Assignment 1: Algorithms Galore!

Computational Statistics
Instructor: Luiz Max de Carvalho

November 23, 2021

Hand-in date: 22/11/2021.

General guidance

- State and prove all non-trivial mathematical results necessary to substantiate your arguments;
- Do not forget to add appropriate scholarly references *at the end* of the document;
- Mathematical expressions also receive punctuation;
- Please hand in a single PDF file as your final main document.

  Code appendices are welcome, *in addition* to the main PDF document.

Background

By now we have seen quite a few methods for computing integrals *via* Monte Carlo. Each method has its own advantages and drawbacks. It is important that we understand these properties in order to apply the methods effectively. In this assignment we will continue studying the problem of computing the average distance between two points on a disk, this time from the perspective of method comparison. That is to say that in this assignment you will experience the microcosm of comparing several methods for solving a problem for which we happen to know the right answer in closed-form.

Recall that we want to compute

$$
\begin{aligned}
I &= \frac{1}{\pi^2 R^4} \int_0^R \int_0^R \int_0^{2\pi} \int_0^{2\pi} \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos\phi(\theta_1, \theta_2)}\, r_1 r_2 \, d\theta_1 \, d\theta_2 \, dr_1 \, dr_2, \\
&= \frac{2^7}{45\pi} R,
\end{aligned}
$$

where $\phi(\theta_1, \theta_2)$ is the central angle between r_1 and r_2.

Here we will do something a bit risky: we will compare a few methods to compute I using a bunch of different methods without knowing in advance which will work best or if there are going to be any differences at all. Welcome to Science!

Methods

Here we will list a selection of methods that will be randomly assigned to each student, along with some questions that need to be answered for that particular method.

- **Rejection sampling**
  - Justify your choice of proposal distribution and show that it conforms to the necessary conditions for the algorithm to work; in particular, try to find a proposal that gives the highest acceptance probability.
- **Importance sampling**
  - Justify your choice of proposal based on the variance of the resulting estimator.
- **Gibbs sampling**
  - Write your full conditionals out and show that they adhere to the Hammersley-Clifford condition.
- **Metropolis-Hastings**
  - Justify your choice of proposal; test different ones if you need to.
- **Static Hamiltonian Monte Carlo**
  - Comment on the choice of step size (ε) and integration time (τ).

Questions

1. You have been (randomly) assigned a method from the previous section. Represent I as $\int_{\mathcal{X}} \phi(x)\pi(x)\,dx$ and justify your choice of ϕ, π and $\mathcal{X}$. Recall that these choices are arbitrary up to a point, but they might lead to wildly different empirical performances **and** theoretical properties for estimators of I. **Justify** your choices in light of the method you have been given to work with. Choose wisely and be rigorous in your justifications.

2. Again, starting from the eventual samples you will obtain with your method, construct a non-empty[1] family of estimators of I and discuss whether it is (strongly) consistent and whether a central limit theorem can be established.

3. Detail a suite of diagnostics that might be employed in your application to detect convergence or performance problems. Extra points for those who design algorithms that exploit the structure of this particular integration problem.

4. For each $R \in \{0.01, 0.1, 1, 10, 100, 1000, 10000\}$, perform $M = 500$ runs from your simulation method and compute: (i) variance (ii) bias (iii) standard deviation of the mean (MCSE).

5. Can you identify one key quantity missing from the previous item? *Hint:* it bears relevance to the real world application of any computational method.

Warning: the questions in this assignment might seem deceptively simple; do not be fooled. I expect a lot of effort from you in making your method work the best it can. This entails loads of failed derivations and experiments, which you are encouraged to report in order to document the discovery process. Also, feel free to include answers to questions that have not been asked, if you feel they are relevant. Make loads of figures and tables and let your scientific imagination run wild! Good luck!

