4 Definite Integrals

4.1 Buffon's needle

Random sampling can be used to obtain statistical estimate of values. Buffon's Needle (1777) is a very early example of this idea, where π is calculated as an integral performed by a purely random process, that of repeatedly dropping a needle of length *l* onto a set of parallel lines of separation *l* (see figure 1) The needle will cross the nearest line, at a

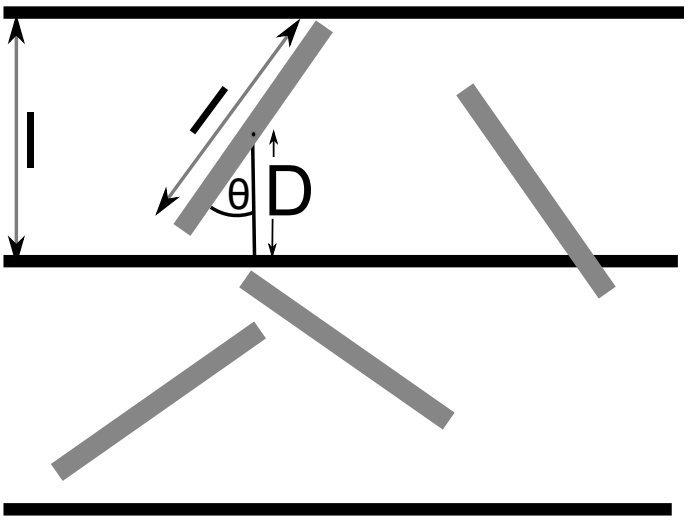

Figure 1: A needle dropped onto some lines

distance D from the centre of the needle, if $D \leq \frac{1}{2}\sin\theta$, then the possible alignments of the needle can be plotted as in figure 2 here the rectangle

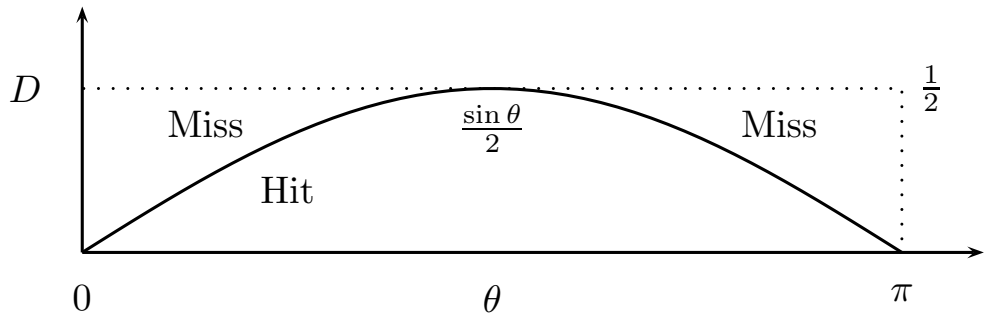

Figure 2: Ratio of 'hits' to total trials.

is every possible position for the needle. Then the fraction of times a randomly dropped needle crosses a line is given by the ratio between the total area and places where a needle counts as a hit event:

$$\int_0^\pi \frac{\sin\theta}{2} d\theta \Big/ \frac{\pi}{2} = 2/\pi$$

This gives a statistical estimate of π.[1]
This approach is very similar to the rejection sampling method we discussed in the last notes. But it is also a sort of numerical integration of the area under a curve.

4.2 Computing integrals

In a more general sense, consider the integral

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

Deterministic integration

Our usual approach would (at its simplest) would be perhaps to adopt the *trapeziodal rule*. Divide the interval $[a, b]$ into n equal intervals such that

$$\Delta x = \frac{b-a}{n} \tag{2}$$

and hence the sequence points

$$\{x_i : x_i = x_0 + i\Delta x, i = 0, \cdots, n-1; x_0 = a; x_n = b\}. \tag{3}$$

Then the approximation to the integral is

$$I = \left[\frac{1}{2}f(x_0) + \sum_{i=1}^{n-1} f(x_i) + \frac{1}{2}f(x_n)\right] \Delta x \tag{4}$$

[1] In 1901, Italian mathematician Mario Lazzarini estimated π this way, obtaining 355/113, which is accurate to the 7$^{\text{th}}$ decimal place. However, there is some evidence that he may have carefully chosen the number of trials to increase the chances of finding this ratio [1].

Monte Carlo Integration

In contrast, a Monte Carlo approach is based on the the *mean value* theorem, which is that the definite integral is equal to the average value of $f(x)$ inside the interval times the width of the interval. That is

$$I = (b-a)\overline{f} \tag{5}$$

We can approximate this by

$$I \approx (b-a)\langle f \rangle = \frac{(b-a)}{n}\sum_{i=1}^{n} f(x_i) \tag{6}$$

where the x_i are random number distributed uniformly in the interval $a \leq x_i \leq b$ and n is the number of trials.[2] Note the similarity to the trapezoidal result, except that in this case the points x_i are randomly positioned not equally spaced. Also note that since the distribution function is uniformly distributed, in the limit of an infinite number of points, the trapeziodal rule and this Monte Carlo approach are identical.
We can quantify the possible error in the sampling using the *variance* in the value

$$(\sigma)^2 = \frac{1}{n}\left(\left\langle f_n^2 \right\rangle - \langle f_n \rangle^2\right) \tag{7}$$

where we have written f_n to mean $f(x_n)$, and the average of a quantity is defined by:

$$\langle A_n \rangle = \frac{1}{n}\sum_{i=1}^{n} A_i \tag{8}$$

Even more useful is the RMS (root mean squared) deviation from the average, σ, which is the one standard deviation error bar for the estimated average value ($\approx 68\%$ confidence that the correct value is with $\pm$ this error range around the mean value).
The error term in the trapeziodal rule scales as $O(n^{-2})$. That for the Monte Carlo method is very much worse, $O(n^{-1/2})$. So why use Monte Carlo methods for integrals at all?

