Probability, Monte Carlo Simulation & Application in Finance – Day 1

1. *Random Hats Problem*

Imagine a group of guests at a party who have all left their hats in a designated area. When they leave, they grab the hats completely at random. How likely it is that at least one guest ends up with their own hat?

a) Let's simulate this process: Create 3 index cards labelled with 3 guest names (Anna, Brian, Colin) and let's assume their original position is ordered alphabetically, that is they placed their hats in alphabetic order. The arrangement A B C would mean all guests grabbed the correct hat.

Interpret the arrangement A C B. How many correct matches does it has?

b) Now randomly shuffle the cards and note the new order, that's the order in which each guests grabbed a hat back. Record the number of guests who grabbed their own hat (a "match")

c) Let's take a vote in class on how many received each number of possible matches. Each student's shuffled order is a 'trial' or 'repetitions'. Determine the proportion of times there are 0 matches, 1 match, and so on, and record these in the "proportion" row of the table above.

No. of Matches =>	0	1	2	3
Count in Class (a)				
Total number of students in class (b)				
Proportion (a/b)				

d) Analyze Proportion row carefully:
   i. Is 1 match a more common outcome than 0 match? Why?

   ii. Is 3 match a common outcome? Why?

   iii. Do we ever get a 2 match? Why?

e) Find: Proportion of 0 * 0 + Proportion of 1 * 1 + Proportion of 2 * 2 + Proportion of 3 * 3. Is it close to 1? Why?

- If a process is random, then individual outcomes are uncertain. People have noticed that for many random processes, the distribution of outcomes has a pattern over many repetitions.
- The probability of any outcome in a random process is the proportion of times that outcome would occur over a large number of repetitions (relative frequency).
- Probability can be approximated by simulating a random process repeatedly and determining the proportion (relative frequency) of occurrences.
- Better approximations come from more repetitions.

Before we get back to the guest picking up their hats, let's learn some vocabulary words.

- The set of all possible outcomes is called the **sample space**.

Example

If a tossed coin lands heads up, we write H. If it lands tails up, we write T. We toss a coin twice. Then the sample space is {HH, HT, TH, TT} which is all the possible outcomes on tossing a coin twice.

Problem (f): We toss a coin three times. Write down the sample space.

- A **probability model** describes all possible outcomes and assigns probabilities to them.

Example

We are tossing a fair coin. Then P(H) = 1/2 and P(T) = 1/2. Please note that a real coin has physical imperfections, so most likely P(H) is not exactly equal to 1/2. This is why we use the word "model". The P(H) = 1/2 is an assumption we make to approximate reality.

- An **event** is a collection of outcomes, i.e., a subset of the sample space.

Example

We toss a coin twice. The event X is: we get at least one head. Then X = {HT, TH, HH}. Note that X is a subset of {HH, HT, TH, TT}.

Problem (g): Find P(X).

- When the occurrence of one event does not affect the occurence of the other events, then the events are called **Independent**.

Example

You toss a coin twice. The outcome of the second toss is independent of the outcome of the first one. Therefore, P(HH) = P(H) times P(H).

Now we are ready to get back to the guest picking up their hats.

h) Theoretical analysis considers all the possible ways to distribute the 3 hats among the 3 guests. Below are all the possible 6 arrangements. Find the number of correct matches for each outcome, i.e., for each arrangement, indicate how many guests receive their correct hat. Determine the (exact theoretical) probability of each outcome by dividing these counts by the number of possible arrangement and recording the results in the last column. Find the sum of probabilities – is it equal to 1?

	# of Match	Probability
ABC		
ACB		
BAC		
BCA		
CAB		
CBA		
		Sum=?

.

i) What is the probability that at least one guest gets their own hat?

j) What is the probability that all guests get their own hats?

k) What is the probability that no guest gets their own hat?

l) Create a bar chart in python with the # of matches on the x-axis and corresponding probability on the y-axis. The chart is called the “Discrete Probability Distribution” for our random variable which is the # of matches. It is called discrete because the # of matches can only take discrete integer values. It cannot take any non-integral values, say 0.5.

m) A probability distribution describes the likelihood of all the possible outcomes for a given event. For example, if we roll a fair six-sided die, we say that the probability distribution is uniform, meaning that each side has an equal chance of coming up. On the other hand, if we roll two six-sided dice, the probability distribution is different, with the sum of ____ being the most likely outcome. Fill in the number in the blank.

