Variance Reduction and Rare event probability estimation using Importance sampling

Monu Kumar Choubey, 2011090, *School of Physical Sciences, NISER.*

1. Introduction

Numerical estimation of expectation values (and more generally, integrals) is a critical calculation and finds very wide applications. Primarily, this is done using Monte Carlo methods. Monte Carlo (MC) is the technique of approximating an expectation by the sample mean of a function of simulated random variables. Consider a random variable X (may be multidimensional), having a probability density function (pdf) $f_X(x)$ such that $f_X(x) > 0 \ \forall \, x \, \epsilon \, A$, where A is the complete set of values for x (sample space). Then the expected value of a function $g(X)$ is given by (Eq^n 1a),

$$\mathrm{E}_f[g(X)] = \sum_{x \epsilon A} g(x) f_X(x)$$

For continuous distributions, we have (Eq^n 1b),

$$\mathrm{E}_f[g(X)] = \int_{x \epsilon A} g(x) f_X(x) dx$$

The Monte Carlo estimator, taking n samples from X $(x_1, \ldots, x_n)$, is given by,

$$\bar{\mathrm{g}}_n(x) = \frac{1}{n} \sum_{i=1}^{n} g(x_i)$$

Alternatively (Eq^n 2),

$$\bar{\mathrm{g}}_n(X) = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$$

If the integral exists, according to the law of large numbers (Central limit theorem), given $\epsilon > 0$,

$$\lim_{n \to \infty} P\left(\left| \bar{\mathrm{g}}_n(X) - \mathrm{E}_f\big(g(X)\big) \right| \geq \epsilon \right) = 0$$

Also, the estimator $\bar{\mathrm{g}}_n(X)$ is unbiased for $\mathrm{E}_f\big(g(X)\big)$:

$$\mathrm{E}_f[\bar{\mathrm{g}}_n(X)] = \mathrm{E}_f\left[\frac{1}{n} \sum_{i=1}^{n} g(X_i) \right] = \frac{1}{n} \sum_{i=1}^{n} E_f[g(X_i)] = \mathrm{E}_f[g(X)]$$

The variance is given by (Eq^n 3a),

$$Var\big(\bar{\mathrm{g}}_n(X)\big) = Var\left(\frac{1}{n} \sum_{i=1}^{n} g(X_i) \right) = \frac{Var(g(X))}{n}$$
$$= \frac{1}{n} \sum_{x \epsilon A} \left[g(X) - \mathrm{E}_f[g(X)] \right]^2 f_X(x)$$

If the random variable is continuous (Eq^n 3b),

$$Var\big(\bar{\mathrm{g}}_n(X)\big) = \frac{1}{n} \int_{x \epsilon A} \left[g(X) - \mathrm{E}_f[g(X)] \right]^2 f_X(x) dx$$

2. Importance Sampling

Numerical estimation using MC has the following challenges:

- pdf of the random variable is very complicated, difficult to sample from; pdf may not be normalized.
- inefficient to sample from the pdf, large variance/slow convergence
- solving the integral is impossible/not viable

Thus, a proxy distribution (called a proposal distribution/sampling distribution) is used, instead of the original pdf, to simulate the random variable. This is the basic principle of importance sampling (IS) and utilizes the mathematical foundations of the MC methods. Consider a sampling distribution $h(x)$ (contained in A) for the random variable X. Then, we have (Eq^n 4),

$$\mathrm{E}_f[g(X)] = \int_{x \epsilon A} g(x) f_X(x) dx = \int_{x \epsilon A} g(x) \left(\frac{f_X(x)}{h_X(x)} \right) h_X(x) dx$$
$$= \mathrm{E}_h\left[g(x) \left(\frac{f_X(x)}{h_X(x)} \right) \right]$$

Eq^n 4 is valid iff $h(x) \neq 0 \ \forall \, x \, \epsilon \, A \ s.t. \, f(x) \neq 0$. Thus, the estimator is given by (Eq^n 5),

$$\bar{\mathrm{g}}_n^h(X) = \frac{1}{n} \sum_{i=1}^{n} g(X_i) \frac{f(X_i)}{h(X_i)}$$

