**Jinsheng Wang (NetID: jwang278)**

**Atomic scale simulation HW4**

**Uniform Random Number Generator**

**1: LCG(16,3,1,2)**

the first 20 steps are:

[ 7. 6. 3. 10. 15. 14. 11. 2. 7. 6. 3. 10. 15. 14. 11. 2. 7. 6. 3. 10.]

After 7 numbers, the random numbers generated started to repeat, and the period is 7,6,3,10,15,14,11,2. Not all the numbers between 0 and 15 are presented.

**2: LCG(2^32,69069,1,0)**

the first 10 values are:

1.00000000e+00 6.90700000e+04 4.75628535e+08 3.27740411e+09 7.72999773e+08 3.87783206e+09 3.82183544e+09 1.66220041e+09 2.04415807e+09 3.78898993e+09

map all the random numbers to range [0,1]

the first 10 values are:

2.32830644e-10 1.60816126e-05 1.10740898e-01 7.63080108e-01 1.79978035e-01 9.02878134e-01 8.89840406e-01 3.87011191e-01 4.75942640e-01 8.82192963e-01

map all the random numbers to range [-0.5,0.5]

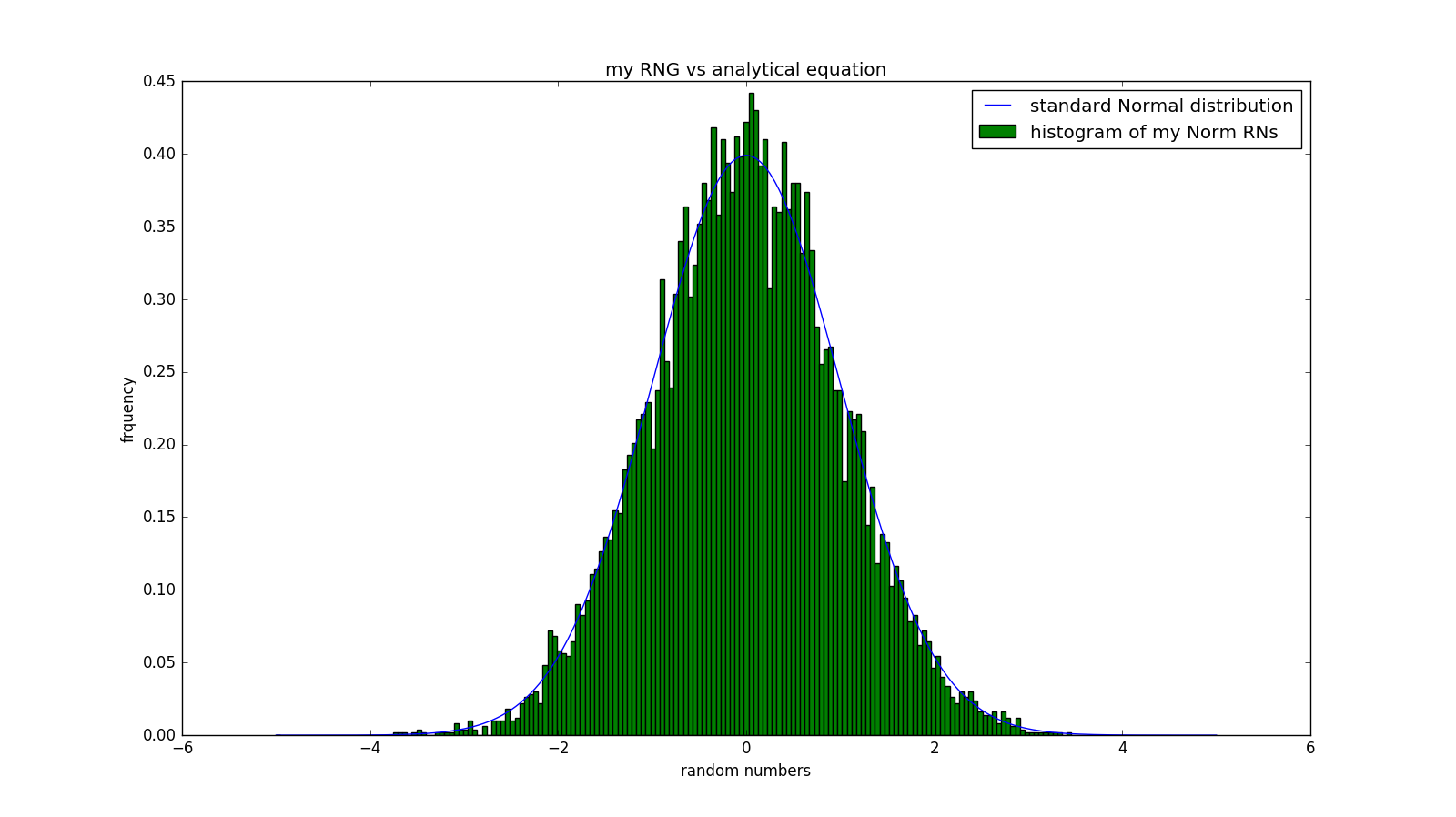
the first 10 values are:

