Lab Assignment #1

STATS 210, Session 1, Fall 2021

Yiyuan Qin

yq74

Content

Problem 1 **Buffon's Needle experiments**

Ideas:

From the reference book, I knew that d ($0 \le d \le \frac{a}{2}$) represents the distance between the midpoint of a needle and the nearest parallel line, and θ ($0 \le \theta \le \pi$)represents the angle between the needle and the parallel line. A is the rectangle whose length and height are respectively π and $\frac{a}{2}$. Condition $d \le \frac{l}{2} sin\theta$ is the shadow area. The probability $p = \frac{The\ Area\ of\ U}{The\ Area\ of\ A} = \frac{0.5\int_0^{\pi} l\ sin\theta\, d\theta}{0.5a\pi} = \frac{2l}{\pi a} = \frac{n}{N}$, where n is the number of needles in U and N is the whole needles number. Simplifying the formula, I got $\pi = \frac{2lN}{an}$, which was used to estimate the value of π.

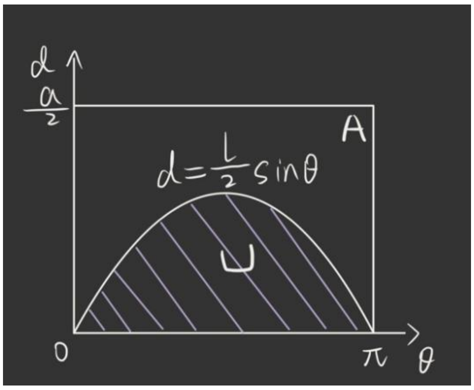


Figure 1- The visualization of how to estimate π by probability

Process:

problem1.m is a function to estimate π by inputting *N*, *l*, *a*, where *N* is the number of trials, *l* the length of needles, and *a* is the length of the parallel lines. *problem1_plot.m* will output 4 value of π with 4 different trials and plot the estimated values versus trial numbers using logarithmic coordinate for the number of trials.

Result:

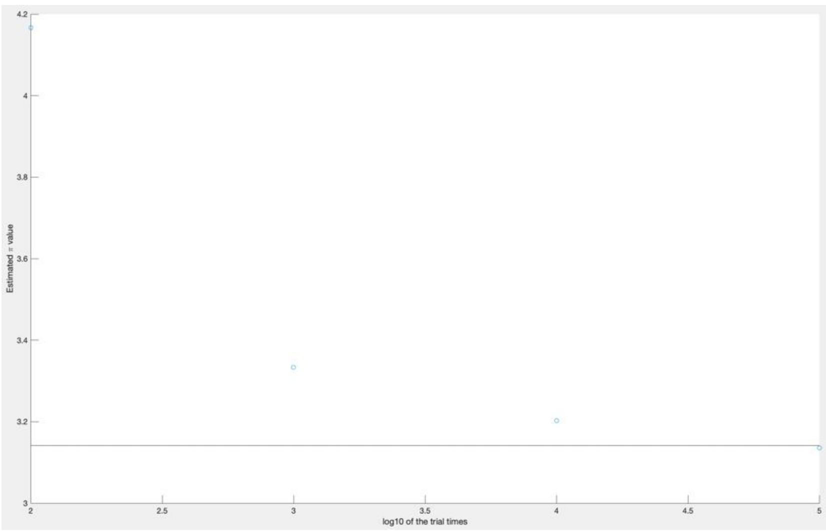


Figure 2- The estimated values versus trial numbers by running 1 time of *problem1_plot.m*

As the figure showing, the trend is converged. The black line is $Estimated\ value = \pi$. In order to minimize the chance error and observe the converge trend more obviously, I ran the *problem1_plot.m* several times and plotted all the results.

