

TOPIC 4 - MONTE CARLO INTEGRATION : 1D HIT & MISS AND SAMPLE MEAN METHODS

*This worksheet accompanies the EJS simulation MCIntegration\_No1\_1DHitMissSampleMean.jar*

---

**The challenge:** Our task is to evaluate the following integral<sup>1</sup>:

$$\int_a^b f(x) dx.$$

Geometrically, this integral gives the area under the curve  $f(x)$  between  $x = a$  and  $x = b$  (regions where  $f(x)$  is negative contribute negatively to the area).

Let us consider the case where  $f(x)$  is non-negative. The simplest Monte Carlo method of integration is to enclose the function  $f(x)$  between  $x = a$  and  $x = b$  by a region of known area and then to determine the fraction of this area beneath  $f(x)$  (one can always choose a rectangular region  $(b - a)$  units wide and  $\max_{a \leq x \leq b} (f)$  high). Next, one randomly generates an  $(x_i, y_i)$  uniformly drawn from the bounding region. If  $y_i < f(x_i)$ , then this point is within the area defined by  $f(x)$  and the  $x$ -axis. By repeating this procedure many times, one can estimate the required fraction and, hence, the required integral. This method amounts to the **hit & miss integration technique**. In fact, this idea may have occurred to you when using the hit & miss sampling method in Topic 1 (there the acceptance rate gave the fraction).

Another technique makes use of the definition of the average of  $f(x)$ :

$$\bar{f} = \frac{1}{(b - a)} \int_a^b f(x) dx.$$

If we randomly generate  $N$  points  $\{x_1, x_2, \dots, x_N\}$  over the interval  $[a, b]$ , then we can use an approximation to the average to estimate the integral:

$$\int_a^b f(x) dx \approx \frac{(b - a)}{N} \sum_{i=1}^N f(x_i).$$

This method is known as **sample mean integration**.

In both the hit & miss and sample mean approaches, points have to be generated a number of times. The number of points (or steps) used in the calculation is known as the number of trials  $N_{\text{trials}}$ . Since both techniques give an estimate of the result, it is best to repeat the whole process a number of times. The number of repeats is known as the number of runs  $N_{\text{runs}}$ . One can then use an average over runs as a best approximation to the integral.

**Questions:**

1. Write pseudo-code to perform the hit & miss integration technique on a general function  $f(x)$  (i.e. do not assume that the function is non-negative).

---

<sup>1</sup>It will be shown that Monte Carlo techniques outperform traditional methods of integration only in higher dimensions. Our approach here, however, is to use the 1D case to introduce the concepts.

2. Write pseudo-code to perform the sample mean integration technique on a general function  $f(x)$ .
3. Consider the integral:

$$\int_0^2 \cos(x) dx.$$

In this case, the integral is easily determined analytically, allowing us to calculate the actual error in our estimates.

- (a) Solve the integral analytically.
  - (b) Use the associated EJS simulation to investigate the use of the hit & miss method to perform this integral. Use the default setting of  $N_{\text{trials}} = 100$  and  $N_{\text{runs}} = 200$ .
    - i. Investigate the functional form of the histogram of run results. Why does the histogram have so many unoccupied bins (provided the bin width is small enough)? How do the uncertainty measures compare to the errors?
    - ii. Incorporating all runs, how would you quote the final result?
    - iii. If you were able to perform just one run, what would you then quote as the result?
    - iv. In this case, it is easy to choose an optimal bounding box. However, suppose that you chose a larger box. What effect would this have on your result?
    - v. Investigate the effect of varying  $N_{\text{trials}}$  and  $N_{\text{runs}}$ .
  - (c) Use the EJS simulation to investigate the sample mean technique of solving this integral.
    - i. If you were able to perform just one run, what would you then quote as the result?
    - ii. Investigate the functional form of the histogram of run results.
    - iii. Incorporating all runs, how would you quote the final result?
    - iv. How do the uncertainty measures compare to the errors?
4. Consider the (again easily analytically evaluated) integral:

$$\int_0^3 x^2 dx.$$

- (a) Compare the results of the hit & miss and sample mean approaches to solving this integral. Use the default setting of  $N_{\text{trials}} = 100$  and  $N_{\text{runs}} = 200$ .
- (b) Given the simple integrand, it is straightforward to calculate its variance. Use this result to derive the uncertainty in a single run result and the uncertainty in the average over multiple runs using the sample mean approach. How do these calculated values compare with the actual errors and uncertainties quoted by the simulation?
- (c) Select **Investigate Uncertainty/Error Dependence on Number of Trials** and generate a set of results with different  $N_{\text{trials}}$  and  $N_{\text{runs}} = 20$  (the simulation automatically doubles  $N_{\text{trials}}$  after completion of each set of runs, but the user can override this). How well do the uncertainties match the actual errors? What is the functional dependence of the errors and uncertainties on  $N_{\text{trials}}$ ? This functional dependence is the same for  $d$ -dimensional integrals, whereas the trapezoidal integration method has an error that varies as  $N^{-\frac{2}{d}}$  (i.e. it depends on the number of dimensions). With this in mind, when does Monte Carlo integration outperform traditional methods?