深 圳 大 学 实 验 报 告

课程名称：概率论与数理统计

实验项目名称：**Conditional Probability and Bayes's Rule**

学院：电子与信息工程学院

专业：电子信息工程

指导教师：陈昌盛老师

报告人：贾苏健 学号：**2022280485**

班级：**06**

实验时间：**2023.10.21**

实验报告提交时间：**2023.10.26**

教务处制

Aim of Experiment:

1)Dice Rolls:

Randomness Principle: This experiment is aimed at simulating the results of rolling dice. Each roll of the dice is inherently random. This randomness can be simulated using a pseudorandom number generator, where the generation of pseudorandom numbers is based on a seed value to ensure reproducibility.

Probability Distribution: The result of each dice roll is a discrete random event with a uniform probability distribution, where each face has an equal probability of 1/6. This is because each face of a fair six-sided die has an equal chance of appearing.

2)Conditional Probability and Baye's Rule:

Conditional Probability Principle: Conditional probability refers to the probability of one event occurring given that another event has occurred. In this context, we consider the conditional probability of selecting a ball from two urns (Urn A and Urn B). Calculating conditional probability involves using known information to infer the probability of events.

Bayes' Theorem: Bayes' theorem is an important tool for calculating conditional probabilities. It describes how to calculate the conditional probability of an event given some known information. In

this experiment, Bayes' theorem is used to calculate the conditional probability of selecting a white ball from Urn A, given that the ball is white.

3)Conditional probability analysis for math scores:

Principles of Data Analysis: In this section, we analyze a dataset of actual student math scores. We focus on two attributes: students' math scores (G3) and their study time (studytime). We use the dataset to estimate the probability of achieving high scores (greater than or equal to 15) in different study time intervals.

Probability Estimation: We estimate various probabilities using statistical information from the dataset, such as the distribution of students' study times, the probability of achieving high scores, and the probability of achieving high scores in different study time intervals. This helps us understand the relationship between students' study habits and their grades.

Experiment Content:

1)Dice Rolls:

This experiment aims to simulate random events, particularly the outcomes of dice rolls. Using Python programming, we generate pseudorandom numbers to simulate dice throws and then create a probability distribution plot of the experimental data.

2)Conditional Probability and Baye's Rule:

This experiment focuses on the principles of conditional probability and the application of Bayes' Theorem. It considers scenarios where balls are chosen from two different urns and then writes a function to calculate the conditional probability of a ball coming from a specific urn when it is known to be white.

3)Conditional probability analysis for math scores:

This experiment uses a real dataset of student math scores to analyze the conditional probability relationship between students' math scores and their study time. We calculate the probability of achieving high scores in different study time intervals and explore the relationship between students' study time and high scores.

Experiment Process：

1)Dice Rolls:

Setting a Random Seed: Firstly, by setting the seed value for the random number generator.

Simulating Dice Rolls: Simulating the roll of a fair six-sided die using random number generation functions in Python. Each simulation produces a random dice outcome.

Plotting Probability Distribution: Using the Matplotlib library to

create a probability distribution plot of the results from the dice rolling experiment. The plot also includes a baseline for theoretical probabilities.

2)Conditional Probability and Baye's Rule:

Conditional Probability Calculation: Writing a `conditional_probability(rA, wA, rB, wB)` function that calculates the conditional probability of a ball coming from urn A based on known conditions.

Application of Bayes' Theorem: The `conditional_probability(rA, wA, rB, wB)` function applies Bayes' Theorem to calculate conditional probability.

3)Conditional probability analysis for math scores:

Importing Student Data: Importing a dataset of student math scores using the Pandas library, which includes attributes such as students' scores (G3) and study time (studytime).

Probability Estimation: Estimating the probability of achieving high scores (greater than or equal to 15) in different study time intervals and exploring the conditional probability relationship between students' study time and high scores by utilizing information from the dataset.

Plotting Probability Distribution: Creating a probability distribution plot using the Matplotlib library to visualize the conditional probability relationship between students' study time and high scores.

Data Logging and Processing:

1)Dice Rolls:

Data Recording: In the dice rolling simulation, at the beginning of the experiment, the outcomes of each simulated dice roll are recorded.

```
# Simulate n die rolls
die_rolls = np.random.randint(1, 7, n)
```

Figure (1) captures random experimental data.

