E2 Monte Carlo integration

To get familiar with the Monte Carlo method you are here asked to first consider a one-dimensional integral. You will implement importance sampling using the Transformation method. You will then implement the Metropolis algorithm and use that to solve a three dimensional integral. You will also consider the evaluation of the statistical inefficiency s, which can be used to determine proper error bars.

This task is paired with code questions in Yata that are mandatory to pass before being allowed to present. The Yata questions will help you verify that your code is correct.

You will use the GNU Scientific Library (GSL) [1] for random number generation, as the random numbers generated by the built-in functions in C are of subpar quality. You can refer to C2 for examples on how to do this.

Task

1. Consider the one-dimensional integral

$$I = \int_0^1 f(x)\,\mathrm{d}x = \int_0^1 x(1-x)\,\mathrm{d}x.$$

   Do E2 task 1 on Yata to implement a routine for Monte Carlo integration, without the use of importance sampling.

2. The efficiency of the Monte Carlo method can be improved using importance sampling. Do E2 task 2 on Yata, where you calculate the same integral with this technique.

   Convince yourself that the importance sampling is working correctly by copying the code to your computer and plotting a histogram of the generated points.

3. Calculate the integral and estimate the error using $N = 10^1$, 10^2, 10^3, and 10^4. Compare your results with and without importance sampling. Compare to the exact value. (1 p)

4. Consider now the three-dimensional integral

$$I = \pi^{-3/2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x^2+x^2y^2+x^2y^2z^2) \exp\left[-(x^2+y^2+z^2)\right] \mathrm{d}x\mathrm{d}y\mathrm{d}z$$

   Use

$$w(x,y,z) = \pi^{-3/2} \exp\left[-(x^2+y^2+z^2)\right]$$

   as weightfunction, where

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z = 1,$$

[1] `https://www.gnu.org/software/gsl/`

and the Metropolis algorithm to generate points according to the above weightfunction. A good choice for the trial changes $\tau_{m \to n}$ are

$$x_n = x_m + \Delta(r - 0.5)$$

and similarly for y and z, and where r is a uniform random number on the interval [0,1] and Δ is a parameter.

Implement the functions necessary for computing the integral in E2 task 4 on Yata.

5. Copy the code that you have verified on Yata to your computer, and compute the integral. You do not need to give error bars. For this task, you need to choose a value of Δ. A "rule of thumb" is to choose Δ such that about half or slightly less of the trials becomes accepted. (1 p)

6. In large scale Monte Carlo simulations it is important to determine the statistical inefficiency s to be able to properly estimate the errors. Consider a dataset of N sequential values f_i, $i = 1, 2, \dots N$. There are two common ways to estimate the statistical inefficiency s.

   One can determine the auto-correlation function

   $$\Phi_k = \frac{\langle f_{i+k} f_i \rangle - \langle f_i \rangle^2}{\langle f_i^2 \rangle - \langle f_i \rangle^2}, \quad \Phi_{-k} = \Phi_k$$

   and estimate s from

   $$\Phi_{k=s} = \exp(-2) \approx 0.135$$

   This is based on the assumption that the correlation function Φ_k decays exponentially.

   Alternatively, one can consider the method of block averaging. In this method new blocked variables

   $$F_j = \frac{1}{B} \sum_{i=1}^{B} f_{i+(j-1)B}$$

   have to be determined for various block sizes B. If the block size becomes sufficiently large ($B > s$) the blocked variables F_j become statistically independent and

   $$s = \frac{B \, \mathrm{Var}[\,F\,]}{\mathrm{Var}[\,f\,]}$$

   Implement the two methods of estimating the statistical inefficiency in E2 task 6 on Yata.

7. On the course Gitlab you find the text file `MC.txt` containing $N = 10^6$ values f_i from a Monte Carlo simulation. Use this data set and determine s using both the method based on evaluation of the correlation function as well as on block averaging.

   Do you get consistent results for s using the two different methods? Notice, if the mean value of f is large compared with the typical deviations from the mean value the above formula for the correlation can introduce numerical inaccuracies. It is then better that first to determine the mean value and then calculate the correlation for the deviation from the mean value. (2 p)