) Importance sampling

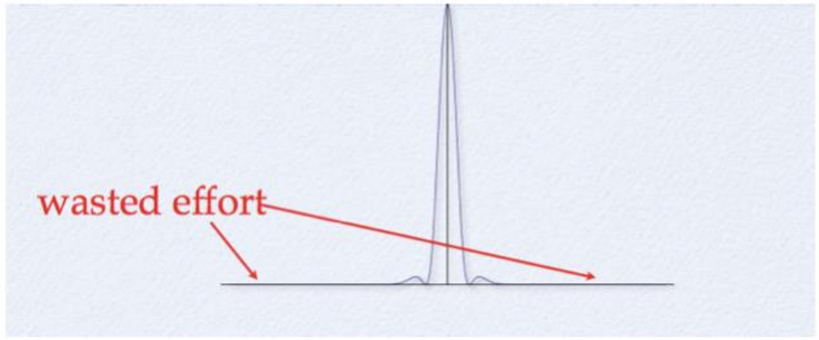


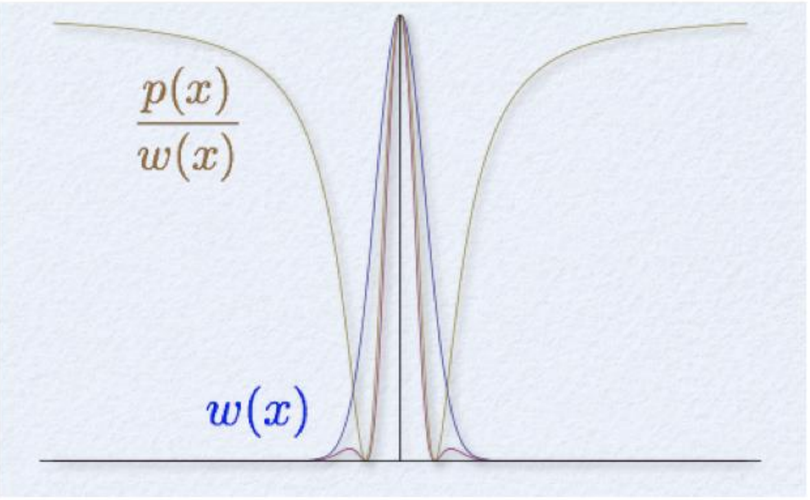
Importance Sampling

Importance sampling allows us to draw samples x that are distributed as probability $w(x)$, instead of a uniform distribution.

Compensate for the **bias** by normalizing $p(x)$ by the very same **"importance" function** $w(x)$ and sample $p(x)/w(x)$ instead, resulting in the following integral:

$$\langle f \rangle_p = \int_a^b f(x) \frac{p(x)}{w(x)} w(x) dx \approx \frac{1}{M} \sum_{i=1}^{M} f(x_i) \frac{p(x_i)}{w(x_i)}$$



