Project: numerical integration with Monte Carlo method

Tomasz Chwiej

6th June 2025

1 Introduction

1.1 Theory

Our aim is to numerically approximate the value of definite integral

$$I = \int_a^b g(x)dx \tag{1}$$

For this purpose we will use the Monte Carlo method which is grounded in computer statistics. Before we go further we need to introduce to the integral the **probability density function** $f(x)$ which is nonegative and normalized

$$\bigwedge_{x\in(a,b)} f(x) \geq 0 \qquad \wedge \qquad \int_a^b f(x)dx = 1 \tag{2}$$

$$I = \int_a^b \frac{g(x)}{f(x)} f(x)dx = \int_a^b h(x)f(x)dx, \qquad \left(h(x) = \frac{g(x)}{f(x)}\right) \tag{3}$$

Now recall some information from statistics about how we calculate the expectation value of random variable. If we think about the variable x to be a random number drown from proability distribution $f(x)$ our integral is exactly the expectation value of random variable $h(X)$ (value of function $h(X)$ becomes random due to randomness of its argument X). For the expectation value marked $E(h)$ in statistics below it is used the most common notion met in physics that is $\langle h \rangle$

$$I = E(h) = \langle h \rangle \tag{4}$$

The expectation value can be estimated (approximated) in an experiment: (i) by drawing many times (N) the random numbers from distribution $X_i \sim f(x)$, (ii) computing the values of $h(X_i)$ and (iii) averaging values of $h(X_i)$

$$\langle h \rangle \approx \overline{h} = \frac{1}{N}\sum_{i=1}^{N} h(X_i), \qquad X_i \sim f(x) \tag{5}$$

and similarly we may calculate any $n-th$ moment

$$\langle h^n \rangle \approx \overline{h^n} = \frac{1}{N}\sum_{i=1}^{N} h^n(X_i), \qquad X_i \sim f(x) \tag{6}$$

In order to perform this experiment with computer we need the **random number generator** for distribution $f(x)$. Above estimation of integral value is done within language of statistics so it is computed not precisely but with some uncertainty which must be also determined. The uncertainty of the expectation value is expressed by means of **the second central moment** i.e. **variance**

$$
\begin{aligned}
var(h) &= \int_a^b (h(x) - \langle h \rangle)^2 f(x) dx \\
&= \int_a^b h^2(x) f(x) dx - 2\langle h \rangle \int_a^b h(x) f(x) dx + \langle h \rangle^2 \int_a^b f(x) dx \\
&= \langle h^2 \rangle - \langle h \rangle^2
\end{aligned} \tag{7}
$$

Variance can be thus estimated with equations 5 and 6 (for $n = 2$). Because variance equals the squared standard deviation we get the measure of uncertainty of single random variable

$$\sigma_h = \sqrt{var(h)} \tag{8}$$

and more importantly the uncertainty for the mean value $\overline{h}$, i.e. the variance

$$var(\overline{h}) = \frac{var(h)}{N} \tag{9}$$

and the standard deviation of the mean value

$$\sigma_{\overline{h}} = \frac{\sigma_h}{\sqrt{N}} \tag{10}$$

The last equation 10 is characteristic for MC integration, although σ_h would tend to finite non-zero value, the standard deviation of mean value asymptotically tends to zero as inverse square root of the total number of trials.

1.2 Numerical aspects of MC integration

Based on a piece of theory given in previous subsection we are ready to construct MC algorithm for integration, we just need an explicit formula for the integrand $g(x)$ (or more precisely $h(x) = g(x)/f(x)$), the **probability density function** function $f(x)$ and corresponding random number generator. Then within a single loop we need to compute the first and the second moments (Eqs. 5 and 6), the variance (Eq.7) as well as the standard deviation of single variable (Eq.8) and the standard deviation of the mean value (Eq.10). Pseudocode for such integration may look as follows

```
initialize: N, k = 1, s_1 = 0, s_2 = 0

for n from 1 to N do

      X ~ f(x) - draw random number from pdf
     s_1 += h(X) - acumulate data for 1-st moment
     s_2 += (h(X))^2 - acumulate data for 2-nd moment

     if n mod 10^k == 0 then
             k++

             h_1 = s_1/n - mean
             h_2 = s_2/n - second moment
```

```
        var = h_2 - (h_1)^2  - variance
    σ = √var  - standard deviation
    σ_mean = σ/√n  - standard deviation of mean
    write to file:   n, h_1, var, σ_mean
end if

end do
```

In pseudocode there is used the **modulo division** that is only for $n = 10^k$, $k = 1, 2, 3, 4, 5, 6$ we calculate the mean and standard deviation of mean and write both values to file, this limits the number od data which will be processed next (e.g. plotted).