<u>Choice of sampling distribution</u>:

From Cauchy-Schwartz inequality (and Eq^n 3b), it follows that $Var\left(\bar{\mathrm{g}}_n^h(X)\right)$ is minimized[1] when $h(x) \propto |f(x)|$. Since $f(x) \geq 0 \ \forall \, x \, \epsilon \, A$ (pdf), we need (Eq^n 6),

$$\frac{f(x) g(x)}{h(x)} = \alpha \ \forall \, \{x : h(x) > 0\}$$

Also, for being a good sampling distribution, it must be easy to simulate values from $h(x)$ and compute $h(x) \ \forall \, x \, \epsilon \, A$. For higher dimensional random variables, this is often non-trivial and difficult.

2.1 Advantages

- Estimate remains unbiased:
$$E_h[\bar{g}_n^h(X)] = E_f[g(X)]$$
- Possibly imporved variance:
$$Var\left(\bar{g}_n^h(X)\right) = \frac{1}{n} Var\left(\frac{g(X)f(X)}{h(X)}\right)$$
- Faster convergence, increased efficiency.

3. Illustrations

3.1 Example of 'Importance'

Consider the problem of estimation of the following integral:

$$I = \int_0^{10} e^{-2|x-5|} dx$$

The function is as shown in Fig 1.

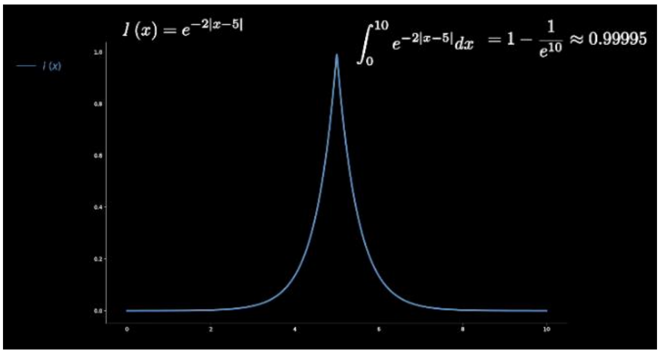

Figure 1: Example of Importance – Integrand

Let's begin with a uniform distribution, given by $p(x)$. Then, $p(x) = 1/10$. The plots for this estimate for the integral and the corresponding variance against the number of samples are as shown in Figs 2,3 respectively.

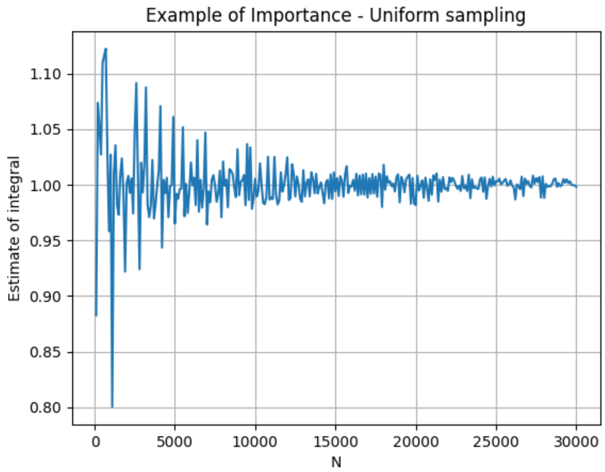

Figure 2: Example of Importance - Estimate (uniform)

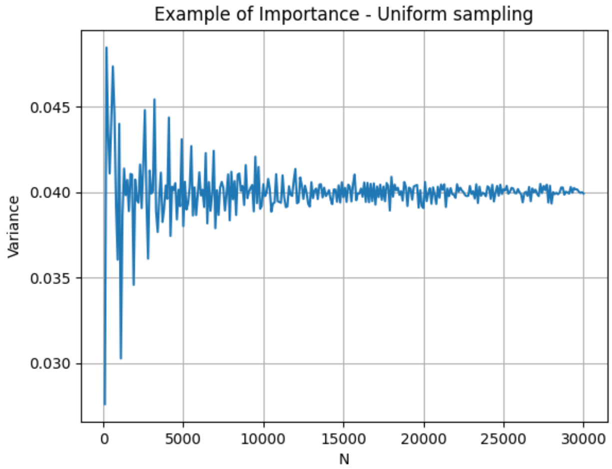

Figure 3: Example of Importance - Variance (uniform)

