Monte Carlo Integration

(When you don't know what's going on, there is always statistics...)

The average value of a function f over the interval $[a, b]$, f_{ave}, is defined by

$$(b-a)f_{ave} = \int_a^b f(x)dx \qquad \text{(width * height = area)}.$$

If f is a fcn of 2 variables the average value of f over Ω is defined by

$$\text{area}(\Omega)f_{ave} = \int_\Omega f(x,y)\ dA \qquad \text{(area * height = volume)}.$$

For higher dimensions, we have

$$\text{volume}(\Omega)f_{ave} = \int_\Omega f(X)\ d\Omega \qquad (d\text{-volume * height} = (d{+}1)\text{-volume}),$$

in each of these, "height" is f_{ave}, and we interpret "volume" in the general sense (and this *may not* be easy to find). If Ω is a k-dimensional "rectangular" region (a "box" in $\mathbb{R}^k$, or a "k-box"), then volume(Ω) is easy to find: length*width*height*... :

$$\text{volume}(\Omega{=}\text{box}) = \int_\Omega 1\ d\Omega = \int_{a_k}^{b_k} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} dx_1 dx_2 \cdots dx_k = \prod_{j=1}^{k} (b_j - a_j).$$

So what? Well, we can turn these definitions around to get expressions for the definite integral, and lacking the true value f_{ave}, we can sample from the domain to get an approximate value for f_{ave} and therefore an approximate value for the integral. If we sample the domain by choosing x randomly, then the technique is called Monte Carlo (MC) integration, and statistical techniques can tell us our approximate error.

If we sample randomly from a uniform distribution on Ω ($x_j \in \Omega$ is just as likely to be selected as any other $x \in \Omega$), and then compute the sample average

$$\hat{f} = \frac{1}{N} \sum_{i=1}^{N} f(x_i),$$

then we have the integral approximation

$$\int_\Omega f\ d\Omega \approx \text{volume}(\Omega)\ \hat{f}.$$

How good is our approximation? Since the variance of the sum of N identically distributed independent r.v.'s is a factor of $1/N$ smaller than the variance of just one, and since the central limit theorem says that as $N \to \infty$, we can interpret this variance as that of a Gaussian distribution, we can therefore consider the error in an MC integration with N samples to behave (statistically) like $O(1/\sqrt{N})$.

Notice that N is the number of function evaluations, and comparing this error estimate to other quadrature rules we have developed, we see that MC integration is very inefficient for low dimensional quadratures. On the other hand, standard grid-type quadrature (e.g. composite Simpson) of functions of more than about $d = 8$ variables is usually more costly than MC techniques. For example, if $d = 20$ (considered very small in MC applications), then 20-dim'l composite Simpson with only 5 grid points per axis would require 5^{20} evaluations of f (a fcn of 20 variables). If we could evaluate f 1000 times per second, then this standard method would take about 3 thousand years. Taking $N = 10^6$, MC might get about 3 digits of accuracy in less than 20 minutes.