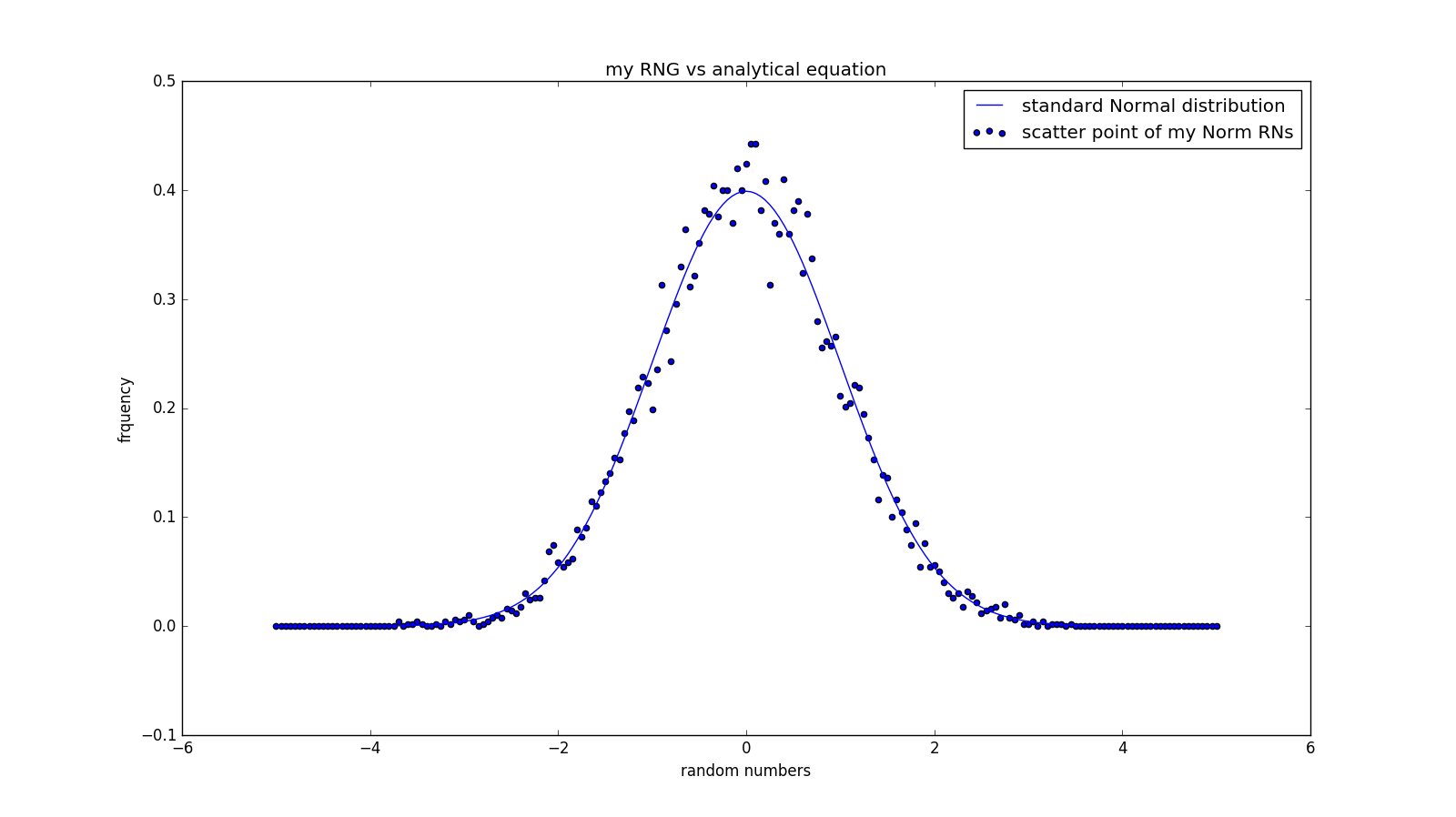
-0.5 -0.49998392 -0.3892591 0.26308011 -0.32002197 0.40287813