Now consider a normal distribution as the sampling distribution, centered at $x = 5$, given by $q(x)$. This is chosen because the region around $x = 5$ is the region which contributes the function's value (over the whole range) the most, as illustrated in Fig 4. This sampling has been done using the truncnorm() method from the scipy's stats library in Python[I].

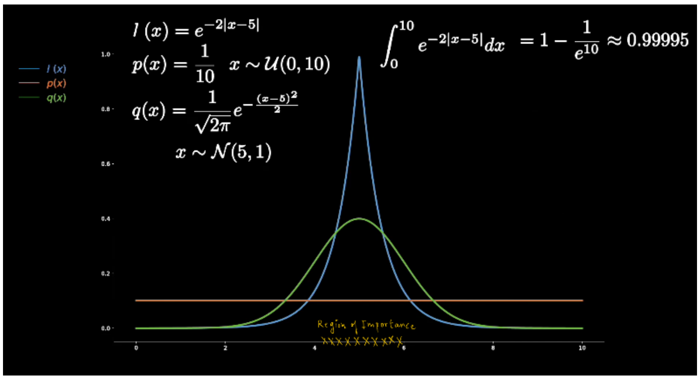

Figure 4: Illustration – Region of Importance

The appropriate plots are as shown in Figs 5 & 6 respectively.

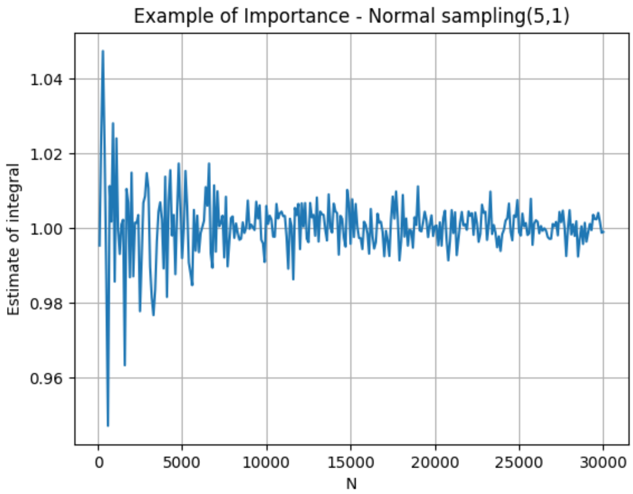

Figure 5: Example of Importance - Estimate (normal)

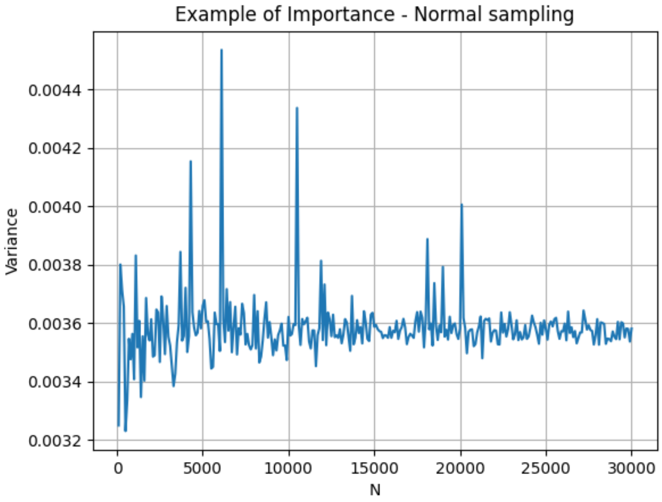

Figure 6: Example of Importance - Variance (normal)

Clearly, there is a significant improvement in terms of numbers of samples needed for convergence as well as the variance. Thus, we have a more efficient estimator using IS.

