Project 2: Monte Carlo Methods

EE 511 – Section Tuesday 5:00pm—5:50pm
Name: Yiqi Feng
Student ID #: 9057129035

● Problem A

1. Problem Statement

Estimate π by the area method including confidence intervals on your estimate. Draw a graph of the successive values of the estimator as the number of samples increases.

How many points do you need to use for your estimate to be within $\pm 1\%$ of the true value of π (with probability 0.95)?

2. Theoretical Analysis

Assuming that the space π is a circle whose formula is $x^2 + y^2 \le 1$, we construct a mathematical model.

Generate pairs of U(0,1) RV's (Xi,Yi).

$$\text{Let } P_i = \begin{cases} 1 & (X_i^2 + Y_i^2 \le 1) \\ 0 & otherwise \end{cases}$$

Than we can get the estimate for $\pi/4$ by

$$\hat{p} = \frac{\sum_{i=1}^{n} P_i}{n}$$

We can also compute the sample variance of the P_i:

$$s_{p_i}^2 = \frac{\sum_{i=1}^{n}(P_i - \hat{p})^2}{n-1}$$

and then:

$$s_{\hat{p}}^2 = VAR\left\{\frac{\sum_{i=1}^{n} P_i}{n}\right\} = \frac{1}{n^2}\sum_{i=1}^{n} s_{p_i}^2 = \frac{1}{n} s_{p_i}^2$$

Then we find the confidence interval using the sample size statistics.

$$\Pr\{p - \beta s_{\hat{p}} \le \hat{p} \le p + \beta s_{\hat{p}}\} = 1 - \alpha$$

Also

$$\Pr\{-\beta s_{\hat{p}} \le \hat{p} - \mathrm{p} \le \beta s_{\hat{p}}\} = 1 - \alpha$$

In this problem, $\alpha = 0.05$, n = 10000. From the standard table of values for the normal distribution, I find that $\beta = 1.96$.

By simulation we can get that $s_{\hat{p}} = 0.004157$, $\hat{p} = 0.7873$. So the confidence interval with probability for 0.95 is that:

$$\Pr\{0.7854 - 1.96 * 0.004157 \le \hat{p} \le 0.7854 + 1.96 * 0.004157\} = 0.95$$

By computing we can get the confidence interval for $\pi/4$ is that

$$\Pr\{0.7854 - \ 0.008148 \le \hat{p} \le 0.7854 + 0.008148\} = 0.95 \quad for\ n = 10000$$

So the confidence interval for π is that

$$\Pr\{3.1416 - \ 0.03256 \le \hat{p} \le 3.1416 + 0.03256\} = 0.95 \quad for\ n = 10000$$

and the estimate of π is 3.1492.

According to the formula above, I find when the estimate is within $\pm 1\%$ of the true value π.

$$\beta s_{\hat{p}} \le \pi/4 * 1\%$$

Also,

$$s_{\hat{p}} = \sqrt{\frac{\sum_{i=1}^{n}(P_i - \hat{p})^2}{n(n-1)}} \le 0.004007$$

3. Simulation Methodology

I set the sample size from 10 to 10000. In each sample, I use the formula $x^2 + y^2 \le 1$ to count the number of points which are in the range of quadrant, and then use the ratio between that number and the sample size to get the estimator of $\pi/4$.

In order to get the number of points when the estimate is within $\pm 1\%$ of the true value π, I use a loop to calculate $s_{\hat{p}}$ with various sample size. The loop ends until

