It is common to find that one needs to compute integrals in various fields of science and engineering, and in this note we will give an overview of their numerical computation.

At high school level, you may have performed such integration using some simple/straightforward schemes like Simpson's Rule; lets recapitulate the underling idea here. Consider a integral $I = \int_a^b \; dx \; f(x)$, with the range of integration (interval) given by $x \in [a, b]$ and integrand $f(x)$. One can split the interval into N equal sub-intervals, and the total integral I is sum of integrals over sub-intervals I_i as follows:

$$I = \int_a^b \; dx \; f(x) = \sum_{i=1}^{N} I_i, \text{ with } I_i = \int_{a+(i-1)h}^{a+ih} \; dx \; f(x), \; i \in [1, N] \tag{1}$$

where $h = (b-a)/N$ is the width of each interval and $i-$th interval has $x_i \in [a + (i-1)h, a + ih]$. Various simple schemes have now approximations for I_i:

- Left Integral: $I_i = h \; f(x_{i,L})$ with $x_{i,L} = a + (i-1)h$, i.e. approximate as rectangle with base h and height $f(x_{i,L})$
- Right Integral: $I_i = h \; f(x_{i,R})$, with $x_{i,R} = a + ih$
- Trapezoidal Rule: $I_i = h \; \frac{1}{2} \left[f(x_{i,L} + f(x_{i,R}) \right]$
- Simpson's Rule (fit function to a parabola within the interval):
  $I_i = \frac{1}{6} \; h \left[f(x_{i,L} + f(x_{i,R}) + 4f(\frac{x_{i,L}+x_{i,R}}{2}) \right]$

Above idea (of using equally spaced grid for segmenting the interval into equally sized sub-intervals) can be extended to 2-dimensions, with $I = \sum_i \sum_j I_{i,j}$, and $I_{i,j} = \int_{A_{i,j}} dx \; dy \; f(x, y)$, with the area of integration $A_{i,j}$ being a rectangular patch of widths of h_x and h_y along $x-$ and $y-$ axis respectively. Thus $A_{i,j} = [x_{i,L} + (i-1)h_x, \; x_{i,L} + ih_x] \bigotimes [y_{j,L} + (j-1)h_y, \; y_{j,L} + jh_y]$. Similar to 1-dimensional integrals, approximation schemes for $A_{i,j}$ can be formulated. Note that the number of sub-intervals is now N^2, if there are N sub-intervals in x and y directions.

Another set of important scheme is the **Adaptive** schemes. Integral over the whole interval can be split into sum of integral over smaller sub-intervals; however it is more efficient to have sub-interval size dependent (adapt) to behaviour of the function in that sub-interval: large sub-intervals where function 'changes' less and smaller sub-intervals where function 'changes' more. For example, the integration schemes of left integral or right integral depends on how fast the f changes; the error in the value of the integral will be given by $h * \frac{df}{dx}(x^*)$. Hence for these evaluations, the interval size has to be inversely related to $f'(x)$. In Simpson's rule (or Trapeziodal rule) which already incorporate the changes of f in the interval, the change is measured by value of $|\frac{d^2 f}{dx^2}|$, and hence large sub-interval size for small $|f''|$ and small sub-intevals for large $|f''|$.

In general for d-dimensional integrals, simplistic schemes outlined above require you to evaluate the function f a total of N^d times, i.e. exponential in the number of intervals in a dimension. This scaling makes the computation of higher dimensional integrals ($d >= 5$) infeasible using such

simplistic schemes, requiring fundamentally new schemes for their evaluation. We will discuss Monte Carlo methods in the remainder of this note.

Monte Carlo methods use ideas from probability theory to generate samples, and calculate approximations to integrals from these samples. Simply put:

$$I = \int_a^b dx\ f(x) = \int_a^b \rho(x) g(x) \text{ with } g(x) = f(x)/\rho(x), \tag{2}$$

and requiring that $\rho(x) \neq 0\ \forall\ x \in [a,b]$. Any function with $\rho(x) \geq 0\ \forall\ x \in [a,b]$ and $\int_a^b dx\ \rho(x) = 1$ can be interpreted as a probability density function and hence

$$I = \int_a^b dx\ f(x) = \int_a^b dx\ \rho(x)\ \frac{f(x)}{\rho(x)} \implies I = \mathbb{E}_{x\sim\rho(x)}\ f(x)/\rho(x) \tag{3}$$

That is, the integral is formally equal to the average value of f/ρ for the samples generated from the probability density function ρ! 'Monte Carlo' refers to the casino city, where random sampling is inherent in all gambling games that city is famous for!

