Monte Carlo Methods and Simulations in Nuclear Technology

General variance reduction techniques

Jan Dufek

2021

KTH Royal Institute of Technology

Contents

Figure-of-merit (FOM)

Figure-of-merit (FOM)

What is the aim of the so-called "variance reduction methods"?

The aim is to improve the efficiency of Monte Carlo simulations–to adress the large computing cost of MC simulations.

How can the efficiency be measured?

It can be measured by the figure of merit (FOM) value

$$\mathrm{FOM} = \frac{1}{\sigma_{m_Y}^2 t},$$

where t is the total computational time or another quantity proportional to the computing time. The $\sigma_{m_Y}^2$ may be exchanged for a square of the real error if it is known.

Figure-of-merit (FOM)

Is the FOM changing during a simple MC simulation, as more samples are being collected?

No, since

$$\mathrm{FOM} = \frac{1}{\sigma^2_{m_Y} t} = \frac{1}{\frac{\sigma^2_Y}{n} t} = const.$$

The FOM is constant since σ^2_Y is, in principle, not growing/decaying with the number of samples n, and the number of collected samples n is proportional to the computing time t.

So, while the variance of the mean value improves while collecting more samples, the computing time increeases as well, and the efficiency of the simulation remains the same.

Can var. red. methods always improve the efficiency of MC simulations?

In principle, all variance reduction methods may improve FOM in exchange for some knowledge about the system. If the knowledge is not sufficient, the FOM may be worsened by the methods.

Moreover, while variance reduction methods may reduce $\sigma^2_{m_Y}$ they usually come at an additional computational cost too. This may lead to worsening (decreasing) the FOM.

Control variate

Control variate

When is it possible to apply the control variate technique?

The control variate technique can be used when the numerical model $Y = g(X)$ with an unknown expectation value $\mathrm{E}[Y]$ may be approximated by a simpler model (*control variate*),

$$Z = g^*(X)$$

with a known expectation value $\mathrm{E}[Z]$.

Control variate

What is the principle of the control variate technique?

The mean value of Y is computed as

$$m_Y = m_{(Y-Z)} + \mathrm{E}[Z]$$

by sampling the **difference between the random variables** $g(x)$ **and** $g^*(x)$ using the same random numbers x, and computing the mean value of the difference $m_{(Y-Z)}$.

Control variate

When will the control variate technique increase FOM?

The FOM will be increased when the control variate Z is strongly positively correlated to Y. We know that the variance of m_Y is equal to the variance of $m_{(Y-Z)}$,

$$\begin{aligned}
\mathrm{Var}[m_Y] =& \mathrm{Var}[m_{(Y-Z)}] \\
=& \frac{\mathrm{Var}[Y-Z]}{n} \\
=& \frac{\mathrm{E}[(Y-Z-\mathrm{E}(Y-Z))^2]}{n} \\
=& \frac{\mathrm{E}[(Y-\mathrm{E}[Y]-(Z-\mathrm{E}[Z]))^2]}{n} \\
=& \frac{\mathrm{E}[(Y-\mathrm{E}[Y])^2-2(Y-\mathrm{E}[Y])(Z-\mathrm{E}[Z])+(Z-\mathrm{E}[Z])^2]}{n} \\
=& \frac{\mathrm{Var}[Y]+\mathrm{Var}[Z]-2\mathrm{Cov}[Y,Z]}{n}.
\end{aligned}$$

so, $\mathrm{Var}[m_Y]$ is improved when $2\mathrm{Cov}[Y,Z] > \mathrm{Var}[Z]$. Evaluating the function $g^*(X)$ must, however, not slow down the calculation, otherwise, the FOM may not increase compared to FOM of simple sampling.

Correlated sampling

Correlated sampling

For what problems can we use the correlated sampling?

Correlated sampling can be used when one has to calculate **the difference** between two very similar models g_1 and g_2,

$$Y_1 = g_1(X),$$
$$Y_2 = g_2(X).$$

Commonly, g_2 is a slightly modified model g_1. Since the difference

$$\mathrm{E}[Y_1] - \mathrm{E}[Y_2]$$

may be very small, the simple sampling would need to calculate $\mathrm{E}[Y_1]$ and $\mathrm{E}[Y_2]$ to a very good accuracy, which would require a large computational time.

Correlated sampling

What is the principle of the correlated sampling?