Solutions:

First of all, let's understand the problem that we want to solve, we have $\mathbf{X}$ and $\mathbf{Y}$ two points on a disk of radius R that we represent as $\mathcal{D}_{\mathcal{R}}$ and suppose these two points are choosed independently uniformly over $\mathcal{D}_{\mathcal{R}}$.
So, we think about them as iid samples:

$$\mathbf{X}, \mathbf{Y} \sim \mathcal{U}(\mathcal{D}_{\mathcal{R}})$$

And we can write their density functions as:

$$\pi_{\mathbf{X}}(x) = \frac{\mathbb{1}(x \in \mathcal{D}_{\mathcal{R}})}{\pi R^2}$$

$$\pi_{\mathbf{Y}}(y) = \frac{\mathbb{1}(y \in \mathcal{D}_{\mathcal{R}})}{\pi R^2}$$

[1]This is a joke. It means you should come up with at least one estimator. But you might, and are even encouraged to, entertain more than one estimator.

So there joint distribution is given by there product because they are independent:

$$\pi_{\mathbf{X},\mathbf{Y}}(x,y) = \frac{\mathbb{1}(x \in D_R)\mathbb{1}(y \in D_R)}{\pi^2 R^4} = \frac{\mathbb{1}(x, y \in D_R)}{\pi^2 R^4} \tag{1}$$

where $\mathbb{1}(x, y \in D_R)$ is the product of indicator functions $\mathbb{1}(x \in \mathcal{D}_\mathcal{R})$ e $\mathbb{1}(y \in \mathcal{D}_\mathcal{R})$.

Based on that samples our objective is to compute the average distance between this two points in our disk D_R, i.e, we have to compute:

$$\begin{aligned}
\mathbb{E}_\pi(||\mathbf{X} - \mathbf{Y}||_2) &= \int_\mathcal{X} \|x - y\|_2 \pi_{\mathbf{X},\mathbf{Y}}(x,y) dx dy \\
&= \frac{1}{\pi^2 R^4} \int_{\mathbb{R}^4} \|x - y\|_2 \mathbb{1}(x, y \in D_R) dx dy
\end{aligned}$$

1. To answer question 1, let's represent this integral:

$$I = \frac{1}{\pi^2 R^4} \int_0^R \int_0^R \int_0^{2\pi} \int_0^{2\pi} \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos\phi(\theta_1, \theta_2)} r_1 r_2 \, d\theta_1 \, d\theta_2 \, dr_1 \, dr_2$$

As $\int_\mathcal{X} \phi(\mathbf{x})\pi(\mathbf{x}) \, d\mathbf{x}$ for a suitable choice of ϕ, π and $\mathcal{X}$ where $\mathbf{x} \in \mathbb{R}^4$. Recall that I is the average distance between the two points but written in polar coordinates, returning to the cartesian coordinates we have:

$$\sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos\phi(\theta_1, \theta_2)} = \sqrt{(x_1^2 - x_2^2) + (y_1^2 - y_2^2)} = \|x - y\|_2$$

where $x = (x_1, y_1)$ and $y = (x_2, y_2)$. And $r_1 r_2$ is just the jacobian matrix determinant of the transformation from cartesian to polar coordinates.

Given the explanation above we can se that $\mathcal{X} = \mathbb{R}^4$(because each point has two coordinates).
And π are given by:

$$\pi \sim \mathcal{U}(\mathcal{D}_R \times \mathcal{D}_R)$$

with $\pi(\mathbf{x}) = \pi(x, y) = \frac{\mathbb{1}(x,y \in D_R)}{\pi^2 R^4}$.
And ϕ:

$$\begin{aligned}
\phi \colon \mathbb{R}^4 &\to \mathbb{R} \\
x, y &\mapsto \phi(\mathbf{x}) = \phi(x, y) = \|x - y\|_2
\end{aligned}$$

So we have:

- $\phi(\mathbf{x}) = \phi(x, y) = \|x - y\|_2$
- $\pi(\mathbf{x}) = \pi(x, y) = \frac{\mathbb{1}(x,y \in \mathcal{D}_\mathcal{R})}{\pi^2 R^4}$
- $\mathcal{X} = \mathbb{R}^4$