$$s_{\hat{p}} \le 0.004007$$

. ***Experiments and Results***

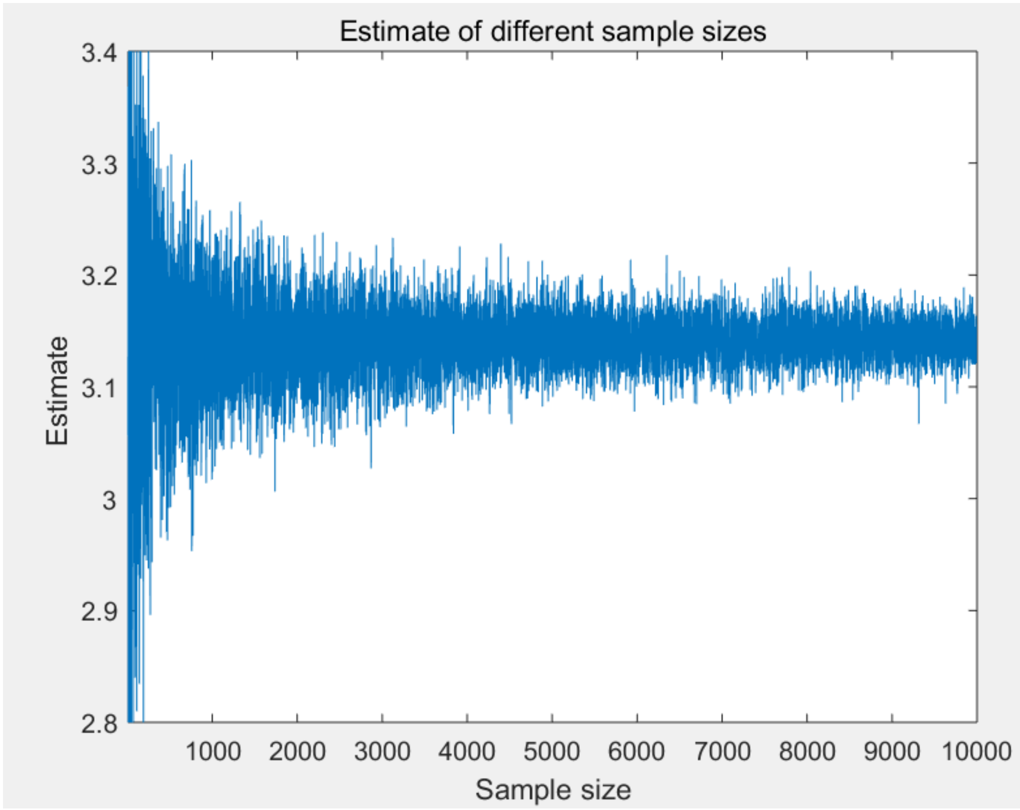

Figure 2.1 Estimate of different sample sizes

From the figure2.1, we can see that the larger the sample size is, the closer to π value of the estimate is. Also, we can see the value of estimate is floating up and down around π, and the range of the floating is becoming smaller and smaller. Therefore, we can predict that when the sample size is infinite, the estimate will be the true value, π.

```
i =

      10068
```

Figure 2.2 Number of points

From the figure2.2, we can see that the number of points is 10068 when the estimate is within ±1% of the true value π.

5. Source Code

The code for Problem A:

Part(1)

```
size = 10000;
a = zeros(1,size);
sample = zeros(1,size);
for j=10:size   %the sample size start from 10 to 10000
    x = rand(2,j);
    y = zeros(1,j);
    quadrant=0;
    for i=1:j
        if (x(1,i)^2+x(2,i)^2)<=1
           y(i) = 1;
           quadrant=quadrant+1;
        end
   sample(1,j)=sqrt(var(y)/j);
   a(1,j)=quadrant/j*4;
end
plot(a(1,10:size));
axis([10 size 2.8 3.4]);
title('Estimate of different sample sizes');
ylabel('Estimate ');
xlabel('Sample size');
```

Part(2)

```
a = zeros(1,100000);

sample = zeros(1,100000);

for i=10:100000

    x = rand(2,i);

    y = zeros(1,i);

    for j=1:i

        if (x(1,j)^2+x(2,j)^2)<=1

           y(j) = 1;

        end

    end

   sample(1,i)=sqrt(var(y)/i);

   if sample(1,i)<=0.004007

       break

   end

end

  i
```

● Problem B

1. Problem Statement

Consider a deck of cards (for simplicity numbered 1.. N). Use a uniform random number generator to pick a card and record what card it is (if you were using actual cards, you would replace the card back into the deck – that is not necessary here since we never really take the card out of the deck). Repeat this N times, recording the number of times that each of the cards is selected. Some cards may not show up (actually, it is very likely that several card numbers will not show up) and some will show up more than once. You can use this data to estimate the following probabilities:

$$p_j = \Pr\{\text{a card will be selected } j \text{ times in the } N \text{ selections}\}$$

It is unlikely that any card will show up more than about 10 times. Run this for $N = 10, N = 52, N = 100, N = 1{,}000, N = 10{,}000$ and verify that $p_0 \simeq 1/e$. Can you also find values for the other p_j based on a mathematical analysis?

2. Theoretical Analysis

For a deck of N cards, the possibility that a card is chosen is $\frac{1}{N}$, and the possibility that a card is not chosen is (1 - $\frac{1}{N}$).

