

Importance sampling and control variates

There exist many problems in science and engineering whose exact solution either does not exist or is difficult to find. For the solutions of those problems, one has to resort to approximate methods. The Variational Monte Carlo (VMC) technique is relatively insensitive to the size of the system, it can be applied to large systems where some other methods are computationally not feasible.

Varvara RUDENKO

17.08.2020

Contents

1	Task	2
2	History and Origin of Monte Carlo Methods	2
3	Monte Carlo Integration Using Importance Sampling - The- ory	2
4	Variance	5
5	Monte Carlo Integration Using Importance Sampling - Prac- tice	6

1 Task

We are estimating $\pi(f) = \int_{\mathbb{R}^d} f(x)\pi(x)dx$ for some measure π with density $\pi(x)$. Using Monte-Carlo approach, we estimate $\pi(f)$ by $\pi_n(f) := \frac{1}{n} \sum_{k=0}^{n-1} f(X_k)$ for i.i.d. random variables X_k distributed according to μ .

If $f(x)\pi(x)$ takes large values at the areas, where $\pi(x)$ is small, it is natural to apply importance sampling to reduce variance of $\pi_n(f)$.

Task:

1. Study the paper Muller et al.
2. Evaluate Importance Sampling + control variates.

2 History and Origin of Monte Carlo Methods

The term ‘Monte Carlo’ has been named after the famous Mediterranean casino town in Monaco. The Monte Carlo (MC) method refers to a class of statistical methods which use random numbers (probabilistic method) similar to those in a casino game of chance, to solve a real and physical (nonprobabilistic) problem. Since the method uses random numbers, it is therefore stochastic in nature and has associated statistical properties. With the advent of modern computers, the random sampling required in most analyses can easily be obtained in a more robust, faster and efficient way. MC methods are widely used in many areas of science and engineering and are particularly useful for solving complicated higher-dimensional integrals and in the study of N-body systems.

3 Monte Carlo Integration Using Importance Sampling - Theory

Take an integral, we want to evaluate it,

$$I = \int_a^b f(x)dx \tag{1}$$

where $f(x)$ is a smooth function defined in the interval $[a,b]$. In the ‘crude’-MC method, the integral is approximated as:

$$I = \int_a^b f(x)dx \approx \frac{b-a}{N} \sum_{i=1}^N f(x_i), \quad (2)$$

where x_i are random numbers, in the interval $[a,b]$. The variance in f is given by:

$$\sigma^2 = \frac{b-a}{N} \sum_{i=1}^N f^2(x_i) - \left[\frac{b-a}{N} \sum_{i=1}^N f(x_i) \right]^2 \quad (3)$$

The integration error can be defined as:

$$\sigma I = \sqrt{\left(\frac{\sigma^2}{N} \right)} \quad (4)$$

There are two ways to reduce the error. First, you can increase the number of tests. Second, reduce the variance. The second option is much more suitable, since it requires less computing time. In the case of ” raw ” integration-MC, it counts the function uniformly, that is, with a uniform distribution. Therefore, if a small area with a large volume of integration makes a significant contribution to the Integral, then there will only be a few samples, which can lead to large statistical errors, even though the number of tests increases. The result of MC integration can be greatly improved with the idea of importance sampling. In this case, the selection points are selected from a distribution that concentrates the points where the integrated function is larger.

Let $g(x)$ be a Probability density function (PDF) defined on $[a,b]$, that has a form close to that of the original function $f(x)$.

$$\frac{f(x)}{g(x)} \approx const \quad (5)$$

Therefore, we can write:

$$I = \int_a^b f(x)dx = \int_a^b \left[\frac{f(x)}{g(x)}\right]g(x)dx \quad (6)$$

The integral can now be calculated with the random numbers generated from the distribution $g(x)$ (weight function) and evaluating $\frac{f(x_i)}{g(x_i)}$ at these points. This is

$$\left\langle \frac{f(x)}{g(x)} \right\rangle = \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{g(x_i)} \quad (7)$$

N - number of steps, x_i - the random numbers distributed as $g(x)$.