In statistical mechanics, it is quite common to evaluate high dimensional integrals. For equilibrium systems, the probability density is given by Boltzmann distribution $p(\vec{R}) \propto \exp\left(-\frac{E(\vec{R})}{kT}\right)$ where $\vec{R} \equiv (\vec{r}_1, \vec{r}_2, \cdots, \vec{R}_N) \in \mathbb{R}^{3N}$ and thus, any property $\langle A \rangle$ can be calculated as

$$\langle A \rangle = \mathbb{E}_{\vec{R}\sim p(\vec{R})}\ A(\vec{R}) = \int\ d\vec{R}\ p(\vec{R})\ A(\vec{R}) = \frac{\int d\vec{R}\ A(\vec{R})\ \exp(-\beta E(\vec{R}))}{\int d\vec{R}\ \exp(-\beta E(\vec{R}))} \tag{4}$$

where $\beta = (kT)^{-1}$. Since $E(\vec{R})$ is highly non-linear and has very large number of minima, $p(\vec{R}) \propto \exp\left(-\frac{E(\vec{R})}{kT}\right)$ is highly multi-modal, and it is near impossible to generate samples from it directly (or brute-force).

When generating independent samples from a distribution is not feasible, it is now common to use Markov Process to generate the samples from $p(\vec{R})$ in the following fashion: start with a configuration $\vec{R}_i$, and then generate a new configuration $\vec{R}_j$ by adding a random vector $\delta\vec{R}$, that is, $\vec{R}_j = \vec{R}_i + \delta\vec{R}$; add $\vec{R}_j$ to chain of configurations with probability $p_{acc} = \min\left(1, \exp\left(-\beta\left[E(\vec{R}_j) - E(\vec{R}_i)\right]\right)\right)$ or $\vec{R}_i$ with probability of $1 - p_{acc}$. This adds a member to Markov Chain, and repeating the step creates a chain of configurations that is theoretically proven to sample $p(\vec{R})$; this scheme was proposed by N. Metropolis and hence is called as **Metropolis Markov Chain Monte Carlo** method. Note that in this Markov Process, two consecutive configurations differ by a random change $\delta\vec{R}$ which in turn is quite small leading to highly correlated configurations in the sampled Markov Chain; thus, so as to get *independent* samples of distribution $p(\vec{R})$, it is necessary to only take configurations that are separated by M steps so that such configurations are de-correlated and likely to be independent samples and discarding the intermediate configurations ($M \sim N$, the number of particles in system under consideration). The term Monte Carlo comes in from the fact that generating $\delta\vec{R}$ requires random sampling.

One final topic is **Importance Sampling**. Often times, it may be required to sample according

to $p'(\vec{R})$ instead of $p(\vec{R})$. For any property $A(\vec{R})$, it can be shown:

$$\langle A \rangle = \int d\vec{R}\; p(\vec{R})\; A(\vec{R}) = \int d\vec{R}\; p'(\vec{R}) A'(\vec{R}) \implies \langle A \rangle = \mathbb{E}_{\vec{R} \sim p'} A'(\vec{R}),\; A' = A\; p/p' \qquad (5)$$

Thus, the samples for computing properties of p comes from the important parts of (or generated from) the distribution p'; that is samples are obviously biased by distribution $p\prime$, leading to the name Importance Sampling. While this is not discussed in this note, one of the important reason to perform Importance Sampling is to reduce the variance in computation of properties.

References:

1. **Simpson Rule**: "Advanced Engineering Mathematics" by Kreyzig and Westwig (10th Edition). Section19.5

2. **Adaptive method:** Simple scheme outlined in this note is almost never used: a more appropriate scheme is given in the same section 19.5 (above).