3.2 Example of 'Expectation'

Consider the quantum mechanical particle obeying 1-D SHO, confined in a box given by $x \, \epsilon \, (-4,4)$. The ground state wave-function in position space is given by,

$$\psi_0(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{4}} e^{-\frac{m\omega x^2}{2\hbar}}$$

The expectation value of position is given by,

$$\langle x \rangle = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{2}} \int_{-4}^{4} x e^{-\frac{m\omega x^2}{\hbar}} dx$$

W.L.O.G, consider $\frac{m\omega}{\pi\hbar} = \pi$.

The integrand is odd and thus, analytical solution is fairly easy. We expect the expectation value to be zero.

The estimates using MC and IS are shown in Figs 7,8,9 and 10. The sampling distribution is taken to be a normal

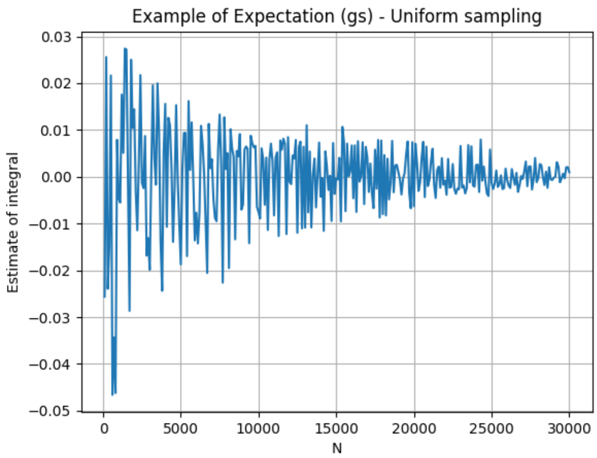

Figure 7: Expectation value of position (gs) - (uniform)

distribution centered at the origin with a std. deviation of 1.

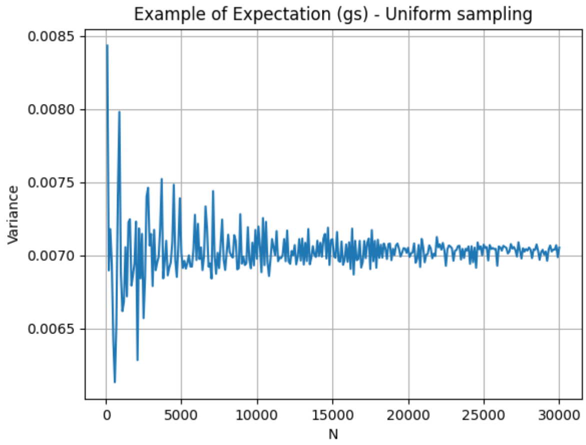

Figure 8: Variance (gs) - (uniform)

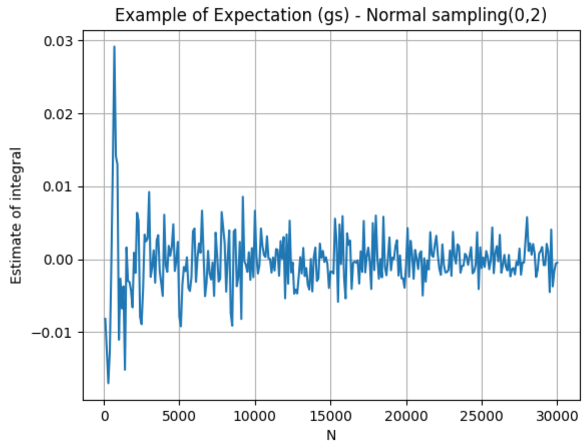

Figure 9: Expectation value of value of position (gs) - normal (0,1)