0.38984041 -0.11298881 -0.02405736 0.38219296

**Gaussian Random Number Generator**

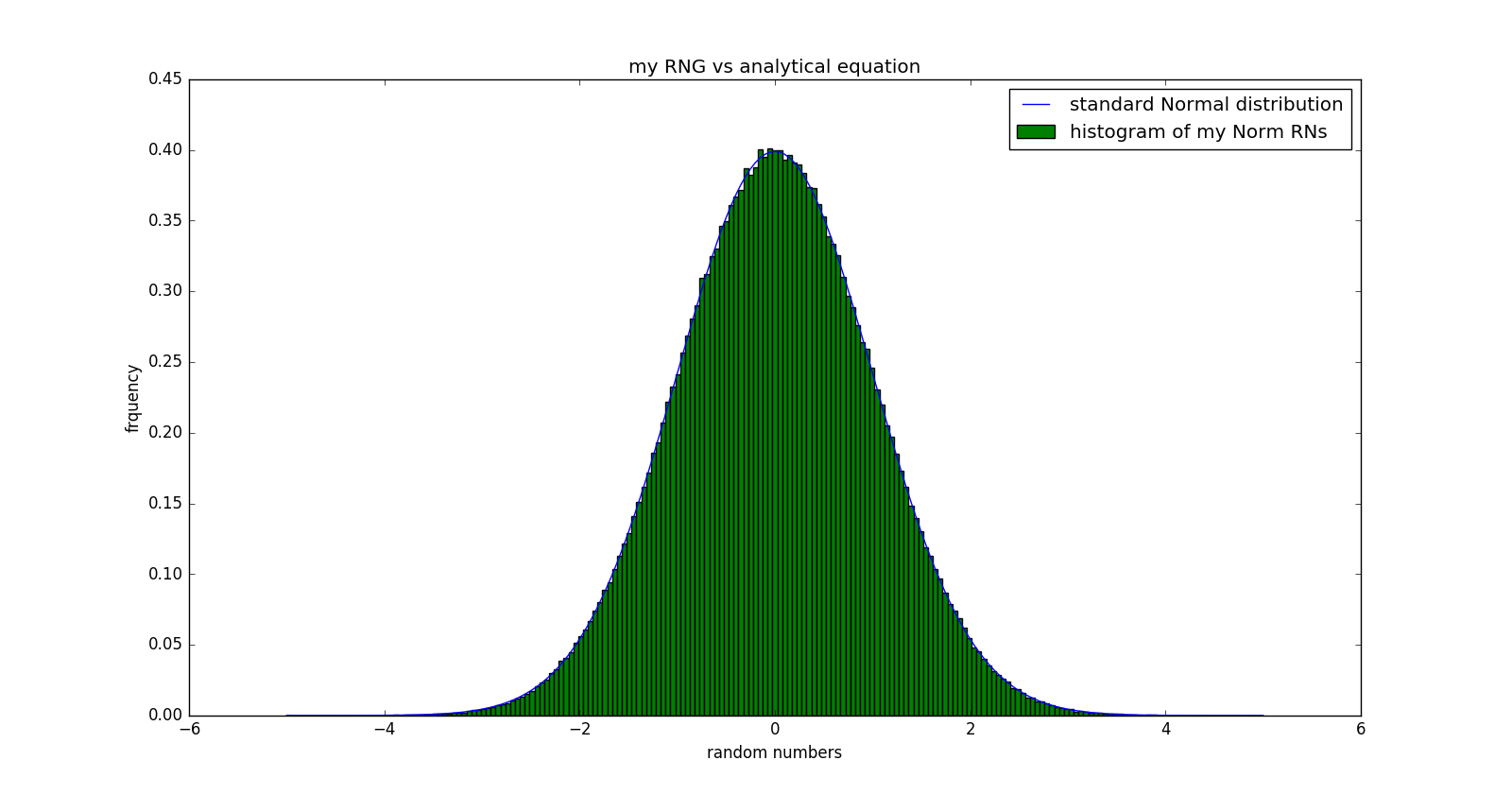
Use 10000 steps and L=5, N\_B=200 plot the histogram and scatter frequency vs analytical values as below:

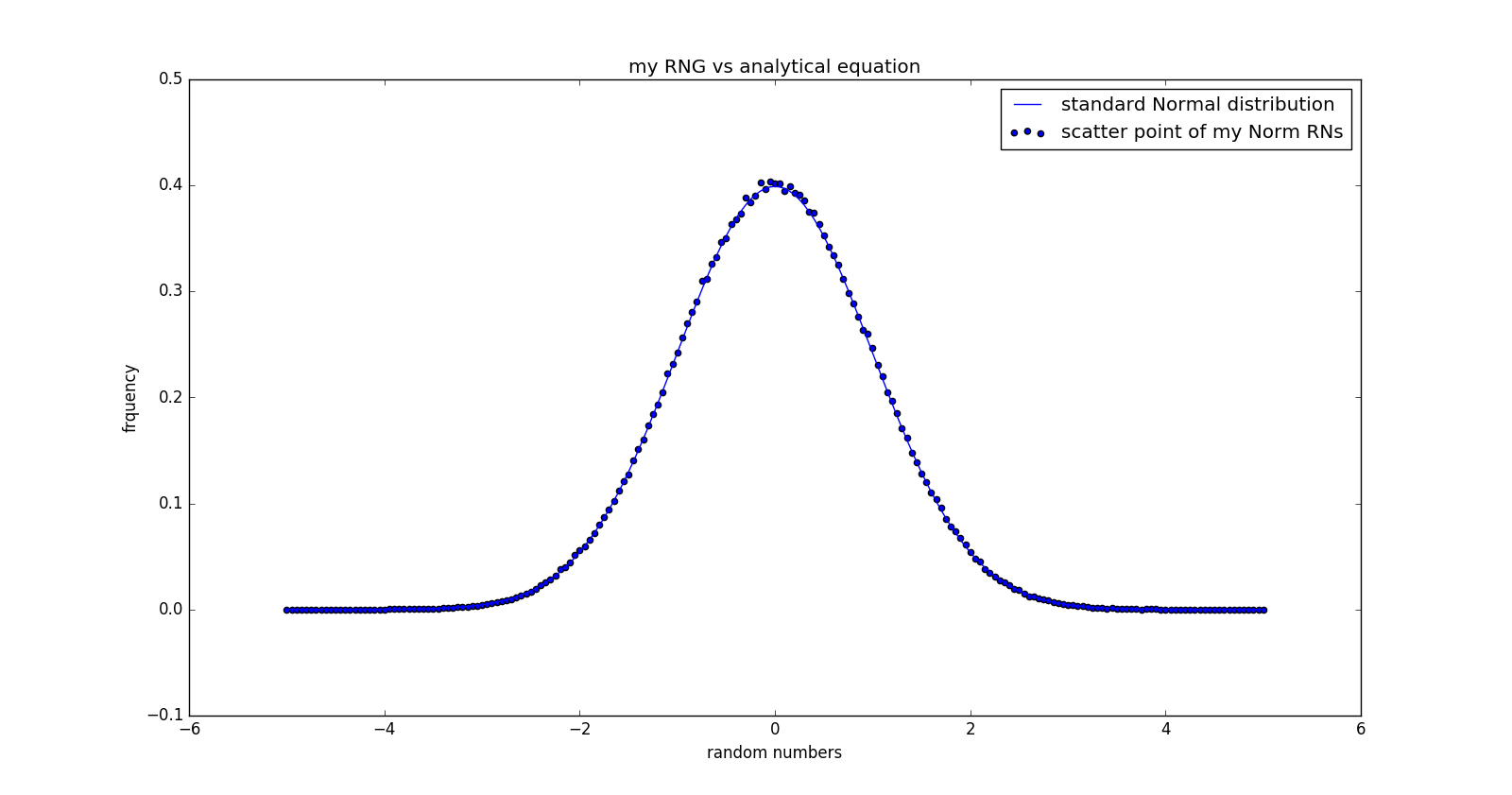
****



as can be seen from the figure above that my normal random numbers agree with analytical equation very well.

If we increase the number of Random Number to 1,000,000, then the histogram and scatter plots will be as follows:





thus we reach a conclusion that my normal RNG is very good.