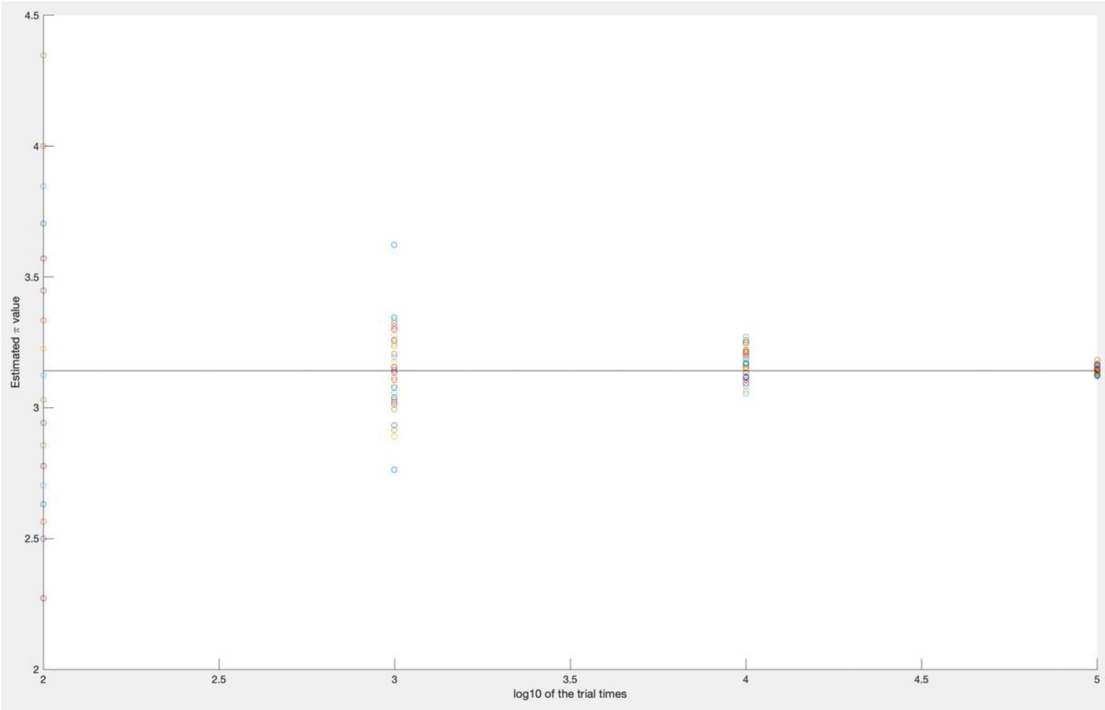


Figure 3- more time running result showing the convergence

In figure 3, it is more easily to get the conclusion that with the trials increasing, the estimated value is more accurate and the whole trend is convergent.

Problem 2 Distinct birthday

Ideas:

First, the problem said that "nobody is born on February 29, we have 365 days for one year", which meant that there were 365 kinds of birthday. Describing each birthday as 1~365 instead of month/day, there were 365 probability for each person. Simulated N = 1000 times and checked whether there were 2 people have had different birthday. If so, this simulation was invalid, vice versa. $P_i = \frac{valid\ simulation}{N}$ Moreover, I needed to draw the PMF when i was less than 200. As a result, added one outer loop to iterate the $P(I = i)$ and summed when I less than i. Finally, drew the plot of PMF. For mathematics part, there were totally 365n probability because everyone had 365 kinds of probable birthday. To meet the requirement, if the first person had 365 choices, the second person only had 364 choices… so the P (n people have distinct birthday) = $\frac{365!}{365^n(365-n)!}$

Process:

problem2.m is a function [pn] = problem2(n)where *n* is the people in one simulation, and *pn* is the probability corresponding to *n*. This file will print out the *pn* and draw the PDF with *pn* on the figure. Problem2_plot.m will draw the PDF of mathematics result.

Result:

To show the result, I input n = 40 and get [pn] = 0.1120

```
>> problem2(40)

ans =

    0.1120
```

Figure 4- P (40) with my function

Also, a PMF is plotted as below. The trend of it is descending and almost 0 when *n* greater than 60.

