3 Monte Carlo methods

In this section we will learn how random numbers are used to calculate integrals, and otherwise simulate random events in physics.

The starting point is to realise that integrals can be recast as an averaging process:

$$I = \int_{x_1}^{x_2} f(x)\mathrm{d}x = (x_2 - x_1)\langle f(x)\rangle, \tag{22}$$

where the angle brackets simply denote the average of the function over that region. Furthermore, we can estimate an integral by evaluating a series of points in the appropriate range x_i, such that

$$\langle f(x)\rangle \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i) \tag{23}$$

Putting these together gives us an approximate form of integration

$$I \approx (x_2 - x_1)\frac{1}{N}\sum_{i=1}^{N} f(x_i). \tag{24}$$

The question is now, how do we choose a sensible set of points x_i to evaluate? The punch-line has already been spoiled at the start of this section: random sampling. There are a few reasons we can give to justify why to do this instead of choosing a grid of sampling points as we did in the previous section.

- Avoiding some systematic bias: imagine attempting to integrate an oscillatory function which grows in frequency, say $cos(x^2)$. One issue with a regular grid is you will inevitably have regions where the grid is *skipping* peaks[1].
- Random samples are additive: you can always improve a previous measurement by repeating the measurement and adding the results in quadrature. On a grid you will get the same results unless you increase the grid size, at which point you repeat all previous progress.
- Doesn't grow with dimensionality: if you choose a grid spacing δx and try to form a grid of N dimensions, you're forced to have $\sim [(x_2 - x_1)/\delta x]^N$ points. This grows the computational cost exponentially. Random samples have the same convergence criteria no matter the dimensionality.
- Above we alluded to a high dimensional albeit square-like region of integration. What if the boundary is non-trivial (e.g. a circular boundary $\sqrt{x^2 + y^2}$)?. Random samples allow us to tackle this easily (see rejection methods later).

[1]This is based on a real example. I spent about a week of head scratching wondering how to resolve all the peaks of a Bessel function. The problem was choosing a logarithmic grid, which means the oscillations appeared to grow exponentially in frequency. No matter how finely I sampled, the peaks would eventually not be resolved. If I had sampled more randomly this could have been avoided. -TK

3.1 Uncertainties

In the limit of large N the central limit theorem tells us that the error on our average Eq. 23 will follow

$$\sigma_I = \frac{V}{\sqrt{N}}\sigma_f, \tag{25}$$

where V is the volume of the integration region ($V = x_2 - x_1$ in our one-dimensional example) and σ_f is the standard deviation of the integrand,

$$\sigma_f^2 = \Big\langle f^2(x) \Big\rangle - \Big\langle f(x) \Big\rangle^2 \tag{26}$$

$$= \frac{1}{N}\sum_{i=1}^{N} f^2(x_i) - \left(\frac{1}{N}\sum_{i=1}^{N} f(x_i)\right)^2. \tag{27}$$

Since the uncertainty reduces as $1/\sqrt{N}$, this leads to an interesting case of "diminishing returns" – very few points will give a disproportionately good estimate compared to the more expensive calculations but, for large enough N, the gains of running longer get progressively less. It is noteworthy that the errors in the integration methods discussed above can be much smaller for comparable N, but with some caveats we will discuss below.

3.2 Importance sampling

There is one powerful way of improving the convergence of Monte Carlo methods known as importance sampling. The idea here is to distribute our random samples in a way that best captures the features of the function in question. In practice this looks like a change of variable in the integral.

Consider that the error in a Monte Carlo integral follows

$$\sigma_f^2 = \frac{1}{V}\int [f(x)]^2 \mathrm{d}x - \left[\frac{1}{V}\int f(x)\mathrm{d}x\right]^2, \tag{28}$$

which is sensitive to how non-uniform the data is (e.g. $\sigma_f = 0$ for a flat function, $f =$ constant). The trick then is to find a change of variables that makes the resulting integral as *flat* as possible:

$$I = \int_{x_1}^{x_2} f(x)\mathrm{d}x = \int_{x_1}^{x_2} \frac{f(x)}{\rho(x)}\rho(x)\mathrm{d}x = \int_{y_2=y(x_2)}^{y_1=y(x_1)} \frac{f(x(y))}{\rho(x(y))}\mathrm{d}y, \tag{29}$$

where $\rho(x)\mathrm{d}x = \mathrm{d}y$. It might be clear that finding such a substitution is a bit more of an art than an exact science, but we will illustrate an example here. Firstly we can list some necessary properties of $\rho(x)$:

- $\rho(x)$ must be non-negative everywhere, and only 0 where $f(x)$ is also 0 (in which case we also demand $\rho(x)$ approach the zero slower, as to give a finite value via L'Hôpital's rule).
- $y(x)$ must be known in closed analytic form. It must also be invertible to produce $x(y)$, unless you can otherwise produce a sample of points according to $y(x)$.
- If $f(x)/\rho(x)$ is less flat than $f(x)$, this was a useless exercise.