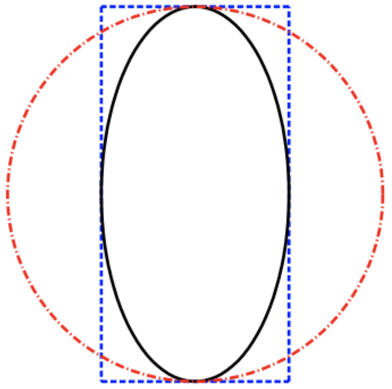


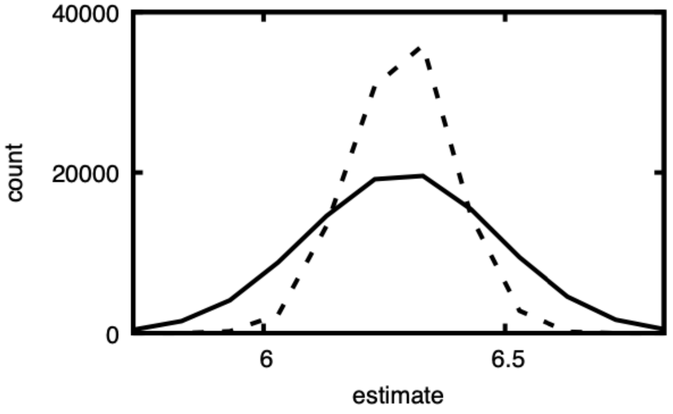
Monte Carlo Integration

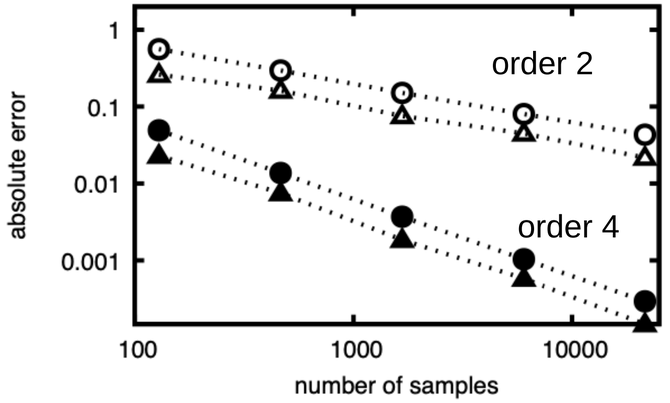
You use the Monte Carlo method to estimate an area of an ellipse. As sampling space you choose two different configurations: a rectangle and a circle.

- You perform Monte Carlo integration 100000 times using the same number of samples every time, and you record a histogram (Histogram of a quantity: Divide the range of values into a series of intervals and count how often the estimate lies in each of these intervals ("bins")) of the area estimates. The histograms for the two sampling spaces are shown below. Which histogram corresponds to which sampling space? Justify your answers.

- Out of the four curves plotted below, only two represent convergence plots for the Monte Carlo method. Identify these two curves. Out of these two, which one corresponds to the rectangular sampling space? Which one corresponds to the circular sampling space? Justify your answers.







Monte Carlo Pseudocode

Assume you are given a closed surface in $\mathbb{R}^3$ described by the relationship $g(x, y, z) = 0$. The interior of this surface is given by $\{(x, y, z)|g(x, y, z) < 0\}$. You assume that your surface lies in the rectangular box $\{(x, y, z)|x_- \leq x \leq x_+, y_- \leq y \leq y_+, z_- \leq z \leq z_+\}$. The following pseudo-code is used to approximate the interior volume defined by the surface based on Monte Carlo. Can you spot the errors?

Algorithm 3 Monte Carlo with bugs

Input:

$x_-, x_+, y_-, y_+, z_-, z_+$, {boundaries of the rectangular box}

Output:

V, {approximation of the interior volume}

Steps:

Compute $|\Omega| = (x_+ - x_-)(y_+ - y_-)$
$K = 0$ # counter
for $i \leftarrow 1, \ldots, N$ **do**
  Sample $x_i \sim uniform(-1, 1)$
  Sample $y_i \sim uniform(-1, 1)$
  Sample $z_i \sim uniform(-1, 1)$
  Compute $\tilde{g} = g(x_i, y_i, z_i)$.
  if $\tilde{g} \geq 0$ **then**
    $K = K + 1$
  end if
end for
$V = K|\Omega|$
return V

1. $|\Omega|$ wrong. $|\Omega| = (x_+ - x_-)(y_+ - y_-)(z_+ - z_-)$
2. Random samples wrong distribution. $x_i \sim uniform(x_-, x_+)$
3. Wrong if-statement. $\tilde{g} < 0$
4. Wrong V. $V = \frac{K}{N}|\Omega|$

MPI Send/Recv example

- **M**essage **P**assing **I**nterface

on Euler:
module load open_mpi

```
 1 #include <mpi.h>
 2
 3 int main(int argc, char **argv)
 4 {
 5     MPI_Init(&argc, &argv);
 6
 7     printf("Hello HPCSE class!\n");
 8
 9     MPI_Finalize();
10     return 0;
11 }
```

include header

Can NOT call MPI routines!

Initialize MPI environment

Can call MPI routines

Finalize MPI environment

Can NOT call MPI routines!

```
> mpic++ hello.cpp -o hello
> mpiexec -n 2 ./hello
Hello HPCSE class!
Hello HPCSE class!
```