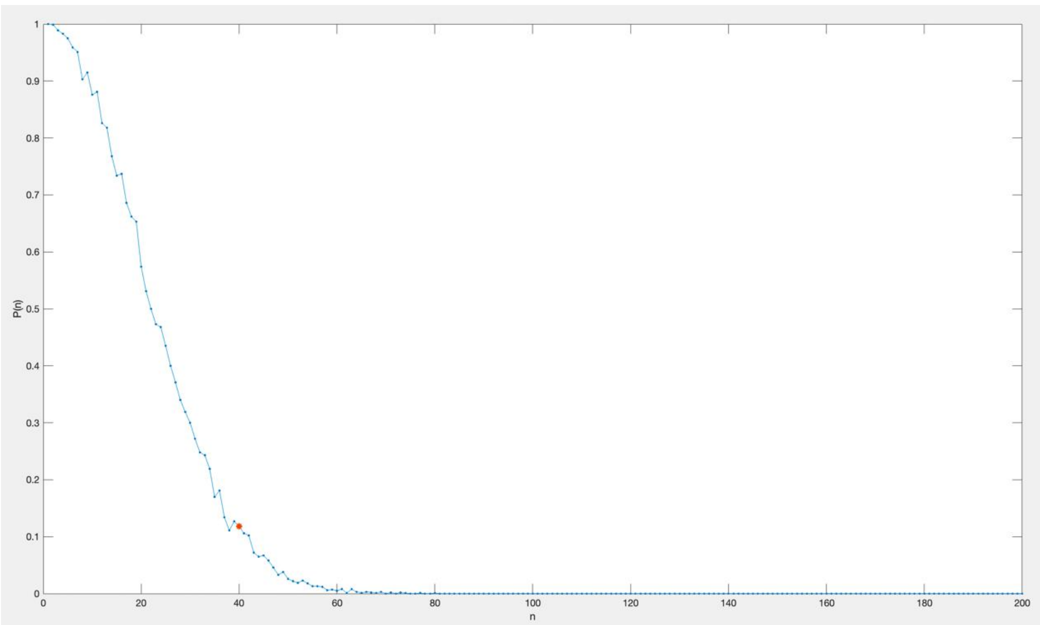


Figure 5- PMF of distinct birthday of simulation

For mathematics part, the figure is more smooth.

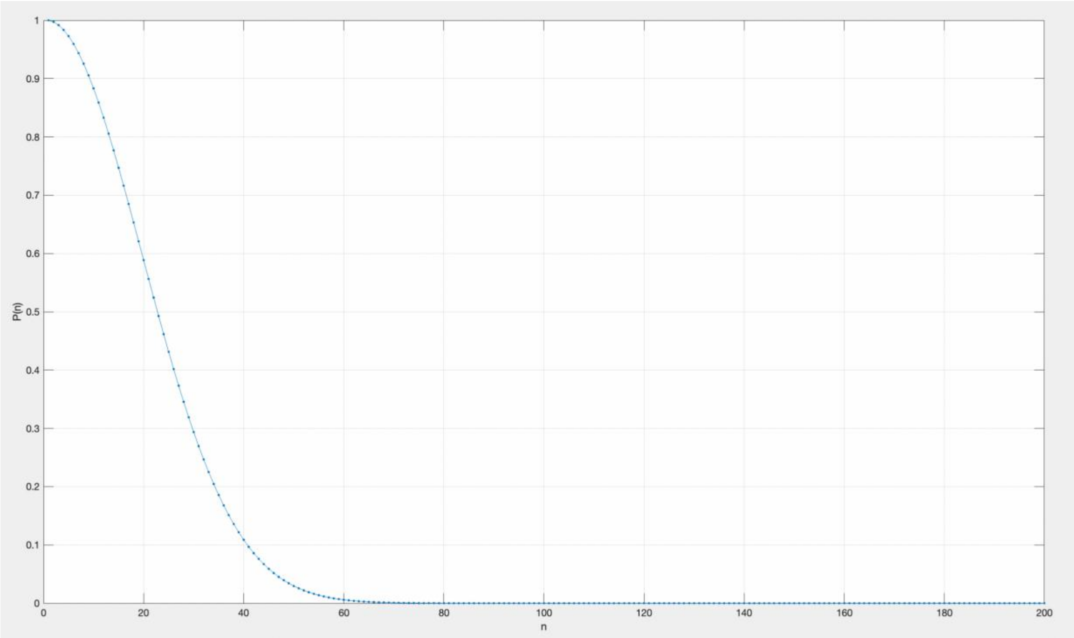


Figure 6- The PMF of distinct birthday of mathematics

Problem 3 7 Cards

Ideas:

Drew 7 cards from the top is not different with drawing 7 cards from the shuffling cards. Although there were 13 kinds of cards, it could be distinguished as 2 kinds: King and the other cards. A is the King, which had 4 elements; B is the other cards, which had 48 elements. The problem required 7 cards, 3 from A and 4 from B. This was a hypergeometric distribution.