Given this definitions out integral I just become our equation $\mathbb{E}_\pi(||\mathbf{X} - \mathbf{Y}||)$.
In light of the Gibb's Sampling method that was chosen for me in this assignment I have to derive the full conditionals of this problem, first

notice that I have to sample from the $(U)(D_R)$ twice, so for both the first and the second point we have our joint distribution as

$$\pi_{\mathbf{x},\mathbf{y}}(x,y) = \frac{\mathbb{1}(x, y \in \mathcal{D}_{\mathcal{R}})}{\pi R^2}$$

Now calculating the marginals, we have:

$$\pi_{\mathbf{x}}(x) = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{dy}{\pi R^2} = \frac{2\sqrt{R^2-x^2}}{\pi R^2}$$

for $x \in [-R, R]$. And the marginal of y is

$$\pi_{\mathbf{y}}(y) = \int_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}} \frac{dy}{\pi R^2} = \frac{2\sqrt{R^2-y^2}}{\pi R^2}$$

Using both of then and with the joint distribution we can calculate the conditionals. We have

$$\pi_{\mathbf{x}|\mathbf{y}}(x|y) = \frac{\pi_{\mathbf{x},\mathbf{y}}(x,y)}{\pi_{\mathbf{y}}(y)} = \frac{\frac{1}{\pi R^2}}{\frac{2\sqrt{R^2-y^2}}{\pi R^2}} = \frac{\mathbb{1}(x^2+y^2 \leq R^2)}{2\sqrt{R^2-y^2}}$$

And,

$$\pi_{\mathbf{y}|\mathbf{x}}(y|x) = \frac{\pi_{\mathbf{x},\mathbf{y}}(x,y)}{\pi_{\mathbf{x}}(x)} = \frac{\frac{1}{\pi R^2}}{\frac{2\sqrt{R^2-x^2}}{\pi R^2}} = \frac{\mathbb{1}(x^2+y^2 \leq R^2)}{2\sqrt{R^2-x^2}}$$

Note that both of them is constant because in the case $\mathbf{x}|\mathbf{y}$, y and R are given and on the other case, x are given. So we have that both as uniform on the circle since if a density function is equal a constant then this distribution is uniform. Then,

$$\pi(\mathbf{x}|\mathbf{y}) \sim \mathcal{U}(\mathcal{D}_{\mathcal{R}})$$

$$\pi(\mathbf{y}|\mathbf{x}) \sim \mathcal{U}(\mathcal{D}_{\mathcal{R}})$$

And we can show that these conditionals adhere to the Hammersley-Clifford theorem given below:

Hammersley-Clifford Theorem: Consider a distribution whose density $\pi(x_1, x_2, ..., x_d)$ satisfies the positivity condition.
Then for any $(z_1, ..., z_d)(\pi)$, i.e. $\pi(z_1, ..., z_2) > 0$, we have

$$\pi(x_1, x_2, ..., x_d) \propto \prod_{j=1}^{d} \frac{\pi(x_j|x_1, ..., x_{j-1}, z_{j+1}, ..., z_d)}{\pi(z_j|x_1, ..., x_{j-1}, z_{j+1}, ..., z_d)}$$

So to show adherense to this theorem lets start by showing that our π satisfy the positivity condition, but this is quite trivial, just see that