**Testing Random Number Generators**

Generate a stream of random numbers on the interval [0,1) using your LCG function and report the chi-squared values in 1,2, and 3 dimensions.

For my own LCG function, I use 1,000,000 RNs between (0,1) to test, 16 points per dimension.

For random.random() mudule, 1,000,000 RNs between (0,1) to test, 16 points per dimension.

For supplied files, use 16 points per dimension.

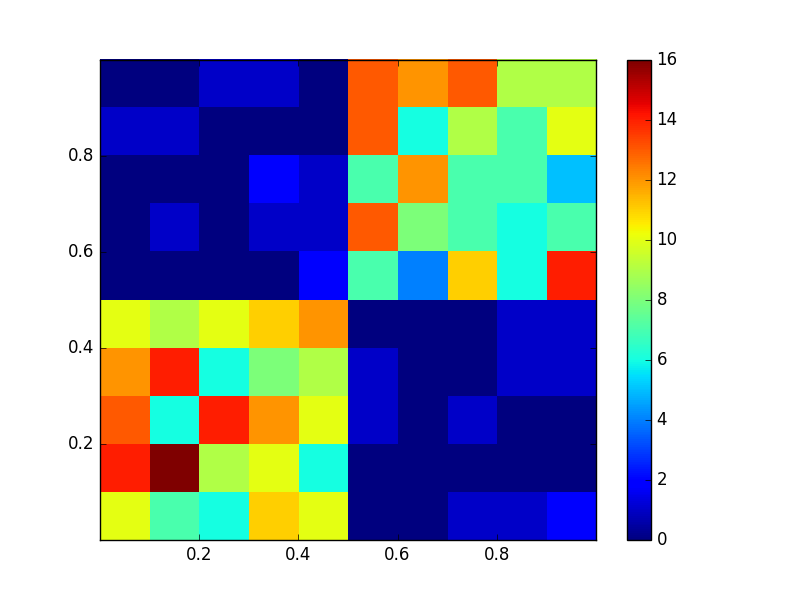
So freedom is 15 for 1D, 225 for 2D and 3375 for 3D.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RNGs | Dimemsion | Chi\_squared | 0.9 upper limit | Pass the test ? |
| My LCG | 1D | 16.898 | 22.307 | yes |
| 2D | 250.504 | 252.578 | yes |
| 3D | 4002.068 | 3480.709 | no |
| random.random | 1D | 9.728 | 22.307 | yes |
| 2D | 259.998 | 252.578 | no |
| 3D | 3955.742 | 3480.709 | no |
| Supplied file1 | 1D | 13.325 | 22.307 | yes |
| 2D | 43208.467 | 252.578 | no |
| 3D | 18385.606 | 3480.709 | no |
| Supplied file2 | 1D | 17.802 | 22.307 | yes |
| 2D | 232.901 | 252.578 | yes |
| 3D | 29013.380 | 3480.709 | no |
| Supplied file3 | 1D | 22.960 | 22.307 | no |
| 2D | 237.709 | 252.578 | yes |
| 3D | 4044.651 | 3480.709 | no |

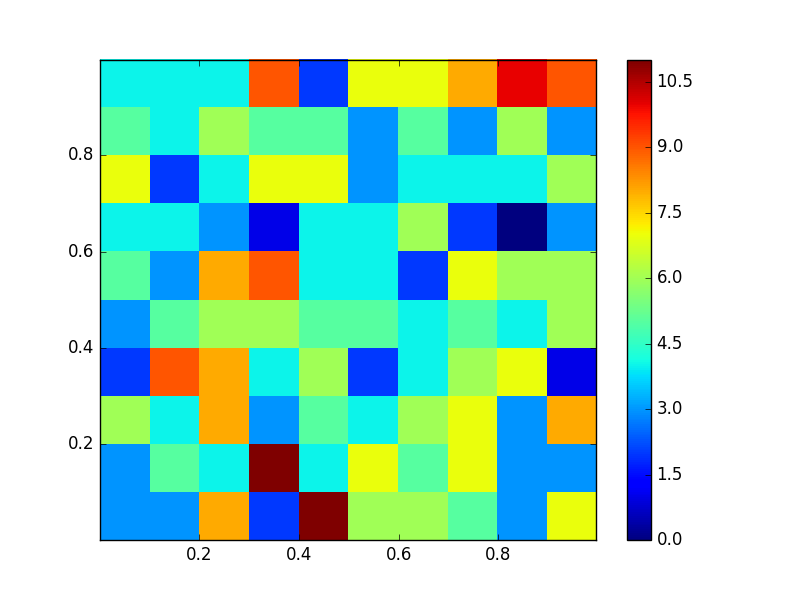
Run it on the first 1000 points of the three supplied datasets.