Thus, after N-time picking,

$$p_0 = (1 - \frac{1}{N})^N$$

Than we can get the estimate for p_0 by

$$\widehat{p_0} = \frac{\sum_{i=1}^{N} p_{0_i}}{N}$$

We can also compute the sample variance of the p_{0_i}:

$$s_{p_i}^2 = \frac{\sum_{i=1}^{N} (p_{0_i} - \widehat{p_0})^2}{N-1}$$

and then:

$$s_{\widehat{\mathrm{p0}}}^2 = VAR\left\{\frac{\sum_{i=1}^{N} p_{0_i}}{N}\right\} = \frac{1}{N^2}\sum_{i=1}^{N} s_{\widehat{\mathrm{p0}}}^2 = \frac{1}{N} s_{p_{0_i}}^2$$

Then we find the confidence interval using the sample size statistics.

$$\Pr\left\{p_0 - \beta s_{\widehat{p0}} \le \widehat{p_0} \le p_0 + \beta s_{\widehat{p0}}\right\} = 1 - \alpha$$

Also

$$\Pr\left\{-\beta s_{\widehat{p0}} \le \widehat{p_0} - p_0 \le \beta s_{\widehat{p0}}\right\} = 1 - \alpha$$

In this problem, $\alpha = 0.05$, n = 10000. From the standard table of values for the normal distribution, I find that $\beta = 1.96$.

By simulation we can get that $s_{\widehat{p0}} = 1.0148\mathrm{e} - 04$, $\widehat{p_0} = 0.3678$. So the confidence interval with probability for 0.95 is that:

$$\Pr\{0.36788 - 1.96 * 1.0148\mathrm{e} - 04 \le \widehat{p_0} \le 0.36788 + 1.96 * 1.0148\mathrm{e} - 04\} = 0.95$$

So the confidence interval for $1/e$ is that

$$\Pr\{0.36768 \le \widehat{p_0} \le 0.36808\} = 0.95 \quad for\ n = 10000$$

and the estimate of $1/e$ is 0.3678.

3. Simulation Methodology

I set the sample size from 10 to 10000. In each sample size I , I count the number of cards (sum) which were chosen 0 time. Thus, p0 = sum / I.

By simulation we can get that $s_{\widehat{p0}} = 1.0148\mathrm{e} - 04$, $\widehat{p_0} = 0.3678$ with the probability 0.95.

Finally, we can obtain the figure bellow.