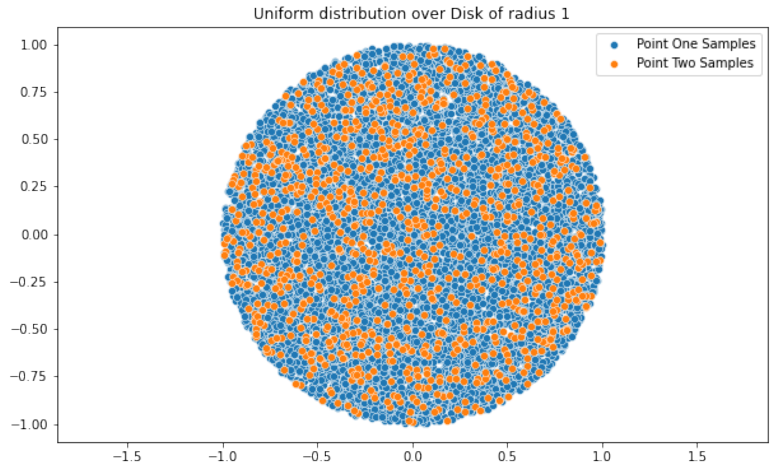

Figure 1: Two points sample generated by Gibbs sampling on disk of radius 1.

$\pi_X(x) > 0$ and $\pi_X(y) > 0$ per definition and $x, y \in [-R, R]$ so $\pi(x, y) > 0$ as well.

Returning to the Hammersley-Clifford condition we have that:

$$\pi(x_1, x_2, ..., x_d) \propto \frac{\pi(x_1|z_2)\pi(x_2|x_1)}{\pi(z_1|z_2)\pi(z_2|x_1)} = \frac{\frac{1}{2\sqrt{R^2-z_2^2}}\frac{1}{2\sqrt{R^2-x_1^2}}}{\frac{1}{2\sqrt{R^2-z_2^2}}\frac{1}{2\sqrt{R^2-x_1^2}}} = 1$$

So our proposed π satify the theorem.

2. Using the Gibb's sampling method we can generate a random sample from both **X** and **Y** in the following way:

Algorithm 1 Gibbs sampling

Require: $p0, M, R > 0$
  $x, y \leftarrow p0$
  $sample \leftarrow zeros(shape = [M, 2])$
  for i=1, 2, ..., M **do**
    $x \leftarrow uniform(-\sqrt{R^2 - y^2}, \sqrt{R^2 - y^2})$
    $y \leftarrow uniform(-\sqrt{R^2 - x^2}, \sqrt{R^2 - x^2})$
    $sample[i] \leftarrow [x, y]$
  end for

This algorithm generates a sample of points in the disk of radius R as we can see in Figure[1].

And we can see the histograms of both sample points as well as we can see in Figure[2].

With these two points generated from the algorithm above we can calculate our distance sample, from where we can get several estimators of interest,

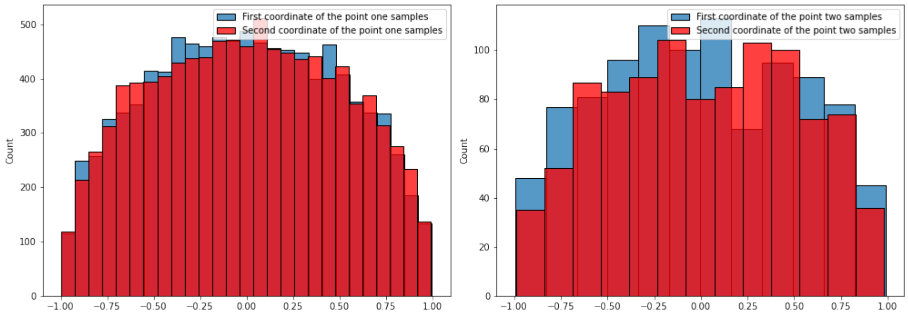

Figure 2: Histogram of the generated point samples.

but now we going to focus on the mean of the distance samples which is our interest from the start.

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} \|x_i - y_i\|_2 \tag{2}$$

With these formula we can compute the distance sample like in Figure[3].

We can check that:

$$\mathbb{E}_\pi\left[\hat{I}\right] = \mathbb{E}_\pi\left[\frac{1}{n}\sum_{i=1}^{n}\|x_i - y_i\|_2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\pi\|x_i - y_i\|_2 = \mathbb{E}_\pi\|x - y\|_2$$

which is our original expectation and using the Strong law of large numbers it's easy to see that the above estimator is consistency with respect to the true distance.