2. *Doughnut Prices*

Suppose that a doughnut shop runs a special promotion where they determine the price of a doughnut by rolling a pair of fair six-sided dice. The price in cents is to be the larger number followed by the smaller number. (For example, if you roll a 3 and a 5, the price would be 53 cents.)

a) Before conducting an analysis, make a guess for the probability that you can afford to buy a doughnut if you walk into the shop having only two quarters.

Let's use Python to simulate this random process and determine an approximate probability that you'll be able to afford the doughnut:

```
import numpy as np
import matplotlib.pyplot as plt
from collections import Counter

# Set N to 10,000
N = 10000

# Generate two sets of dice rolls, each ranging from 1 to 6
d1 = np.random.randint(1, 7, N)
d2 = np.random.randint(1, 7, N)

# Calculate price as 10 * max(d1, d2) + min(d1, d2)
price = 10 * np.maximum(d1, d2) + np.minimum(d1, d2)

# Plot histogram of the price
plt.hist(price, bins=range(min(price), max(price) + 2), edgecolor="black")
plt.xlabel("Price")
plt.ylabel("Frequency")
plt.title("Histogram of Price")
plt.show()

# Display frequency table of price
price_counts = Counter(price)
print("Price Frequency Table:")
for k, v in sorted(price_counts.items()):
  print(f"{k}: {v}")
afford = (price<=50)
print("Probability of Affording the doughnut:",sum(afford)/N)
```

b) Describe what each line of this program does.

c) Run this code and report the approximate probability that you can afford the doughnut.

A (rough) estimate of accuracy is that the approximate probability will very likely fall within ± 1/SQRT(N) of the actual probability, where N represents the number of repetitions.

d) Calculate and interpret this accuracy estimate for your simulation results.

e) How could we improve the accuracy of our approximate probability? Do this, and then report the new approximate probability along with the improved measure of its accuracy.

Next, we will determine the exact probabilities involved with this random process.

f) Write out the sample space for this random process.

g) How many outcomes are in the sample space? Is it reasonable to assume that these outcomes are equally likely?

h) Create a discrete probability distribution table for the random variable which is the price of doughnut. Take the sum of probability – is it equal to 1? Create a bar chart using python of price vs probability of occurrence.

Price of Doughnut	Probability
	Sum=?

i) Determine the probability that the price will be 33 cents.

j) Determine the probability that the price will be 32 cents.

k) Determine the probability that the price will be 23 cents.

l) What is the (exact) probability that you can afford the doughnut (when you only have two quarters)?

m) Is this (exact) probability within the error bound of the approximate probability from our simulation analysis?

n) Now suppose that I offer to buy your doughnut if the price turns out to be an odd number. What is the probability that I'll buy your doughnut?

o) What is the probability that you can either afford to buy your own doughnut (assuming that you have two quarters) or that I will buy one for you?

p) Is this probability in (n) equal to the sum of the two individual probabilities? Explain why this makes sense.

3. Pascal's Problem (from Reference 1)

Most of the early work on probability theory revolved around games using dice. Reputedly, Pascal's interest in the field that came to be known as probability theory began when a friend asked him whether or not it would be profitable to bet that within twenty-four rolls of a pair of dice, he would roll a double six. This was considered a hard problem in the mid-17th century. So, what is the probability to get a double six within twenty-four rolls of a pair of dice?

Solution:

- On the first roll the probability of rolling a six on each die is 1/6, so the probability of rolling a six with both dice is 1/36.
- Therefore, the probability of not rolling a double six on the first roll is 1 - 1/36 = 35/36.
- Therefore the probability of not rolling a double six twenty-four consecutive times is (35/36)^24, nearly 0.51, and therefore the probability of rolling a double six is 1 - (35/36)^24, about 0.49.
- So, in the long run it would not be profitable to bet on rolling a double six within twenty-four rolls.