The multi-dimensional integral formula

If we are interested in doing integrals in more than one dimension (let's call the number of dimensions d) then the Monte-Carlo integration will still

[2] Dure to the "law of large numbers" the value of $(\sum_{i=1}^{n} f(x_i))/n$ will eventually give $\overline{f}$.

yield an error which behaves like $O(n^{-1/2})$, but the trapezoidal rule will give an error which goes like $O(n^{-2/d})$. It can be readily seen that when $d > 4$ that the Monte-Carlo integration overtakes the trapozoidal rule in terms of accuracy per point. We note that simpsons rule (and other deterministic techniques) can give better errors than the trapazoidal rule but that it (or they) still become less accurate with the number of dimensions in a similar way.
As we would expect, multi-dimensional integrals such as

$$I = \int_a^b \int_c^d f(x, y)\, dxdy \tag{9}$$

are given by

$$I = \frac{(b-a)(d-c)}{n} \sum_{i=1}^{n} f(x_i, y_i) \tag{10}$$

and similarly for higher dimensional integrals.

5 Importance sampling

Consider the multi-dimensional integral,

$$I = \int_D F(\mathbf{R})\, d\mathbf{R} \tag{11}$$

where $\mathbf{R}$ is an N-dimensional vector and D is the domain of integration. Lets look now at the behaviour of $F(\mathbf{R})$. If it were constant with $\mathbf{R}$ then only one random sample would be required to calculate the integral, I. Similarly, if it were smoothly varying then fewer points would be required. If we choose our random numbers so that they aren't uniform but follow some distribution, $W(\mathbf{R})$ which mimics the changes in the function $F(\mathbf{R})$, then we would expect much faster convergence in integrals. The integral can now be approximated by

$$I = \frac{1}{N} \sum_{i=1}^{M} \frac{F(\mathbf{R}_i)}{W(\mathbf{R}_i)} \tag{12}$$

where $\mathbf{R}_i$ is now sampled according to the distribution function, $W(\mathbf{R}_i)$. For this to work, $W(\mathbf{R})$ needs to be normalised correctly so that

$$\int_D W(\mathbf{R})\, d\mathbf{R} = 1 \tag{13}$$

We can re-write our integral as:

$$I = \int W(\mathbf{R})\, G(\mathbf{R})\, d\mathbf{R} \tag{14}$$

where $W(\mathbf{R})$ satisfies the normalisation condition that

$$\int W(\mathbf{R})\, d\mathbf{R} = 1 \tag{15}$$

and $G(\mathbf{R})$ is given by

$$G(\mathbf{R}) = \frac{F(\mathbf{R})}{W(\mathbf{R})} \tag{16}$$

and the problem is solved if $G(\mathbf{R})$ is a smooth function of $\mathbf{R}$ (i.e. nearly a constant). The integral is given by:

$$I = \frac{1}{N} \sum_{i=1}^{M} G(\mathbf{R}) \tag{17}$$

The key to importance sampling is picking the function $W(\mathbf{R})$ so that $G(\mathbf{R})$ is as smooth as possible.
As an example consider the (one dimensional) integral,

$$I = \int_0^1 e^{-x^2} dx, \tag{18}$$

we will denote our sampling function by $p(x)$. We can either use standard sampling (i.e. p(x) = 1) or use importance sampling, a suitable function in this case is p(x) = Ae^{-x} where we pick A so that $p(x)$ is normalised over the range $[0, 1]$.
Analogously to equation (8) in the last notes, we can write that:

$$y_i = -\ln\left(1 - \frac{x}{a}\right) \tag{19}$$

where x is a random number uniformly distributed in the range $[0, 1]$. We can find A by ensuring that the normalisation holds, this yields:

$$A = \frac{1}{1 - e^{-1}} \tag{20}$$

and so the value of the integral is given by

$$I = \frac{1}{n} \sum_{i=1}^{n} \frac{e^{-y_i^2}}{p(y_i)} \tag{21}$$

The weighting function, $W(\mathbf{R})$ used for sampling the integral, **must** be correctly normalised to evaluate the integral correctly, can we remove this requirement?

$$I = \frac{\sum_{i=1}^{M} F(\mathbf{R}) / W(\mathbf{R})}{\sum_{i=1}^{M} 1/W(\mathbf{R})} \tag{22}$$

References

[1] Lee Badger, "Lazzarini's Lucky Approximation of π", Mathematics Magazine 67, 83–91 (1994).

[2] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, "Numerical Recipes" (Cambridge, 1986).