3. To check for convergency I write a function that iterates through both sample sizes and radius to find if our method converge in all cases, first I plot the traceplot of the estimator as we can see in Figure[4] for $R = 1$. We can see that we have convergence to the true distance given by $\frac{128R}{45\pi} \approx 0.90$.

   This convergence can be seeing more clearly on the running mean plot of the estimator if Figure[5].

   and it's nice to see that our generated samples are very uncorrelated as you can see in Figure[6].

   The resulting table is on item 4. but we can see here a traceplot vision about the estimator when we vary both sample size and radius(until now all results was calculated varying the sample size but maintaining radius fixed at 1.) at Figure[7]. It's intereseting to see that even when we vary the radius we still have convergence to the true distance value.

   and we can see too traceplots for the variance and the bias when we vary the radius of the circle in the Figure[8]. We can see there that as we vary the radius and the sample size increases our estimator become unbiased and even if the variances increases as we vary radius, it become stabilized

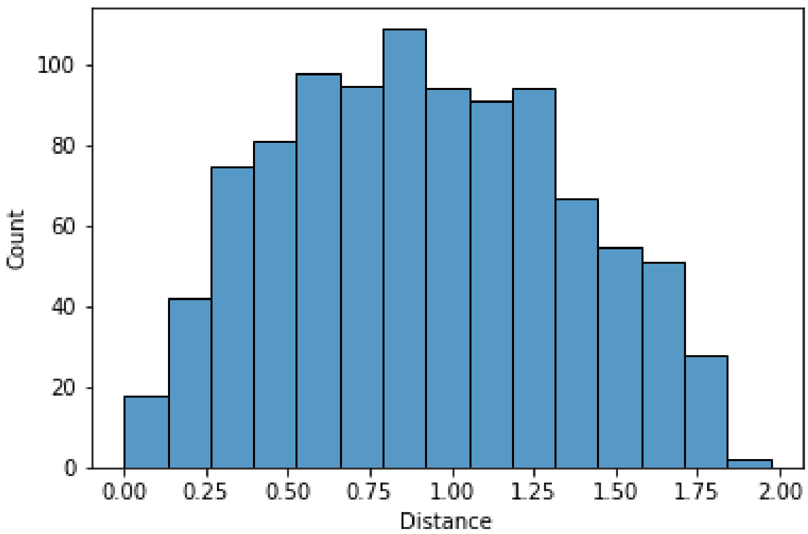

Figure 3: Histogram of the distance samples.

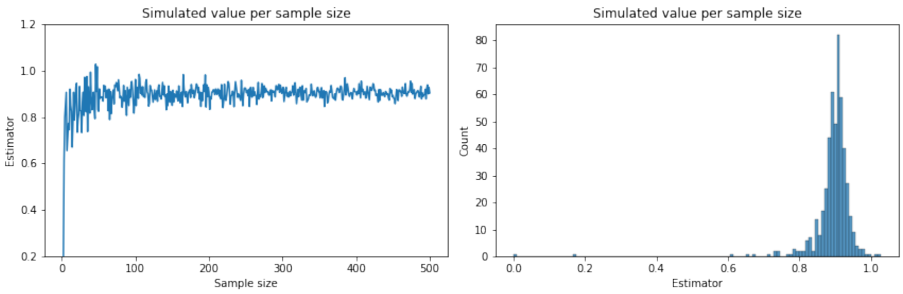

Figure 4: Traceplot of estimator on the right and histogram of the estimator on the left.