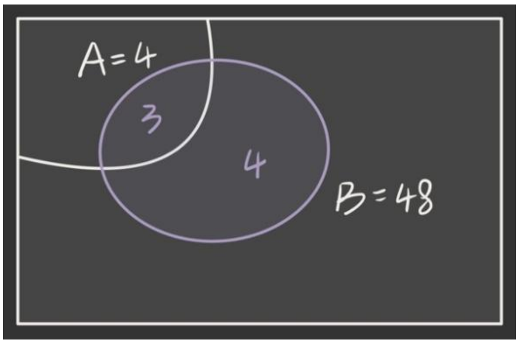


Figure 7- The illustration of how to pick cards

To simplify the case, set King as 1, the other cards as 0. The 52 cards were represented by 4 ones' and 52 zeros'. Shuffled the 52 number, did simulation

1000000 times, and picked up the first 7 numbers. If the sum of the 7 number was 7, there were 3 King in the first 7 numbers, which satisfied the case. Counted the satisfied number as *n*. Finally, $p = \frac{n}{1000000}$

Result:

The probability calculated by mathematics:

$$p = \frac{C_3^4 C_4^{48}}{C_7^{52}} = 0.005818$$

The probability simulated by Matlab:

```
>> problem3
   0.0058
```

Figure 8- The result of problem 3 simulated by Matlab

Compared the two probabilities, the mathematics answer is proven.

Problem 4 7 Value of ln2

Ideas:

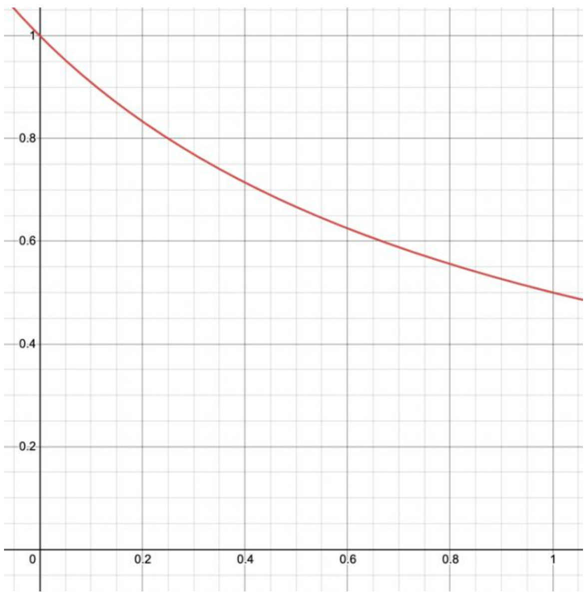


Figure 9- The figure of $y = \frac{1}{x+1}$

Drew the curve of $y = \frac{1}{x+1}$ to visualize the area and made sure the curve was under y=1. Randomly generated 100000 points, which x and y were both between (0,1).

Count the points under the curve as n. $p = \frac{n}{100000}$ Mathematics method was taking the integral from zero to one of this function.

Result:

The area under $y = \frac{1}{x+1}$ is 0.6923 by simulation.

```
>> problem4
    0.6923
```

Figure 10- The result of problem4 by simulation

Calculated the area by mathematics:

$$\int_0^1 \frac{1}{x+1} = ln(x+1)|_0^1 = ln2$$

ln2 is about 0.6923 by my simulation.