2 Practical part

1. Implement on a computer the MC integration pseudocode given in Sec.1.2
2. Use your program to estimate the value of the following integral

$$I_1 = \int_a^b g(x)dx \tag{11}$$

   Conduct MC integration for the following cases (a) and (b), then make plots of the integral values in dependence of n marking the uncertainties $\sigma_{\overline{h}}$ (errorbars) and the plots of variance, in plots mark the exact values for comparison.
   In (a) and (b) use uniform random number generator $U(a, b)$ (it requires the use of **cstdlib** header)

```
double uniform(double a, double b){
        double u=(double)rand()/RAND_MAX
        double x=a+(b-a)*u
        return x
}
```

   (a) simple integration for: total number of trials $N = 10^6$, $a = 0$, $b = \pi/2$, $g(x) = \sin(x)$, $f(x) = \frac{1}{b-a} = \frac{2}{\pi}$, $h(x) = \frac{g(x)}{f(x)}$. Calculate: $\overline{h}$, $var(\overline{h})$ and $\sigma_{\overline{h}}$, then compare these values with exact ones: $I_1 = 1$, $var(h) = (\pi^2/8) - 1$ for the number of trials $n = 10^k, \quad k = 1, 2, 3, 4, 5, 6$.

   (b) integration using antithetic variables (we modify the integrand by using variables X and $b - X$ which are highly correlated): total number of trials $N = 10^6$, $a = 0$, $b = \pi/2$, $g(x) = (\sin(x) + \sin(b - x))/2$ [x and (b-x) are antithetic variables], $f(x) = \frac{1}{b-a} = \frac{2}{\pi}$, $h(x) = \frac{g(x)}{f(x)}$. Calculate: $\overline{h}$, $var(\overline{h})$ and $\sigma_{\overline{h}}$, then compare these values with exact ones: $I_1 = 1$, $var(h) = (\pi^2/16) + (\pi/8) - 1$ (variance will change value) for the number of trials $n = 10^k, \quad k = 1, 2, 3, 4, 5, 6$.

3. Modify the code to calculate the integral

$$I_2 = \int_a^b g(x)\, f(x)dx \tag{12}$$

for: total number of trials $N = 10^6$, $a = 0$, $b = \infty$, $g(x) = \log(x)$ and $f(x) = \exp(-x)$. Our probability density function $f(x)$ is explicitly given and is normalized for $x \in (0, \infty)$ so we only need the random number generator for this exponential distribution

```
double exponential(){
        double u=(double)rand()/RAND_MAX
        double x=-log(U)
        return x
}
```

The exact value of integral equals $I_2 = -\gamma = -0.5772156649$ (Euler constant) whereas the variance is $var(g) = \pi^2/6$. Again make plots of the integral values and the variance for the number of trials $n = 10^k, \quad k = 1, 2, 3, 4, 5, 6$.

4. Report is not required for this project, the partial grade will be granted based on the student's activity during the classes.

