Monte Carlo integration and Importance sampling

Yuning Chen

The purpose of this project is to use Monte Carlo to calculate the approximate value of integral, then use importance sampling to increase the precision.
The integral I'm going to calculate is

$$I = \int_a^b \frac{e^{-x}}{1+(x-1)^2}\,dx$$

Because this integral can't be solved analytically, and therefore we need to approximate it numerically. First, I used the basic Monte Carlo method to approximate the integral.

Crude Monte Carlo

This method is very easy to understand, we can use this formula to estimate the integral:

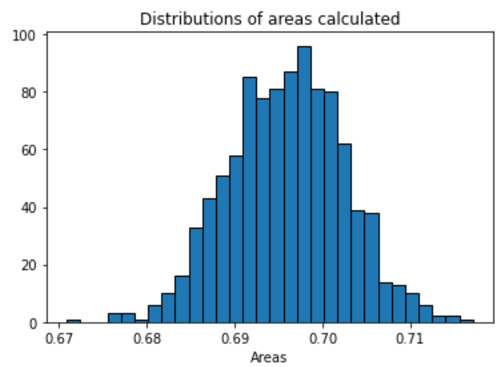
$$f_{ave} = \frac{1}{b-a}\int_a^b f(x)\,dx$$

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

Where f_{ave} denotes the average value of $f(x)$. So we only need to know the f_{ave}, then we can know the value of the integral. The method of Monte Carlo is to generate N random numbers x among [a, b], then calculate the value of f(x). So we can calculate the sample mean and use it to approximately represent f_{ave}.
We need to calculate the variance to observe the accuracy of the estimate.
I used the sample variance represent the variance. Since the N is very big, the corrected sample variance is basically equal the sample variance.

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

Set a=0, b=5, N=20000
We can see that the estimated value is clustered between 0.69 and 0.70, but the estimated distribution is still relatively wide, and the result is not accurate enough.



I will use importance sampling to make the estimation more accurate.

Importance sampling

We can consider sampling from another suitable density function g(x), if we can find a g(x) such that

$$\frac{f(x)}{g(x)} \approx k$$

Basically, we want g(x) to look like a scaled version of f(x).

We'll also need g(x) to satisfy a few criteria:
1. g(x) is integrable
2. g(x) is non-negative on [a,b]
3. The indefinte integral of g(x), which we'll call G(x), has a real inverse
4. The integral of g(x) in the range [a,b] must equal 1
We'll define G(x) as follows, and we'll also perform a change of variables to r.

$$r = G(x) = \int_0^x g(x)\,dx$$

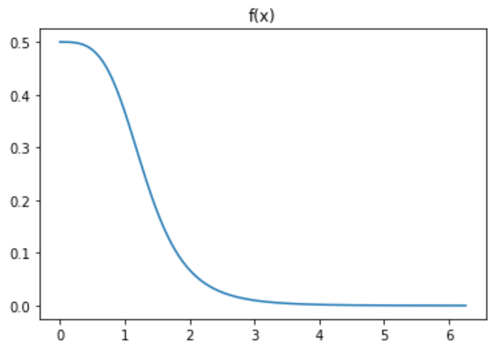
Use these definitions, we can produce the following estimation:

$$I \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(G^{-1}(r_i))}{g(G^{-1}(r_i))}$$

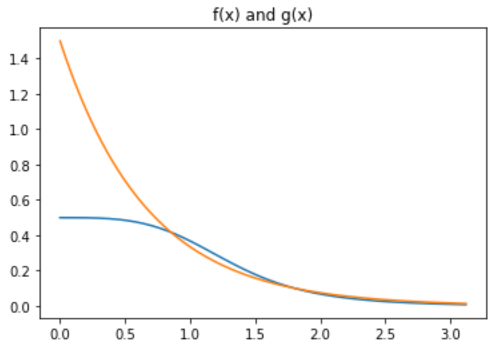
So, we need to find a g(x) is suitable for our integral

$$I = \int_a^b \frac{e^{-x}}{1+(x-1)^2}dx$$

First, I plot the function, and we can see that the function is active between [0 ,3] and inactive between [3, inf]. We need to find a function template that can be parameterized to replicate this quality.



I try to use $g(x) = Ae^{-\lambda x}$ to replicate the quality.



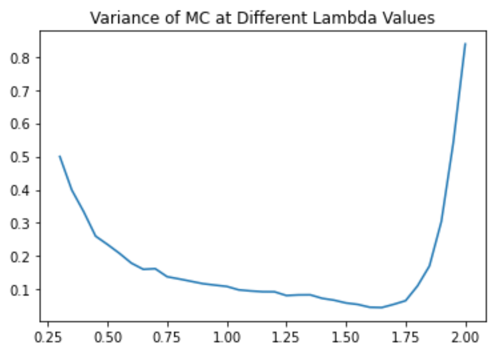
As we can see the graph, g(x) also active between [0, 3], though g(x) does not ideally replicate the shape of f(x), it can still will help us for decreasing the estimation's variance.
Use the normalization condition on g(x), $1 = \int_0^\infty g(x)\,dx$, we can prove that $A = \lambda$.
So the only thing we need to do left is to find the optimal λ that minimize the variance.
I calculate the variance like that:

$$\widehat{\sigma^2} = \left[\frac{1}{N}\sum_{i=1}^{N}\frac{f^2(x_i)}{g^2(x_i)}\right] - \left[\frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{g(x_i)}\right]^2$$

To test the optimal λ, I do like this:

1. Start at λ = *0.05*
2. Calculate the variance
3. Increment λ
4. Repeat steps *2* and *3* until you reach the last λ
5. Pick the λ with the lowest variance — this is your optimal λ
6. Use importance sampling Monte Carlo with this λ to calculate the integral

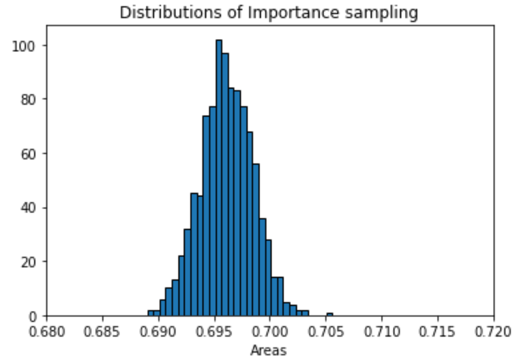
As we can see, when $\lambda = 1.65$, the variance is the lowest. So I use this λ as the optimal λ.



And then I use this optimal λ to calculate the importance sampling Monte Carlo.
After calculating 20,000 of random x, the value of importance sampling is about 0.69543, and the variance is 0.044. While the estimation of basic Monte Carlo is 0.692978, and the variance is 0.8491, which is much bigger than the importance sampling.

And we can see the distribution of the importance sampling Monte Carlo, it's very clustered, as we compared it with the basic Monte Carlo.

```
Importance Sampling Approximation: 0.6954377966042469
Variance: 0.04409868282991081
Error: 0.001484902064614209
```




```
Monte Carlo Approximation of f(x): 0.69297841682135
Variance of Approximation: 0.8497509698052066
Error in Approximation: 0.006518247348042289
```