Here, the absolutely error is 0.12%.

```
>> area = 0.6923

area =

    0.6923

>> abs_error = abs(area - log(2))/log(2)

abs_error =

    0.0012
```

Figure 11- Absolutely error calculation

Problem 5 Geometric random variable

Ideas:

CDF of geometric distribution was $F_X = P(X \le x) = 1 - (1-p)^x$. Since $P(X \ge 4) = 1 - P(X \le 3)$, I only needed to know P (X <=3), which meant the first success happened within 3 times. Then I did not need to consider the situation after 3. Calculated the P (success happened within 3 times) had many situations, so I decided to calculate the backside, P (the first success happened 4 or more 4 times). Generated

3 points between 0 and 1 to represent the first 3 times probability. If the point is in (0,0.25), as I drew in the purple line at Figure 12, this point is "success".

Figure 12- probability corresponding to point illustration

If these 3 numbers were all greater than 0.25, the first success must happen after 3 times. Did simulation 100000 times and counted the case that 3 numbers were all greater than 0.25 as *n*. P (success happened within 3 times) = $\frac{100000-n}{100000}$ and $P(X \geq 4) = \frac{n}{100000}$

Result:

Calculate the probability by mathematics:

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - (1 - (1 - 0.25)^3) = 0.75^3 = 0.421875$$

Simulate the probability by Matlab:

```
>> problem5
    0.4215
```

Figure 13- The result of problem5

Compared the two probabilities, my mathematics result gets proven.

Problem 6 Binomial and Poisson

Ideas:

a) By searching online, I found that binopdf and poisspdf could draw the PMF of Binomial distribution and Poisson distribution directly.

b) $\lambda = 1, n = 100, p = 0.01$ so $\lambda = np = 1$

To generate Poisson by Binomial random variable, used binopdf to draw Poisson with n =100, p = 0.01

Process: *problem_PMF.m* can draw the PMF of Binomial and Poisson distribution directly, while *binomial_approximation.m* will draw a Binomial distribution and a Poisson distribution by binomial approximation.

Result:

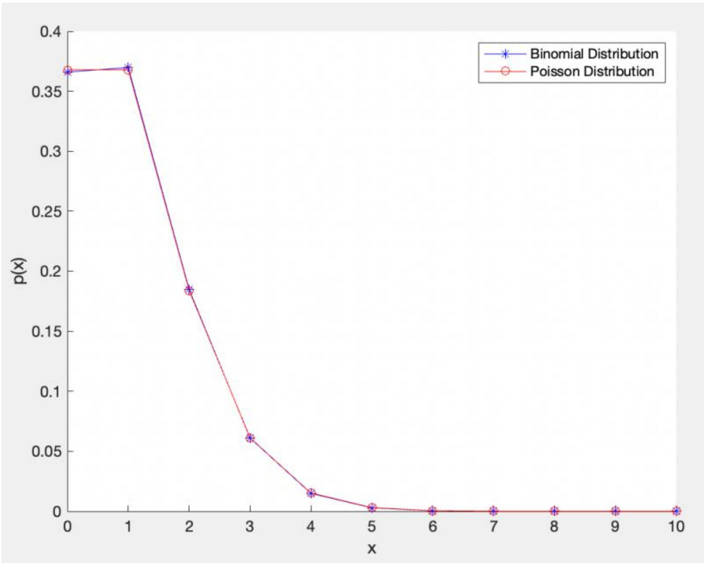
a) 

Figure 14- PMF of Binomial Distribution and Poisson Distribution drawn with respectively binaopdf and poisspdf

From Figure 14, I know that the PMF of these 2 distributions is similar when $\lambda = 1, n = 100, p = 0.01$.

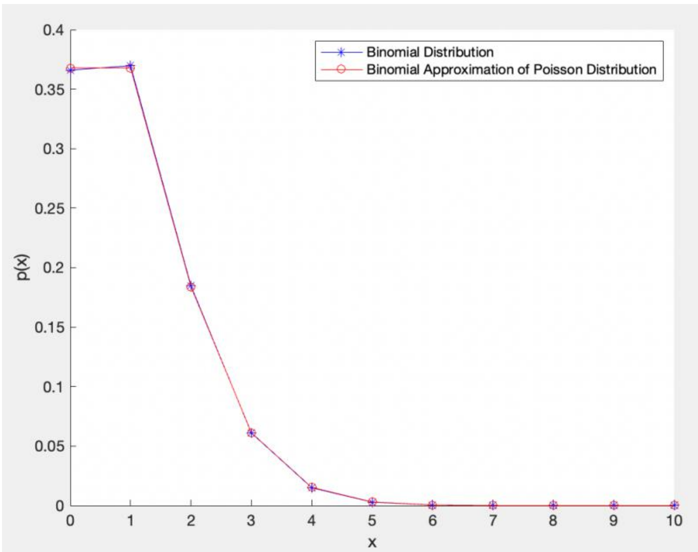
b) 