The number of bin is 10 for each dimension. 1000 data points will only generate 500 pairs of (x,y) in our 2D plot.

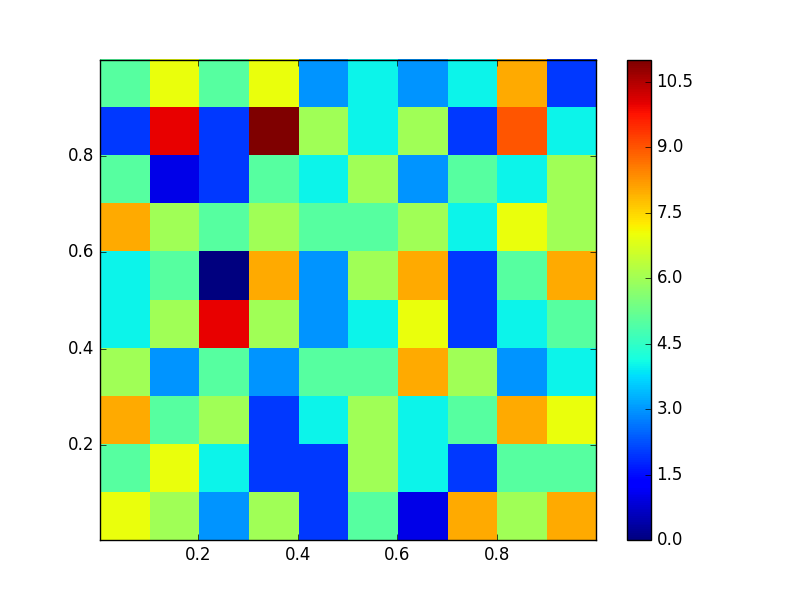
Plot2d for the first 1000 data points in supplied Generator1.dat:



Plot2d for the first 1000 data points in supplied Generator2.dat:



Plot2d for the first 1000 data points in supplied Generator3.dat:



Judging from the 3 plots above, we can easily seen that the x an y for the first data is positively correlated with each other given that populations are much dense either for large a and y, or small x and y.

But for dataset 2 and 3, they are much averaged with small dense cells shown. But over all it can be seen that data3 and data2 are better than data1.

If we combine the three plots with the chi\_squared values calculated above, it can be easily seen that data3 is better than data2 while data2 is better than data1.

**Importance sampling**

A. Numerical calculus

Find the correct expression for Ω, the normalization.

So we have

Choose finite liits of integration as [-L, +L], which you believe are a good approximation to the infinite limits.

First we set N=10,000, which is large enough to ensure precise enough delta\_x. now we change the L values with {1,2,3,4,5,6,7} and extract table and plot by calculus vs L.

calculus = 1.382788 when L = 1.0

calculus = 1.625739 when L = 2.0

calculus = 1.642962 when L = 3.0

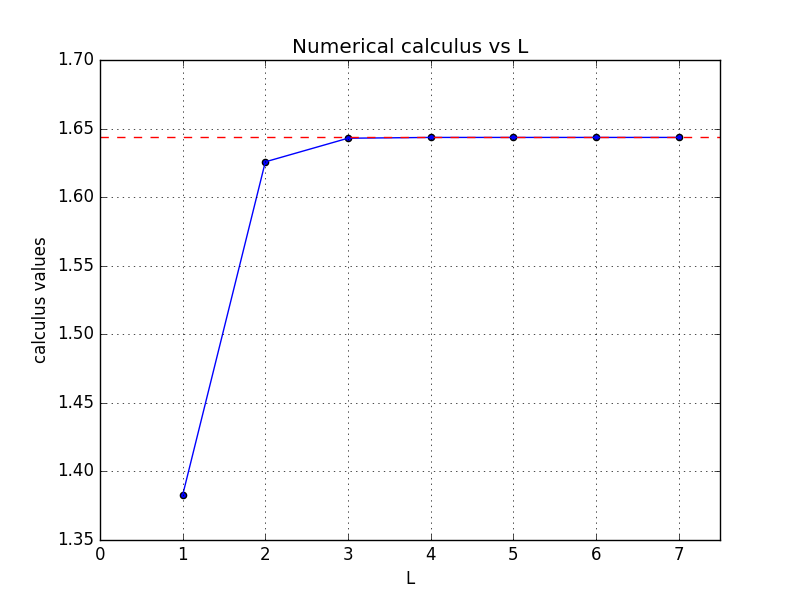
calculus = 1.643536 when L = 4.0

calculus = 1.643545 when L = 5.0

calculus = 1.643545 when L = 6.0

calculus = 1.643545 when L = 7.0

calculus vs L:



so it can be seen that L = 4 would be enough for a good approximation of the original calculus.

To conclude, I used N = 10,000 and L = 4 and finally the calculus will converge to 1.643545

B. importance sampling

Determine an expression for K that normalizes p(x).

Due to a fact that

so

thus

thus and we have

Using the functional forms of p(x) and f(x) given above, write down an analytic expression for the estimator g(x).

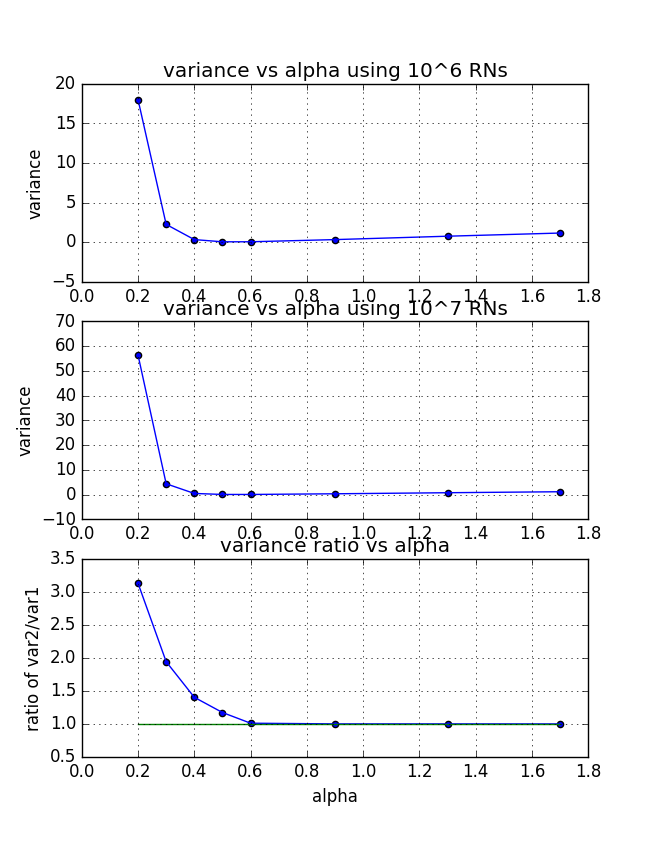
Also write down an expression for the estimator of the variance of I computed with Monte Carlo integration sampling the distribution p(x).

Graph the variance versus alpha and submit this with your code.

Now, I use 1,000,000 and 10,000,000 random numbers and choose 9 alpha and calculate the corresponding integral and variance of each.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| N | alpha | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.9 | 1.3 | 1.7 |
| E6 | Inte. | 1.562245 | 1.628027 | 1.639547 | 1.642349 | 1.643220 | 1.643566 | 1.643925 | 1.6441 | 1.644243 |
| Var. | 189.46 | 18.02 | 2.252 | 0.3329 | 0.05403 | 0.05691 | 0.3324 | 0.7602 | 1.1607 |
| E7 | Inte. | 1.60095 | 1.638466 | 1.642650 | 1.643325 | 1.643487 | 1.643551 | 1.643619 | 1.64364 | 1.643654 |
| Var. | 1002 | 56.44 | 4.375 | 0.4671 | 0.06334 | 0.05750 | 0.3324 | 0.7602 | 1.1603 |

Here is the plot of variance vs alpha for 10^6 RNs and 10^7 RNs and their variance ratios:



**Comment:**

it can be seen that for alpha less than 0.6, the variances are bigger than larger alphas. Moreover, when random numbers are increasing from 10^6 to 10^7, those variances are increasing which can be easily seen from the third subplot about the variance ratios. Thus we reach a conclusion that these variances are infinite. When alphas are greater than 0.6, although there are a little bit increasing of variance trend, but the ratios are converging to 1, showing good variance stability.