The method can efficiently resolve this problem by studying the random variable

$$\begin{aligned} Z &= Y_1 - Y_2 \\ &= (g_1 - g_2)(X). \end{aligned}$$

Variance of Z is

$$\mathrm{Var}[Z] = \mathrm{Var}[Y_1] + \mathrm{Var}[Y_2] - 2\mathrm{Cov}[Y_1, Y_2].$$

Thus, if Y_1 and Y_2 are strongly positively correlated to each other, which is likely when g_1 and g_2 are very similar, then simple sampling of Z is more efficient than sampling Y_1 and Y_2 separately using independent input random variables X_1 and X_2.

Stratified sampling

Stratified sampling

What is the principle of the stratified sampling?

- X is divided into L strata (value groups, X_i, $i = 1, \ldots, L$), and Y is sampled within these groups (which also divides Y into strata).
- Strata may be designed so that the variance of Y_i within the strata are small (some knowledge is needed here).
- Dividing X into strata X_i causes Y be also divided into corresponding strata Y_i; i.e., samples x from X_i will generate samples $y = g(x)$ from Y_i.
- Since we know exactly the probability that a random number X falls into some group X_i we also know that with the same probability ω_i samples of Y fall into Y_i.
- It can be shown that

$$\mathrm{E}[Y] = \sum_{i=1}^{L} \omega_i \, \mathrm{E}[Y_i],$$

and

$$\mathrm{Var}[m_Y] = \sum_{i=1}^{L} \omega_i^2 \, \frac{\mathrm{Var}[Y_i]}{n_i},$$

where n_i is the total number of samples x in X_i.

Stratified sampling

Principle - continuation

- Since $\mathrm{E}[Y_i]$ is not known for any i, the expectation value must be estimated by

$$m_Y = \sum_{i=1}^{L} \omega_i \, m_{Y_i}.$$

- If strata are designed well then it is likely that $\mathrm{Var}[m_Y]$ get smaller compared to the value obtained from simple sampling. On the other hand, a poor stratification will result in an increased variance.

Stratified sampling

How can we decide how many samples we generate in each group?

- The variance obtained by stratified sampling is minimised when the samples are distributed in strata according to the *Neyman allocation*

$$n_i = n \frac{\omega_i \sigma_{Y_i}}{\sum_{j=1}^{L} \omega_j \sigma_{Y_j}},$$

  where σ_{Y_i} is the standard deviation of Y_i.

- These deviations σ_{Y_i} are not known at the beginning of the calculation, therefore, the simulation is commonly divided into batches; the strata in the first batch may contain equal number of samples.
- In the following batches, σ_{Y_i} can be estimated for all i, and samples can be distributed more efficiently.
- Nevertheless, there is a risk of estimating σ_{Y_i} incorrectly in the first batch, which could result in a wrong distribution of the following batches!

Importance sampling

Importance sampling

What is the principle of the importance sampling?

- The method samples X from a new pdf f_Z, which is corrected by multiplying the sampled values of Y by a correction factor

$$\frac{f_X(x)}{f_Z(x)}.$$

- The computed random variable Z

$$Z = g(x)\frac{f_X(x)}{f_Z(x)}$$

has the expectation value $\mathrm{E}[Z]$ equal to $\mathrm{E}[Y]$:

$$\mathrm{E}[Z] = \int_x g(x)\frac{f_X(x)}{f_Z(x)} f_Z(x)\mathrm{d}x = \int_x g(x) f_X(x)\mathrm{d}x = \mathrm{E}[Y].$$

What is the ideal pdf f_Z?

- The variance of Z is

$$\begin{aligned}\mathrm{Var}[Z] &= \mathrm{E}[Z^2] - (\mathrm{E}[Z])^2 \\ &= \int_x g^2(x) \frac{f_X^2(x)}{f_Z(x)} \mathrm{d}x - (\mathrm{E}[Y])^2.\end{aligned}$$

- Thus, if the importance function is chosen such that

$$f_Z(x) = \frac{g(x) f_X(x)}{\mathrm{E}[Y]}$$

  then $\mathrm{Var}[Z] = 0$!

- In this ideal case, only one sample would be sufficient to obtain the precise value of $\mathrm{E}[Y]$ (with a zero variance). Therefore, this method is sometimes called the *zero-variance scheme*.

Importance sampling

We can't know the ideal pdf f_Z. How can we then apply the method?

- Computing the precise importance function $f_Z(x)$ is computationally equivalent to computing the value of $\mathrm{E}[Y]$.
- Therefore, in practice, the importance function must be derived using numerical models much simpler than $g(x)$, so the zero variance is never obtained (although the variance can be sufficiently reduced).
- The correction factors must be computed for each sample, which may be computationally expensive, and it may negatively reflect into FOM.
- There is a risk that this technique will increase the variance if the importance function is wrong!