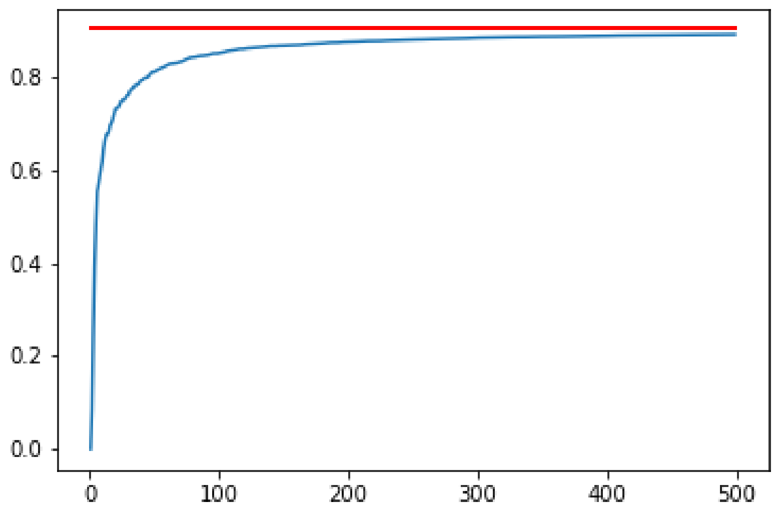

Figure 5: Running mean of the estimator. The red line is the true value of the distance.

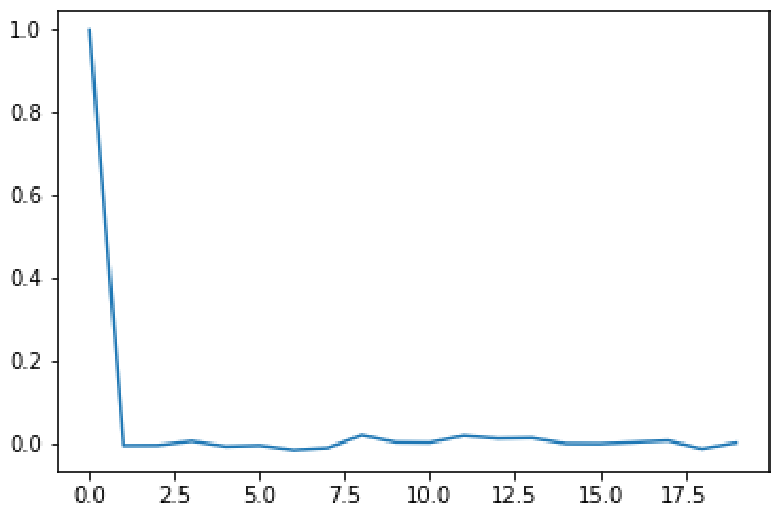

Figure 6: Auto correlation of generated samples.

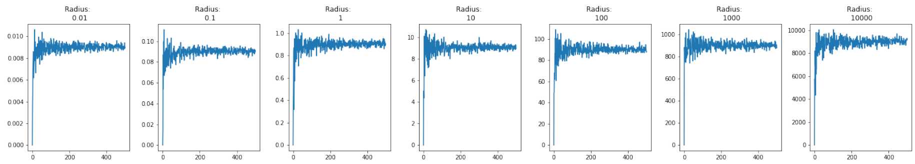

Figure 7: Estimator trace plots for diferent radius.

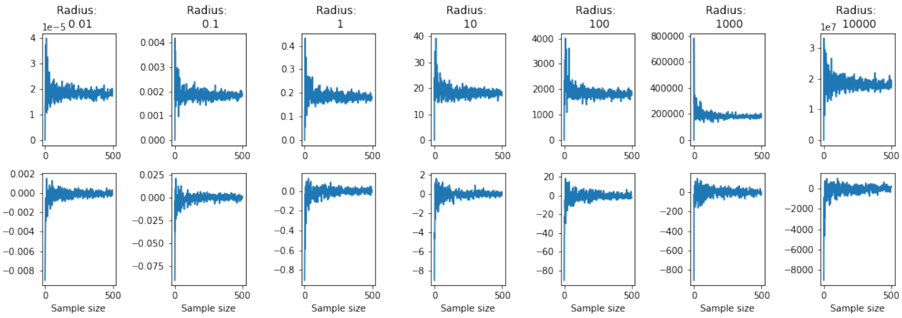

Figure 8: Variance trace plots in the first row and bias traceplot in the second row for diferent radius.

when we inscreases the sample size which is pretty good result for this method.