Based on the above table and plots, the optimal choice of alpha could be 0.6.

Submit a graph that graphs alpha versus (the ratio your calculated variance with 4× as many steps to the original calculated variance)). Also indicate where you have infinite variance.

The second subplot used 10 times more RNs than the first subplot above (from 10^6 to 10^7) and the third subplot gave the ratio of the two variance sets.

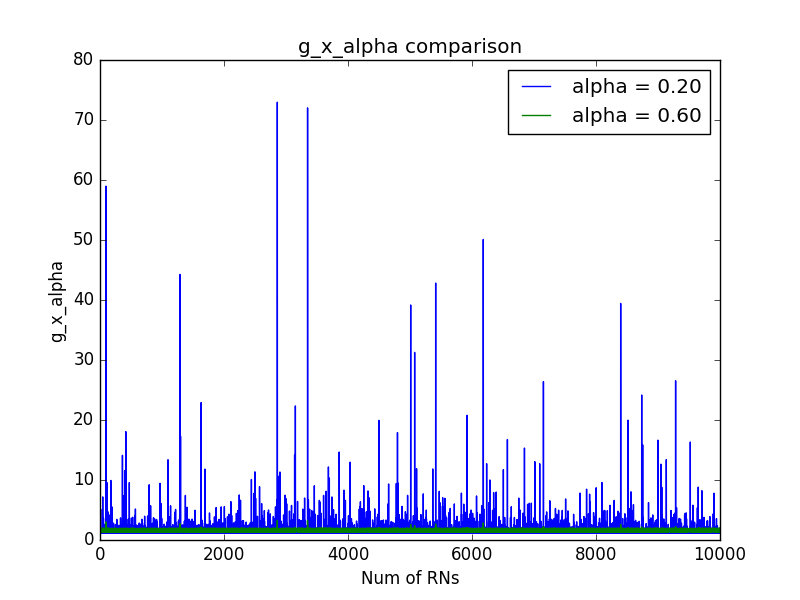
Calculate the variance analytically, expressing your answer as a function of α.

State the range of values of alpha for which the variance is infinite.

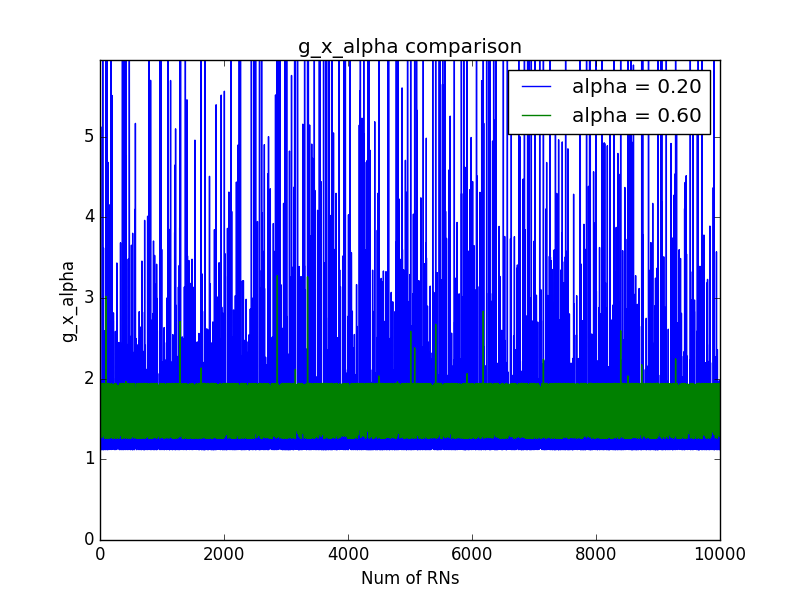
The equation above shows the dependence of variance on that of alpha. If we want the calculus to be finite, then the integrator must be finite, which means is finite. Further more, we need >0 and thus a >0.5. On the other hand, when a < 0.5, we will end up with infinite variance.

Make a trace plot of g(x,α) versus sample number for a value of α that has infinite variance and a value of α that has finite variance. Describe the qualitative difference between these two situations.

Here is the g(x,alpha) for alpha =0.2 and 0.6:



if we zoom in we will find that



we find that g(x, alpha =0.2) has a lots of spikes, causing infinite variance, while g(x, alpha =0.6) is very stable and convergent, with finite variance.