4. Experiments and Results

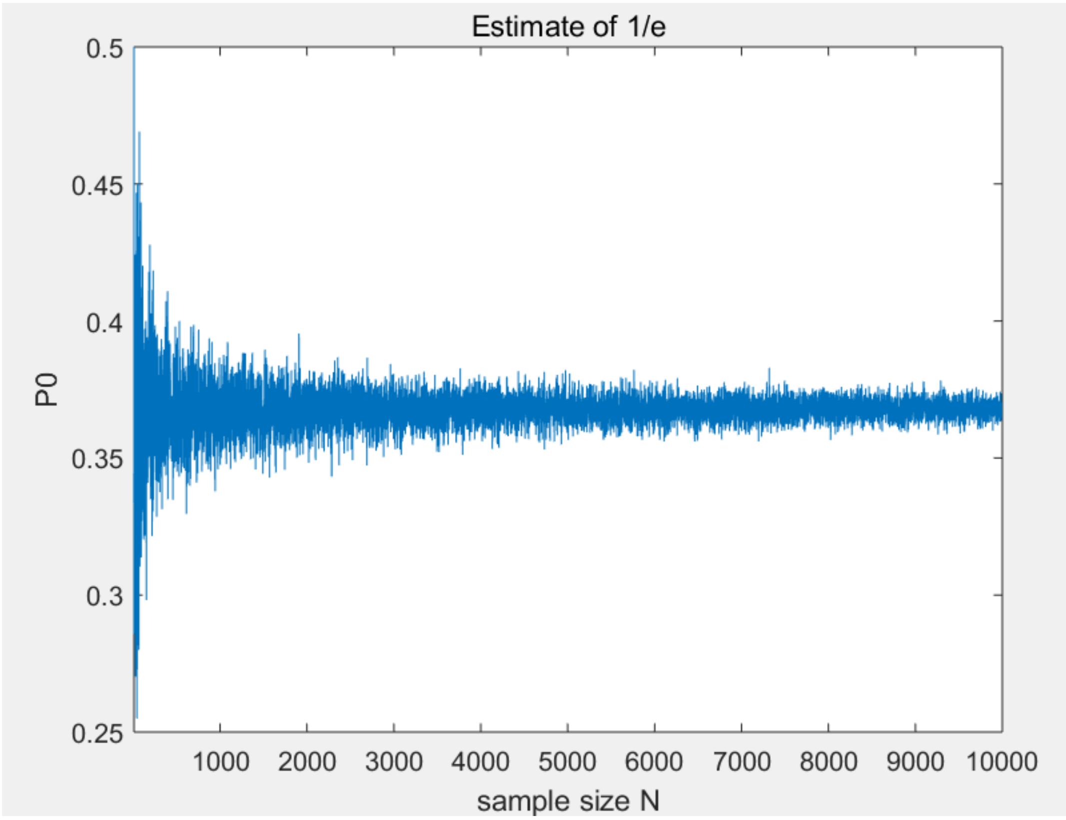

Figure 2.3 Estimate of different sample sizes

From the figure2.3, we can see that the larger the sample size is, the closer to true value 1/e the estimate is. Also, we can see the value of estimate is floating up and down around 0.3678, and the range of the floating is becoming smaller and smaller. Therefore, we can predict that when the sample size is infinite, the estimate will be the true value, 0.3678.

5. Source Code

The code for Problem B:

```
a=zeros(1,10000);
for i=1:10000
    N=zeros(1,i);
    for k=1:i
        x=ceil(rand(1,1)*i);
        N(1,x)= N(1,x)+1;
    end
    sum=0;
    for j=1:i
     if N(1,j)==0
       sum=sum+1;
     end
    end
    a(1,i)=sum/i;
end
sample=sqrt(var(a)/10000)
figure(1);
plot(a(1,10:10000));
axis([1 10000 0.25 0.5]);
title('Estimate of 1/e');
ylabel('P0');
xlabel('sample size N');
```

● Problem C

1. Problem Statement

Use the method discussed in class to find $\hat{y}$, an estimate for Y and find a 95% confidence interval for the value of the integral.

$$Y = \int_0^{\pi} \frac{\sin(x)}{x}\, dx$$

2. Theoretical Analysis

Suppose we want to evaluate

$$I = \int_0^1 g(x)dx$$

if the RV U is uniform (0,1) then

$$\bar{U} = E[g(U)] = I$$

So we can evaluate the integral I by generating a sequence {Ui}. And then by using the strong law of large numbers with probability we can get that

$$\sum_{i=1}^{k} \frac{g(u_i)}{k} \to E[g(U)] = I \; as \; k \to \infty$$

in order to evaluate the integral

$$Y = \int_0^{\pi} \frac{\sin(x)}{x} dx$$

We make a substitution

$$y = \frac{x}{\pi} \; and \; dx = \pi dy$$

and then the integral becomes

$$Y = \int_0^1 \frac{\sin(\pi y)}{\pi y} \pi dy = \int_0^1 \frac{\sin(\pi y)}{y} dy$$

Therefore,

$$Y = E\left[\frac{\sin(\pi y\ )}{y}\right], y \sim U(0,1)$$

Besides,

$$\hat{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

We can also compute the sample variance of the Y_i:

$$s_{y_i}^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{y})^2}{n-1}$$

and then:

$$s_{\hat{y}}^2 = VAR\left\{\frac{\sum_{i=1}^{n} Y_i}{n}\right\} = \frac{1}{n^2}\sum_{i=1}^{n} s_{y_i}^2 = \frac{1}{n} s_{y_i}^2$$

Then we find the confidence interval using the sample size statistics.

$$\Pr\left\{y - \beta s_{\hat{y}} \le \hat{y} \le y + \beta s_{\hat{y}}\right\} = 1 - \alpha$$

Also

$$\Pr\left\{-\beta s_{\hat{y}} \le \hat{y} - \mathrm{y} \le \beta s_{\hat{y}}\right\} = 1 - \alpha$$

In this problem, $\alpha = 0.05$, n = 10000. From the standard table of values for the normal distribution, I find that $\beta = 1.96$.

By simulation we can get that $s_{\hat{y}} = 0.0101$, $\hat{y} = 1.8451$. So the confidence interval with probability for 0.95 is that:

$$\Pr\{1.8519 - 1.96 * 0.0101 \le \hat{y} \le 1.8519 + 1.96 * 0.0101\} = 0.95$$

By computing we can get the confidence interval is that

$$\Pr\{1.832104 \le \hat{y} \le 1.871696\} = 0.95 \quad for\ n = 10000$$

and the estimate is 1.8451.