Just to be safe, let's write a little program to simulate Pascal's friend's game and confirm that we get the same answer as Pascal.

```
# Pascal Problem
import random

def rollDie():
  return random.choice([1,2,3,4,5,6])

def checkPascal(numTrials):
  """Assumes numTrials an integer > 0
  Returns an estimate of the probability of winning"""
  numWins = 0.0
  for i in range(numTrials):
    for j in range(24):
      d1 = rollDie()
      d2 = rollDie()
      if d1 == 6 and d2 == 6:
        numWins += 1
        break
  return numWins/numTrials
```

```
# Check the probability of winning
print('Probability of winning =', checkPascal(numTrials=1000000))
Probability of winning = 0.49117
```

After thought 1: The above simulation was for 1 million trials. Repeat the simulation for 1000 trials. Repeat it multiple times with same number of trials, i.e. 1000

- Will the winning probability be same, more or less than the theoretical value we calculated earlier? Why?

After thought 2: Repeat the simulation with higher number of trials: 10,000 and then 100,000

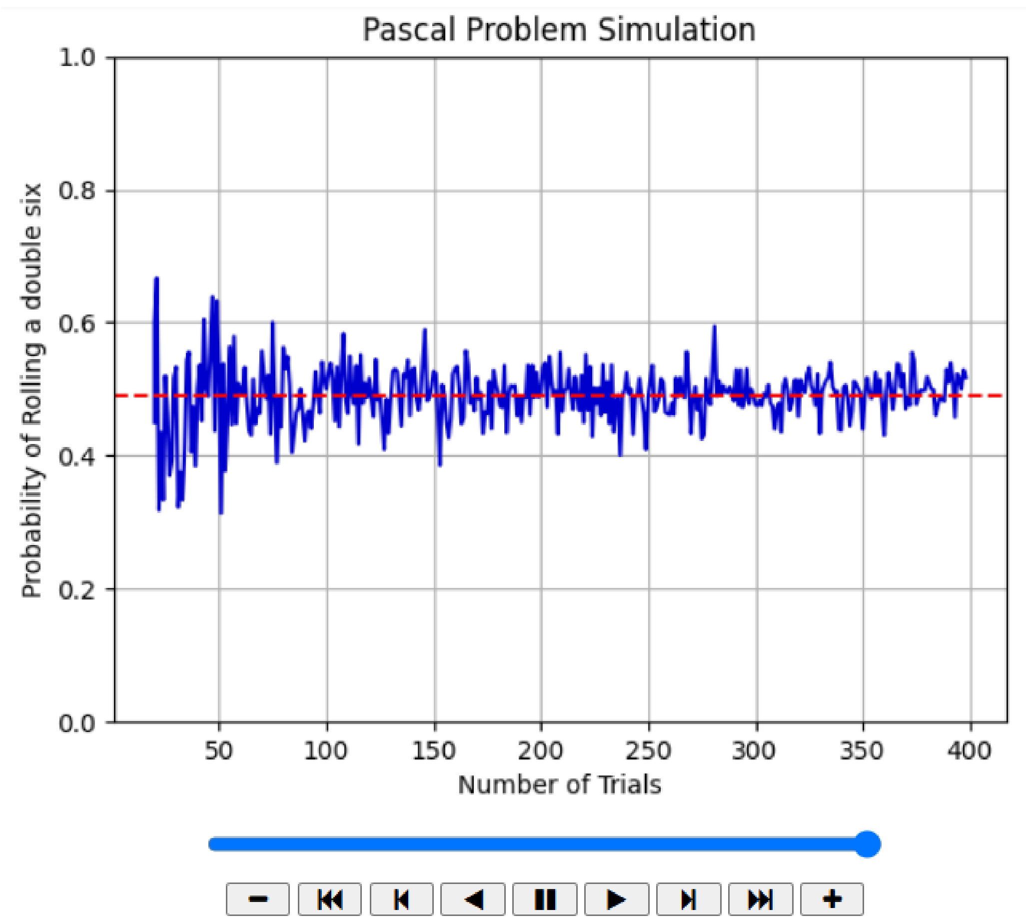
- Will the winning probability increase or decrease as you increase the number of trials? Why?

- Will the winning probability increase or decrease as you decrease the number of trials? Why?