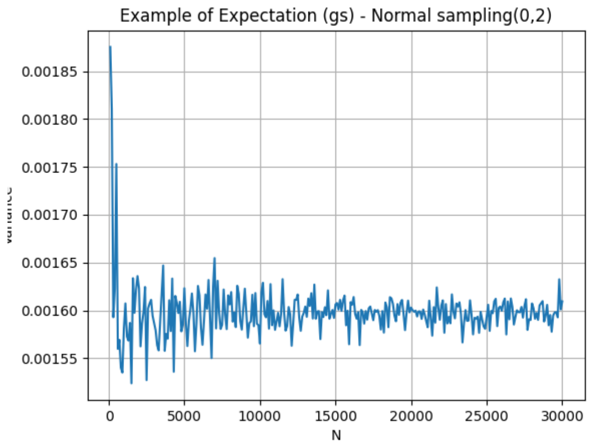

Figure 10: Variance (gs) - normal (0,1)

This is an example of the sampling distribution function being in the same family of the main function ($I(x)$). This condition leads to significant agreement with Eq^n 6, and thus, is a preferred strategy for choosing proxy distributions for IS. This is called the Maximum principle.

The use of IS becomes more advantageous as we try to compute the expectation values for higher excited states, for which, analytical solutions are not as simple. Estimates and corresponding variances for the 1^{st} excited states are as shown in Figs 11,12.

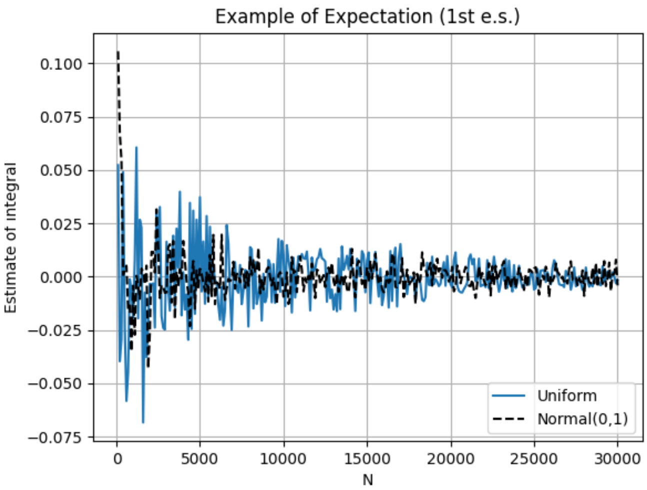

Figure 11: Expectation value of position (1^{st} e.s.) - Estimate

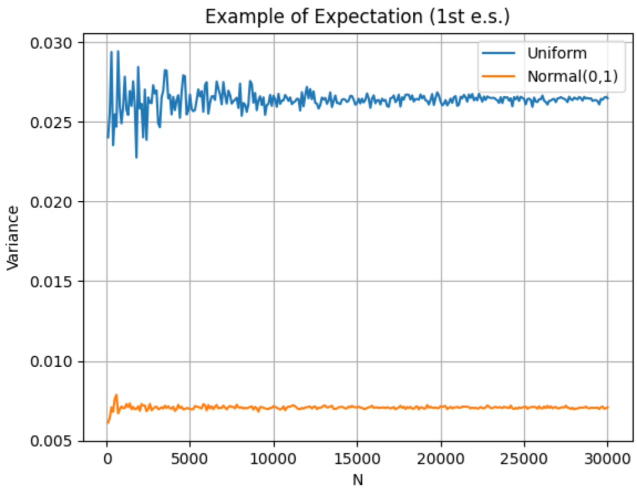

Figure 12: Expectation value of position (1^{st} e.s.) - Variance

3.3 Example of ‘Rare event estimation’

The relatively slow convergence of Monte Carlo methods is significantly more important for calculation of probabilities for events occurring rarely. These kinds of calculations find applications in modelling and estimation (some fields – quantum mechanics, statistical mechanics), data handling (and machine learning in general), reliability testing, finance etc. Thus, consider a general quantity with functional dependence on parameter x (can be generalized to higher dimensions). Let it be $I(x)$, such that the desired value of the quantity is attained in extremely low probability zones of the distribution function for the parameter, i.e., it’s highly unlikely.

As an example, consider the following,

$$I(x) = x^4 e^{\frac{x^2}{4}}$$

The task is to find the probability of the quantity’s value being greater than 1 (Pr $(I(X) \geq 1)$). To illustrate that the normal distribution is not always helpful, consider the original distribution (pdf) for the variable to be a gaussian centered at the origin, given by $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Now, the proxy distribution function is determined using the Maximum principle (may not always work), which turns to be a normal distribution centered at $\sqrt{8}$. Let it be $q(x)$. The functions are as illustrated in Fig 13.