3. Simulation Methodology

In Matlab, we can use function sinint(pi) to obtain the true value , 1.8519. I set the sample size from 1 to 10000. In each sample, I generate k numbers(a) between 0 and 1 and get the average of all value y=sin(pi*a)/a, in which k means sample size。

4. Experiments and Results

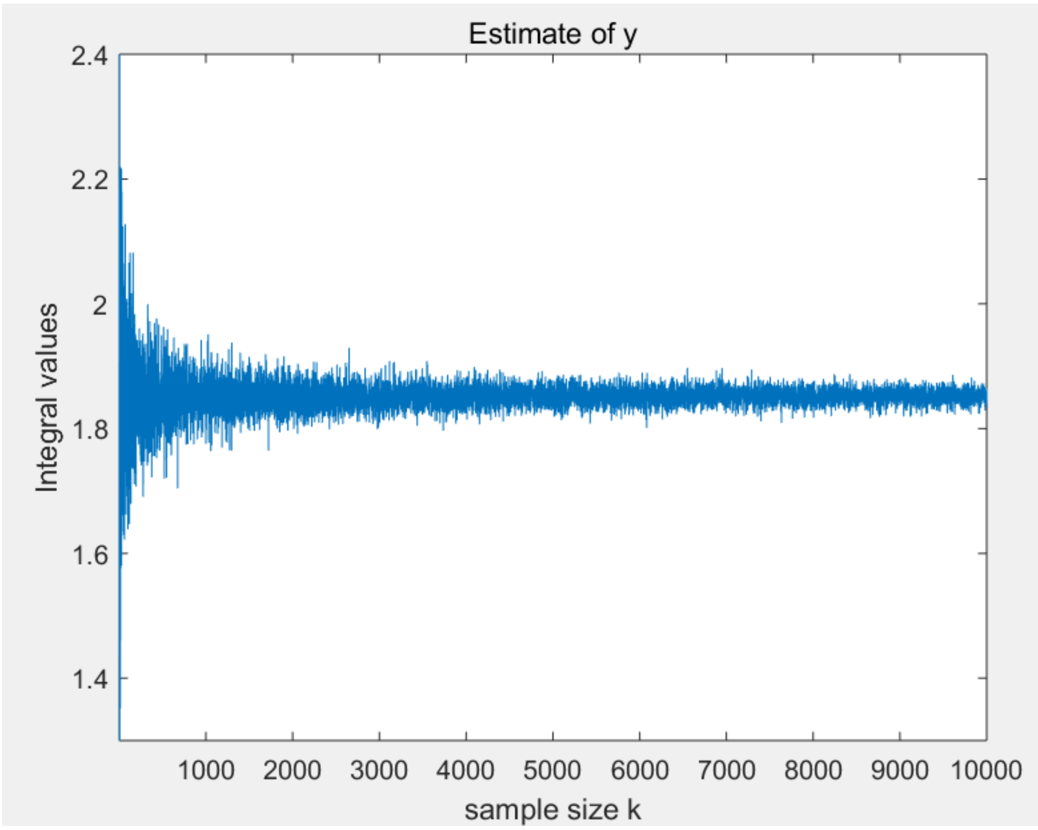

Figure2.3 Integral interval from 0 to π of different sample size

From the figure2.3, we can see that the larger the sample size is, the closer to true value 1.8519 the estimate is. Also, we can see the value of estimate is floating up and down around 1.8519, and the range of the floating is becoming smaller and smaller. Therefore, we can predict that when the sample size is infinite, the estimate will be the true value, 1.8519.

5. Source Code

The source code for Problem C:

```
sum = zeros(1,10000);
    for k = 1:10000
        a = rand(1,k);
        sum(1,k) = 0;
        y = zeros(1,k);
        for j = 1:k
            y(1,j) = sin(pi*a(1,j))/a(1,j);
            sum(1,k) = sum(1,k) + sin(pi*a(1,j))/a(1,j)/k;
        end
    end
    sample = sqrt(var(y))/100;     %s_ŷ when k=10000
figure(1);
plot(sum(1,:))
title('Integral interval from 0 to pi');
ylabel('Integral values');
xlabel('sample size k');
```