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 $[a, b]$ is defined by

$$(b - a)f_{ave} = \int_a^b f(x)dx$$

(width * height = area). If f is a fcn of 2 variables the average value of f over Ω is

$$\text{area}(\Omega)f_{ave} = \int_\Omega f(x, y)dA$$

(area * height = volume). For higher dimensions, we have

$$\text{volume}(\Omega)f_{ave} = \int_\Omega f(X)dV,$$

where we interpret "volume" in the general sense (and this *may not* be easy to compute). If Ω is a "rectangular" region, then volume(Ω) is easy to compute, and we write:

$$\left[\prod_{j=1}^{k}(b_j - a_j)\right] f_{ave} = \int_{a_k}^{b_k} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2, \ldots, x_k)dx_1dx_2 \cdots dx_k.$$

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 integration.

If we sample from the domain uniformly ($x_j \in \Omega$ is just as likely to be selected as any other $x_k \in \Omega$), and then compute the sample average

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

then we have the integral approximation

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

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 a Monte Carlo simulation 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 Monte Carlo integration is extr*eeee*mly inefficient for low dimensional quadratures. This technique is an important tool in scientific computation, but is only used for very high dimensional problems.