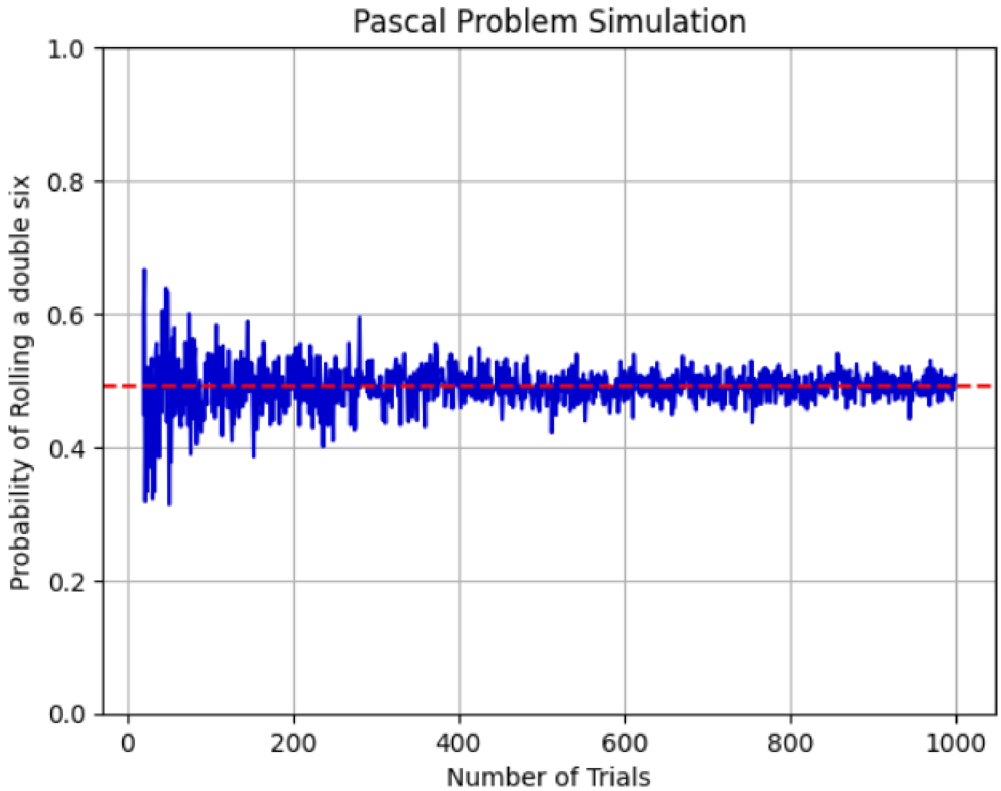
Fun activity: Use below python program to see the animation of how the simulated result varies as number of trials increases

```
# Pascal Problem Animation: As number of trials increases, the simulated
results converge to actual theoretical result
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.animation as animation
plt.rcParams["animation.html"] = "jshtml"
plt.rcParams['animation.embed_limit'] = 2**128

fig = plt.figure()

ax1 = fig.add_subplot(1,1,1)
xs = []; ys = []

no_of_trials=1000
def animate(i):
  ys.append(checkPascal(numTrials=i))
  xs.append(i)
```

```
  ax1.clear()
  ax1.plot(xs, ys,color='mediumblue',label='Monte-Carlo Estimate')
  plt.axhline(y=1-(35/36)**24, color='r', linestyle='--') # Add actual
value to compare with simulated result

  plt.title("Pascal Problem Simulation")
  plt.xlabel("Number of Trials")
  plt.ylabel("Probability of Rolling a double six")
  plt.ylim(0,1)
  plt.grid(True)

ani=animation.FuncAnimation(fig, animate,
frames=np.arange(20,no_of_trials+1))
ani
```



You will observe that the winning probability will converge (move closer) to theoretical value we calculated earlier (denoted by red line in figure below), as the number of trials increase.


Pascal Problem Simulation
Probability of Rolling a double six
Number of Trials
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000

4. Homework

1. Review the Random Babies applet and explain the numbers to your family members: https://www.rossmanchance.com/applets/2021/randombabies/RandomBabies.html

Specify process:

Number of babies 4

Number of trials 1000

Randomize

☑ Animate

Results

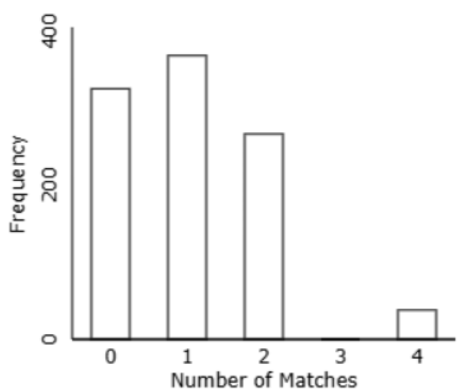
Number of trials: 1000

Cumulative results

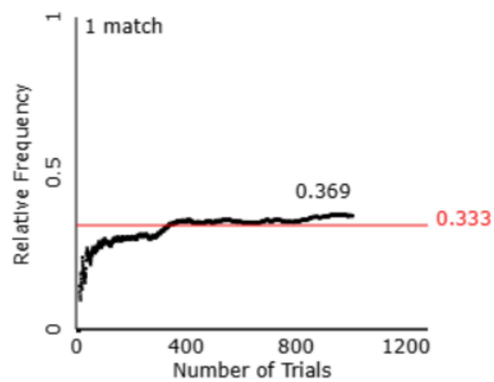
Matches	Count	Prop	Theo
0	326	0.326	0.375
1	369	0.369	0.3333
2	267	0.267	0.25
3	0	0	0
4	38	0.038	0.0417

Most recent
Number of Matches: 1

☑ Show Theoretical



Select display:
○ Average ◉ Relative Frequency

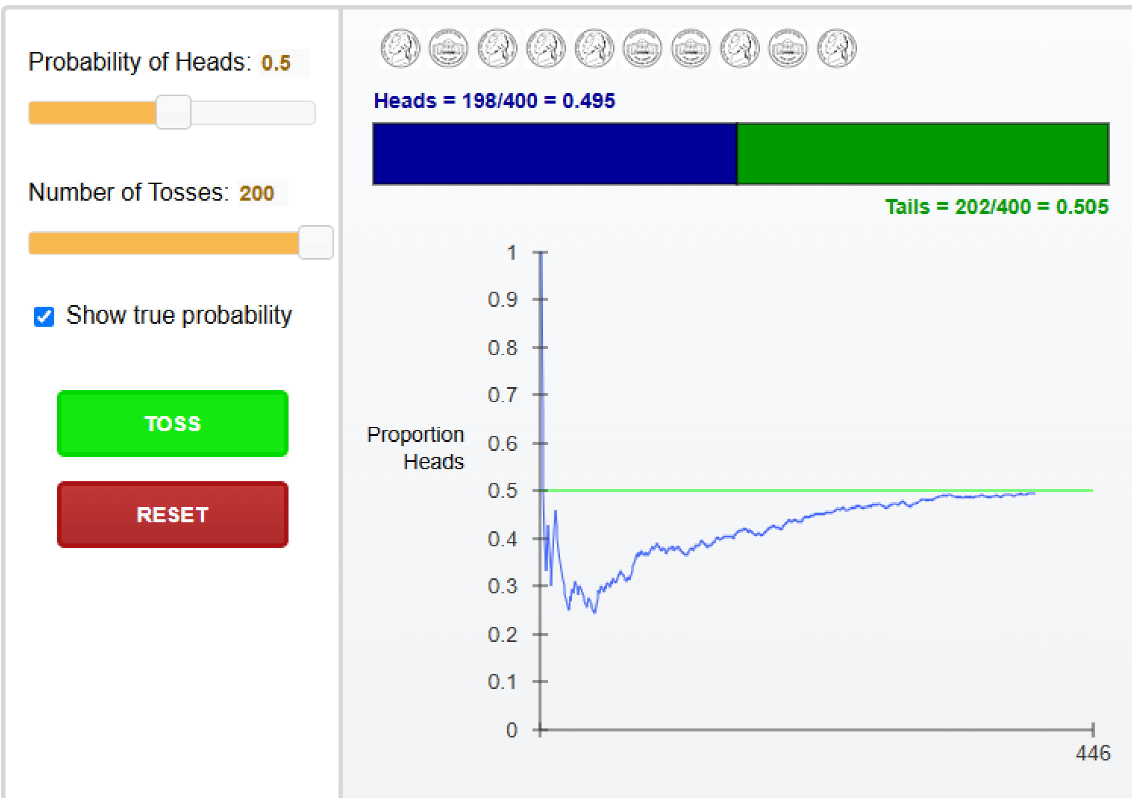


2. Review the Coin Toss applet and explain the concept to your family members: https://digitalfirst.bfwpub.com/stats_applet/stats_applet_10_prob.html

When you toss a coin, there are only two possible outcomes, heads or tails. On any one toss, you will observe one outcome or another —heads or tails. Over a large number of tosses, though, the percentage of heads and tails will come to approximate the true probability of each outcome.