Figure 15- PFM of Binomial Distribution and Poisson Distribution by binomial approximation

Problem 7 Normal distribution probability

Ideas:

This case asked to simulate the probability, but it was not the same as the problem4. problem 4 could use "needle" to simulate because it had a closed section of x, while in normal distribution it was complex to draw the x section and y section, because there were "sqrt", "fraction" and so on. The best way to solve this problem was find a group of number that satisfied normal distribution and count the number that greater than $\mu + a\sigma$.

Process:

The function [a1,a2,a3] = problem7(mu,sigma) where mu is the average number of this group of number, and sigma is the standard deviation, will output 3 probability. $P\ (X > \mu + a\sigma)$ when a = 1,2,3. Also, the function will print a frequency histogram of the group of number. I also write the ordinary method to simulate the probability. The function is [a1,a2,a3] = normal_way(mu,sigma)It is necessary to estimate the area in the middle,($\mu - a\sigma$,$\mu + a\sigma$) first, and $P\ (X > \mu + a\sigma) = \frac{1 - middle\ area}{2}$ based the characteristic of normal distribution by Monte Carlo Procedure as the forth problem above.

Result:

1. First method:

To simplify the case, I choose X~N (μ=0, σ^2=1)

```
>> [a1,a2,a3] = problem7(0,1)

a1 =

    0.1596


a2 =

    0.0230


a3 =

    0.0013
```

Figure 16- The frequency histogram

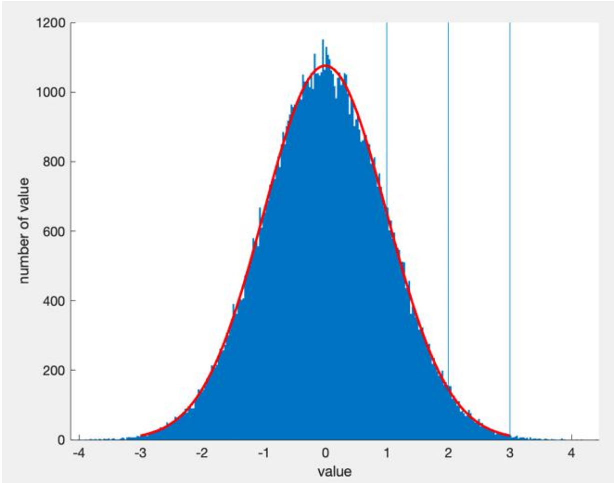


Figure 17- Three probability of problem7

2. Second method:

This is the result of the ordinary method, and the solution is approximately to the first one.

```
>> [a1,a2,a3] = normal_way(0,1)

a1 =

    0.1586


a2 =

    0.0234


a3 =

    0.0016
```

Figure 18- The simulated result by ordinary method

3. Mathematics

$P\ (X > \mu + a\sigma)$

a) $P(X > 1) = 1 - P(X \leq 1) = 1 - 0.8413 = 0.1587$

b) $P(X > 2) = 1 - P(X \leq 2) = 1 - 0.9772 = 0.0228$

c) $P(X > 3) = 1 - P(X \leq 3) = 1 - 0.9987 = 0.0013$

Reference

z表	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0:	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1:	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2:	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3:	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4:	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5:	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6:	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7:	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8:	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9:	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0:	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1:	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2:	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3:	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4:	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5:	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6:	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7:	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8:	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9:	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0:	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1:	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2:	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3:	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4:	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5:	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6:	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7:	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8:	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9:	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0:	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Figure 19- The normal distribution z-score table