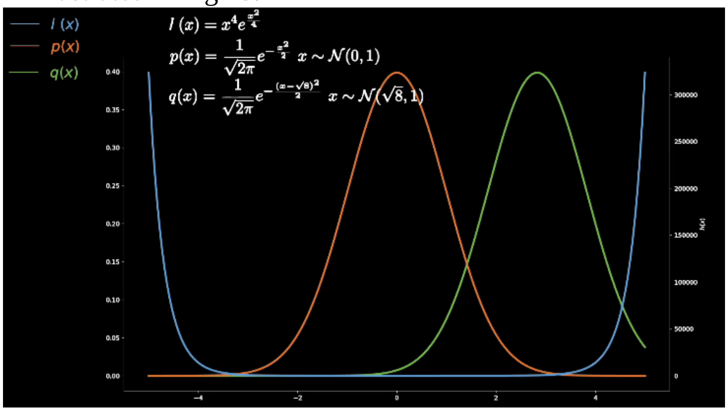

Figure 13: Example of Rarity - Illustration of pdfs

Probability estimates for both distributions are as shown in Fig 14.

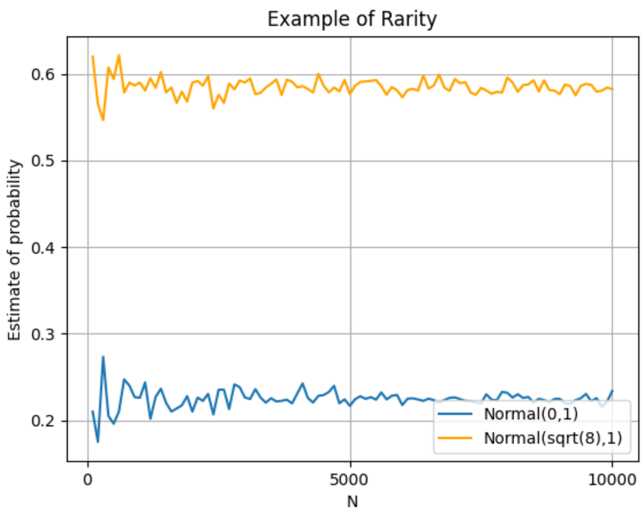

Figure 14: Example of Rarity - effect of sample distribution

Thus, even though a normal distribution is often good for sampling, here it’s a bad indicator of probability since the region of importance (in which $I(x)$ attains high values) lies in the low probability zone of $p(x)$.

Therefore, even though the sample space is same ($x \epsilon (-2,2)$), the probability estimates for the quantity's value are different.

4. Discussions & Conclusions

We've seen various cases of applicability of IS to realize faster, more efficient numerical estimation. However, appropriate choice of sampling distribution function is crucial.

The choice of a proxy function is very essential and depends on the problem. A bad choice may lead to worse variance and slower convergence. Though there is no absolute technique for finding the best proxy, conditional techniques like the Maximum principle are helpful. Such techniques are needed while dealing with higher dimensional and correlated random variables.

5. Notes

[I]The sampling is done using stats.truncnorm() from scipy to obtain better samples (than created using self-made function, included in support library for completeness) and to maintain the applicability of proposed code for generalized problems(higher dimensions).

All the relevant codes and the support library (of defined functions) used may be found here: GitHub

6. References

- [1]Rubinstein, B. Y., 1981 *Simulation and the Monte Carlo Method*. New York: Wiley & Sons, p 123.
- Hammersley, J. M. and D. C. Handscomb, 1964 Monte Carlo Methods. London: Methuen & Co Ltd.
- E. Anderson, "Monte Carlo Methods and Importance Sampling", https://ib.berkeley.edu/labs/slatkin/eriq/classes/guest_lect/mc_lecture_notes.pdf
- K. P. Murphy. Machine Learning: A Probabilistic Perspective, MIT Press, 2012