Set the probability of heads (between 0 and 1.0) and the number of tosses, then click "Toss". The outcomes of each toss will be reflected on the graph. Check the box to show a line with the true probability on the graph. Click "Reset" at any time to reset the graph.

In this applet, you can set the true probability of heads for your virtual coin, then toss it any number of times. Notice how the proportion of tosses that produce heads can be quite variable at first, but will eventually settle down to the true probability.



3. **Multiple Choice Tests**: Follow the instructions below (from Reference 2):

Guessing on a Test

Name........ ....

A multiple-choice question with four choices can be answered in four ways: a, b, c, or d. If each choice is equally likely to be picked, the probability of answering the question correctly is ¼. Thus if we guess on, say, 50 such questions, we would expect to answer 12 or 13 of them correctly.

Investigate how well the probability of guessing the correct answer to a four-choice, multiple-choice question agrees with what actually happens when you guess on a multiple-choice test with four choices per question. Take the 50-question final exam for MA 699 QUASI MATHENOMIAL COMPLEXULUS. Since you do not have the exam, just circle the a's, b's, c's, and d's of your choice on the answer sheet below. Then check your answers using the key at the bottom of this page and record your results in the table. Then complete the table and answer the questions below.

PROBABILITY

ANSWER SHEET

1. a b c d	11. a b c d	21. a b c d	31. a b c d	41. a b c d
2. a b c d	12. a b c d	22. a b c d	32. a b c d	42. a b c d
3. a b c d	13. a b c d	23. a b c d	33. a b c d	43. a b c d
4. a b c d	14. a b c d	24. a b c d	34. a b c d	44. a b c d
5. a b c d	15. a b c d	25. a b c d	35. a b c d	45. a b c d
6. a b c d	16. a b c d	26. a b c d	36. a b c d	46. a b c d
7. a b c d	17. a b c d	27. a b c d	37. a b c d	47. a b c d
8. a b c d	18. a b c d	28. a b c d	38. a b c d	48. a b c d
9. a b c d	19. a b c d	29. a b c d	39. a b c d	49. a b c d
10. a b c d	20. a b c d	30. a b c d	40. a b c d	50. a b c d

Outcome	Tally	Fre-quency	Relative frequency	Proba-bility	Difference between relative frequency and probability
Correct				$\frac{1}{4}$	
Incorrect				$\frac{3}{4}$	
	Sum	50	$\frac{50}{50} = 1$	$\frac{4}{4} = 1$	

1. What was your score on the exam?

2. Did anyone in your class score more than 50%? Less than 10%?

3. Which would you recommend for getting a high score on a multiple-choice test, guessing or studying?

4. Charley Brown uses the following strategy for taking a multiple-choice test:

> "Let's see now. In a multiple-choice test, the answer to the first question is almost always C. Then A to sort of balance the C. Then A again to trick you. Then B and D to break the pattern. They never go too long without a D. Then three B's in a row. They always have three B's in a row someplace. Then C and A .... If you're smart, you can pass a multiple-choice test without being smart."

Take the 50-question final exam for MA 699 QUASI MATHEMOMIAL COMPLEXULUS again using Charley Brown's strategy for taking a multiple-choice test and see if you can improve on your score.

Key: caabdbbbca dbcccabdda bcaabbcdda ccbadcabcd badbbddcac

4. **Gambler's Fallacy**: The gambler's fallacy, also known as the Monte Carlo fallacy or the fallacy of the maturity of chances, is the mistaken belief that, if something happens more frequently than normal during some period, it will happen less frequently in the future, or that, if something happens less frequently than normal during some period, it will happen more frequently in the future (presumably as a means of balancing nature).
   a. During a fair coin flipping exercise, we got a sequence of coin flips like this: H H H H T H H H H. What would you expect the next flip to be? What will be the probability that next flip will give a head?

   b. In a local casino, each roulette wheel has an electronic display showing the last ten or so spins. Suppose the roulette wheel stops on same number for 7 times in a row. Imagine you were there when the wheel stopped on the same number for the sixth time. How tempted would you be to make a huge bet on it not coming up to that number on the seventh time?