Another way is to define $dG(x) = g(x)dx$, where

$$G(x) = \int_a^x g(x)dx \quad (8)$$

You can replace variables, $r = G(x)$, where r is a sequence of random numbers evenly distributed in the interval (0,1). Therefore:

$$I = \int_a^b \left[\frac{f(x)}{g(x)}\right]dG(x) = \int_0^1 \frac{f(G^{-1}(r))}{g(G^{-1}(r))}dr \approx \frac{1}{N} \sum_{i=1}^N \frac{f(G^{-1}(r))}{g(G^{-1}(r))}dr \quad (9)$$

The form $g(x)$ must be chosen so that it minimizes the variance of integral $\frac{f(x)}{g(x)}$, you can choose it so that the integral of $\frac{f(x)}{g(x)}$ is almost flat, and therefore the variance will be formed to a large extent.

The variance is calculated from:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N \tilde{f}(x_i)^2 - \left(\frac{1}{N} \sum_{i=1}^N \tilde{f}(x_i) \right)^2 \quad (10)$$

Then the integration error is equal to:

$$\sigma_1 = \sqrt{\frac{\sigma^2}{N}} \quad (11)$$

4 Variance

$$Var[\pi_n(f)] = \frac{1}{n^2} \sum_{k=0}^{n-1} Var[f(X_k)] = \frac{\sigma_\pi^2(f)}{n}, \quad (12)$$

where $\sigma_\pi^2(f) = Var[f(X_0)] = \pi(f^2) - (\pi(f))^2$. What options to fight variance?

- Increase n . Not an option in many situations.
- Control variates: replace f by $f - g$, where $\pi(g) = 0$. Denote by $G := \pi(g) = 0$. Find

$$\hat{g}_n := \operatorname{argmin} V_n(f - g), \quad (13)$$

where

$$V_n(f - g) = \frac{1}{n-1} \sum_{k=0}^{n-1} (f(X_k) - \pi_n(f) - g(X_k) + \pi_n(g))^2 \quad (14)$$

Let $\pi(x) = e^{-U(x)}$. Take Stein's control variates

$$g_\phi(x) = - \langle \phi(x), \Delta U(x) \rangle + \phi'(x), \quad (15)$$

then by construction we get zero expectation. Let $X = R$. Then easy to believe $E_{g_\phi}(x) = 0$. Idea is to take $\phi(x) = \phi(x, \theta)$ - neural network.

Setup Our setup for experiments:

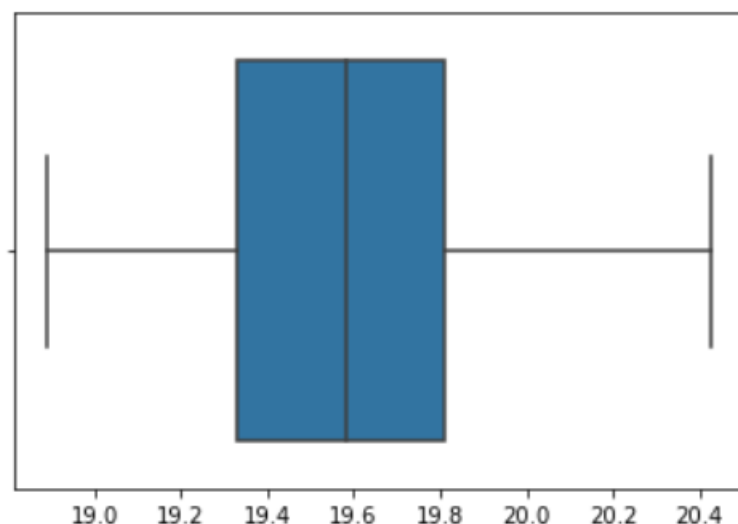
- For one dimension $f(x) = x^3$, various polynomials for $d = 2$; Tested distributions:
- Gaussian Mixtures in dimensions 1, 2;
- Banana shaped distribution;
- Donut distribution:

5 Monte Carlo Integration Using Importance Sampling - Practice

Now, let's illustrate the algorithm with one of the examples. The integral that we will calculate:

$$I = \int_{-\infty}^{\infty} \frac{1}{2 * 0.4 * \sqrt{2\pi}} (x^4 + (x - 2)^2) (e^{-\frac{(x-1)^2}{0.32}} + e^{-\frac{(x+2)^2}{0.32}}) \quad (16)$$

For the truth, we take the value calculated in WolframAlpha = 19.64. Applying a crude MS, we get the result: 19.59 ± 0.25



Let's try to improve the accuracy of measurements using IS. We select the parameters that give the lowest variance.

