



# Indian Institute of Technology Mandi भारतीय प्रौद्योगिकी संस्थान मण्डी

## IC252-Data Science 2

### Assignment– 02

---

#### General instructions:

- Utilize Python programming language for implementation.
- Ensure the program is well-documented to enhance comprehension.
- Employ functions and loops for efficient code organization.
- Implement error handling to manage invalid inputs or unexpected scenarios.
- Optimize the code for performance and readability where applicable.

**Question 1:** Suppose there are  $k$  people in a room. Assume each person's birthday is equally likely to be any of the 365 days of the year (we exclude *February 29*), and that people's birthdays are independent. What is the probability that at least one pair of people in the group have the same birthday? Plot the graph of the same as  $K$  grows from 2 to 100

**Question 2:** Stirling's formula is given below is used to approximate the factorial of a given number.

$$n! \sim \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n$$

Write a program to plot the ratio given below as  $n$  grows from 1 to 20.

$$\frac{n!}{\sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n}$$

**Question 3:** Consider a well-shuffled deck of  $n$  cards, labelled 1 through  $n$ . You flip over the cards one by one, saying the numbers 1 through  $n$  as you do so. You win the game if, at some point, the number you say aloud is the same as the number on the card being flipped over (for example, if the 7<sup>th</sup> card in the deck has the label 7). What is the probability of winning? What should be the strategy for choosing  $n$  to maximize your win probability? Show the same using appropriate visualizations. Take  $n \geq 2$

**Question 4:** Develop a Python program to simulate the throwing of two dice and compute the probability distribution of the sum of the numbers obtained on the two dice.

The program should perform the following tasks:

- (a) Calculate the sum of the numbers obtained on the two dice.

- (b) Repeat the process for a substantial number of trials (e.g., 10,000 trials).
- (c) Tabulate the frequency of each possible sum (2 through 12).
- (d) Compute the probability of each sum by dividing its frequency by the total number of trials.
- (e) Display the probability distribution of the sum of the numbers obtained on the two dice.

**Question 5:** Develop a Python program to simulate the Monty Hall problem, a probability puzzle inspired by a game show scenario.

- (a) Simulate a scenario where a contestant is presented with three closed doors.
- (b) Behind one door is a valuable prize, while the other two doors conceal worthless items.
- (c) Initially, the contestant selects one door (without it being opened).
- (d) The game show host, who knows what is behind each door, then opens one of the other two doors that doesn't contain the prize.
- (e) The contestant is given the option to stick with their original choice or switch to the other unopened door.
- (f) Repeat the experiment for a specified number of trials (e.g., 10,000 trials).
- (g) Record whether the contestant wins the prize or not for each trial, based on their final decision (stick or switch).
- (h) Calculate the probability of winning the prize when the contestant sticks with their initial choice and when they switch doors.
- (i) Display the results and probability distributions in a clear and organized format.