3 Example results

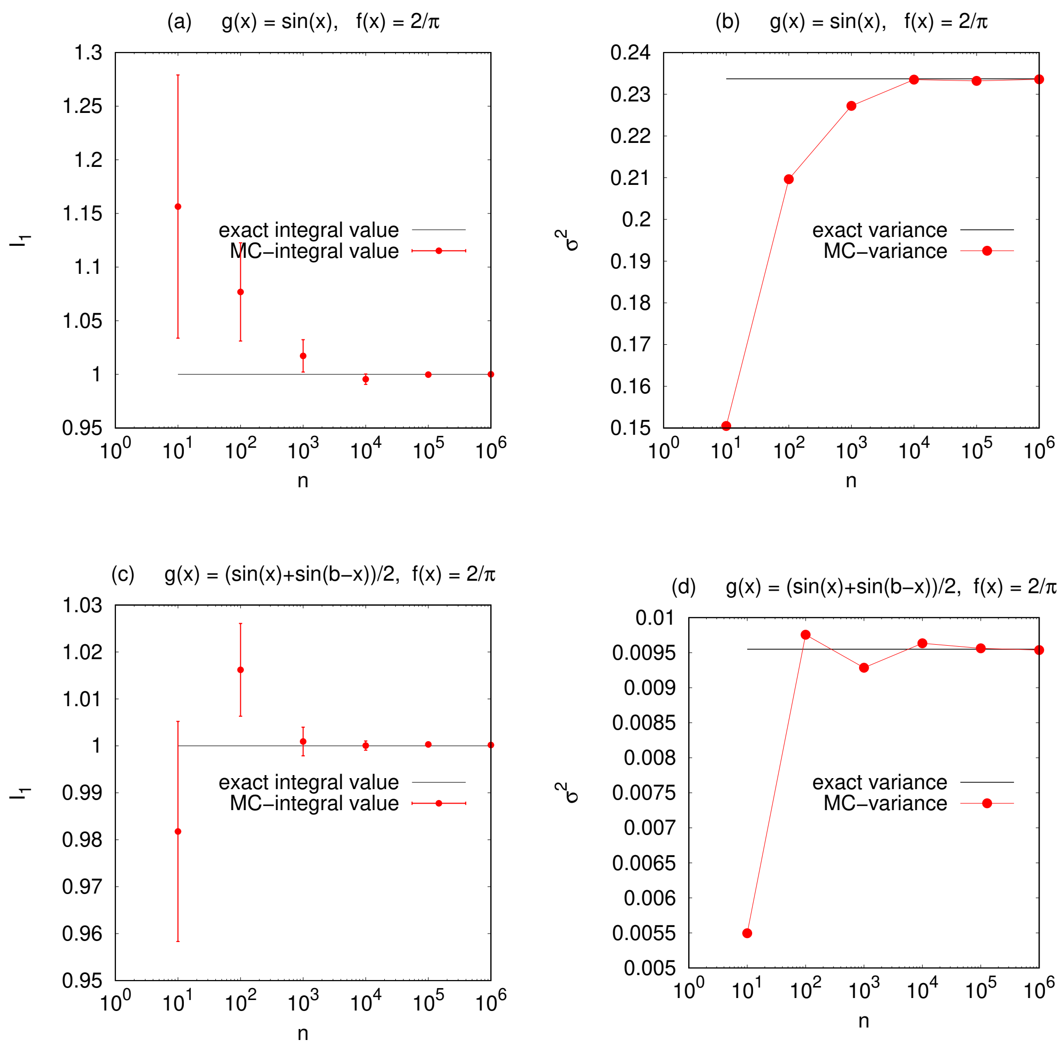

Figure 1: Estimation of integral I_1 with MC method: (a) and (b) obtained with standard approach, in (c) and (d) antithetic variables were used. Note that the errorbars in (c) are one order of magnitude smaller than in (a) - small modification of integrand $g(x)$ makes that the errors are partly canceled. The lengths of errorbars are equal to standard deviation of mean value $\sigma_{\overline{h}}$.

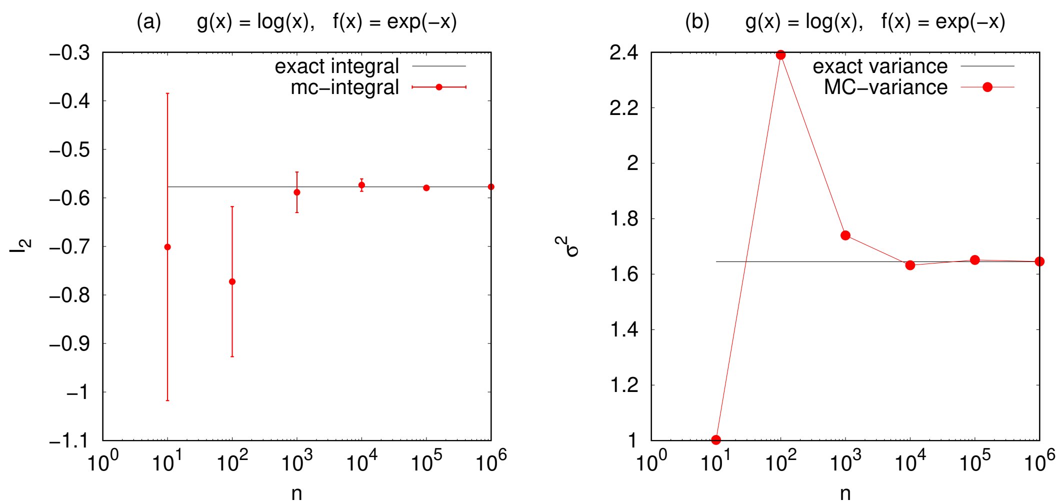

Figure 2: MC estimation of integral I_2 (a) and variance (b). The lengths of errorbars in (a) are equal to standard deviation of mean value $\sigma_{\overline{g}}$.