4. The results of the sampling for $R = [0.01, 0.1, 1, 10, 100, 1000, 10000]$ and $M = 500$ are presented below:

Radius	**I**	**Estimator**	**Variance**	**Bias**	**MCSE**
0.01	0.009054	0.009144	1.806568e-05	0.000089	0.000009
0.10	0.090541	0.090740	1.912285e-03	0.000198	0.000087
1.00	0.905415	0.909606	1.788386e-01	0.004191	0.000846
10.00	9.054148	9.028753	1.879007e+01	-0.025395	0.008670
100.00	90.541479	91.251696	1.728992e+03	0.710217	0.083162
1000.00	905.414787	904.994200	1.786637e+05	-0.420587	0.845372
10000.00	9054.147874	9041.915680	1.923669e+07	-12.232193	8.771930

5. Along with the sampling results and diagnostics I find useful to calculate the execution time of all sampling pipeline, I understand that for a computational sampling method be efficient, it must be accurate in terms of convergence to the target distribution and by efficient in terms of computational complexity. So I re-run the code that generate the table above but now with an aditional column named Time (in seconds).

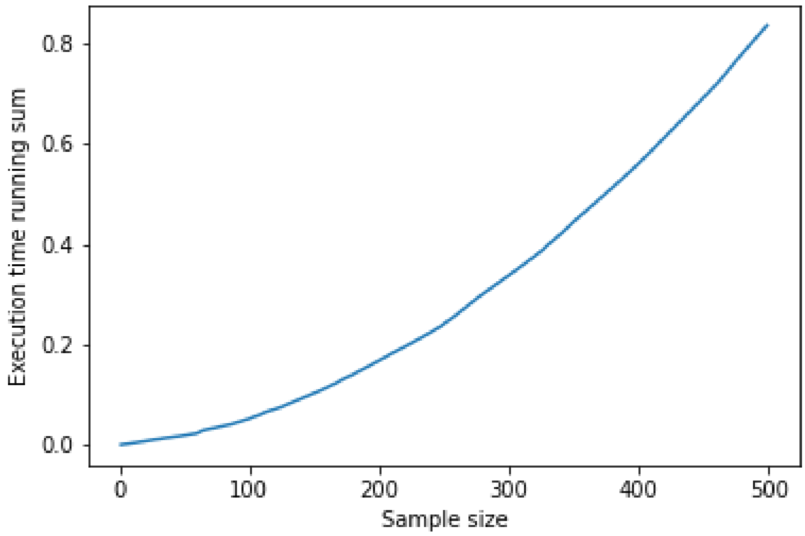

Figure 9: Running time sum when we vary the sample size.

Radius	**I**	**Estimator**	**Variance**	**Bias**	**MCSE**	**Time**
0.01	0.009054	0.009004	1.903609e-05	-0.000051	0.000009	0.006087
0.10	0.090541	0.088841	1.840069e-03	-0.001701	0.000086	0.005868
1.00	0.905415	0.913883	1.687984e-01	0.008469	0.000822	0.008612
10.00	9.054148	9.118159	1.837349e+01	0.064011	0.008573	0.006845
100.00	90.541479	90.705381	1.846345e+03	0.163903	0.085938	0.008689
1000.00	905.414787	900.506327	1.854449e+05	-4.908461	0.861266	0.007783
10000.00	9054.147874	9192.231899	1.716505e+07	138.084025	8.286144	0.007407

And if we fix the radius to 1 for example and only vary the sample size we can see that the execution time running sum increases quadraticaly with the sample size as we can see in the Figure[9].