Data Processing: By recording the results of dice rolls, data processing may involve calculating the frequency and probability distribution of the dice outcomes.

```
# Count the occurrence of each face
Count = np.zeros((6, n))
for i in range(6):
    Count[i] = (die_rolls == i + 1).cumsum()

# plot the empirical values
for i in range(6):
    Prob = Count[i,1:]/np.arange(1,n)
```

```
# Count the occurrence of even numbers (2, 4, 6)
even_count = np.sum(die_rolls % 2 == 0)

# Calculate the empirical probability of even numbers
Prob_E = np.cumsum(die_rolls % 2 == 0) / np.arange(1, n + 1)
```

Figure (2) calculates the probability of dice rolls.

2)Conditional Probability and Baye's Rule:

Data Recording: In this section, data recording includes the number of red and white balls and the color of the chosen ball under known conditions.

```
rA, wA, rB, wB = 1., 2., 2., 1.
```

Figure (3) Probability of Red and White Balls in Urns A and B

```
# Calculate the probability of selecting Bag A
prob_A = 1 / 2  # Assuming equal probability of choosing Bag A or Bag B
```

Figure (4) Probability of Drawing from Urn A

Data Processing: Data processing occurs within the conditional_probability function, which calculates conditional probabilities.

```
# Calculate the probability of selecting a white ball from Bag A
prob_white_given_A = wA / (rA + wA)

# Calculate the probability of selecting a white ball from Bag B
prob_white_given_B = wB / (rB + wB)

# Use Bayes' rule to calculate the conditional probability
conditional_prob_A_given_white = (prob_A * prob_white_given_A) / ((prob_A * prob_white_given_A) + ((1 - prob_A) * prob_white_given_B))
```

Figure (5) Calculates Conditional Probability.

3)Conditional probability analysis for math scores:

Data Recording: In this section, data recording involves importing a dataset of student scores and study times.

```
data_math = pd.read_csv("./dataset/student-mat.csv")
```

```
attributes = ["G3", "studytime"]
data_math = data_math[attributes]
```

Figure (6) Retrieves Data from the Dataset.

Data Processing: Data processing includes statistical analysis of the dataset, such as estimating the probability of achieving high scores in different study time intervals. This also includes calculating the conditional probability relationship between students' study time and high scores to understand the probability relationship between these two variables.

$$P(\text{study interval}) = \frac{\text{Number of students with this study interval}}{\textit{Total number of students}}$$

Figure (7) Formula for Calculating the Proportion of Different Study Durations to the Total Number of Students.

```
data_temp = data_math["studytime"].value_counts()
P_studytime = pd.DataFrame((data_temp/data_temp.sum()).sort_index())
P_studytime.index = ["< 2 hours", "2 to 5 hours", "5 to 10 hours", "> 10 hours
P_studytime.columns = ["Probability"]
P_studytime.columns.name = "Study Interval"
```

Figure (8) Process for Calculating the Proportion of Different Study Durations to the Total Number of Students.

$$P(\text{high score}) = \frac{\text{Number of students with high scores}}{\text{Total number of students}}$$

Figure (9) Formula for Calculating the Proportion of High-Scoring and Low-Scoring Students to the Total Number of Students.

```
data_temp = (data_math["G3"] >= 15).value_counts()
P_score15_p = pd.DataFrame(data_temp/data_temp.sum())
P_score15_p.index = ["Low", "High"]
P_score15_p.columns = ["Probability"]
P_score15_p.columns.name = "Score"
```

Figure (10) Process for Calculating the Proportion of High-Scoring and Low-Scoring Students to the Total Number of Students.

$$P(\text{study interval} \mid \text{highscore}) = \frac{\text{Number of students with study interval AND highscore}}{\text{Total number of students with highscore}}$$

Figure (11) Formula for Calculating the Proportion of Different Study Durations to High-Scoring Students.

```
score = 15
data_temp = data_math.loc[data_math["G3"] >= score, "studytime"]
P_T_given_score15 = pd.DataFrame((data_temp.value_counts()/data_temp.shape[0]).sort_index())
P_T_given_score15.index = ["< 2 hours", "2 to 5 hours", "5 to 10 hours", "> 10 hours"]
P_T_given_score15.columns = ["Probability"]
```

Figure (12) Process for Calculating the Proportion of Different Study Durations to High-Scoring Students.

$$P(\text{student getting a highscore} \mid \text{study interval}) = \frac{P(\text{study interval} \mid \text{the student scored high})P(\text{highscore})}{P(\text{study interval})}$$

Figure (13) Formula for Calculating the Proportion of High-Scoring Students among Different Study Durations.

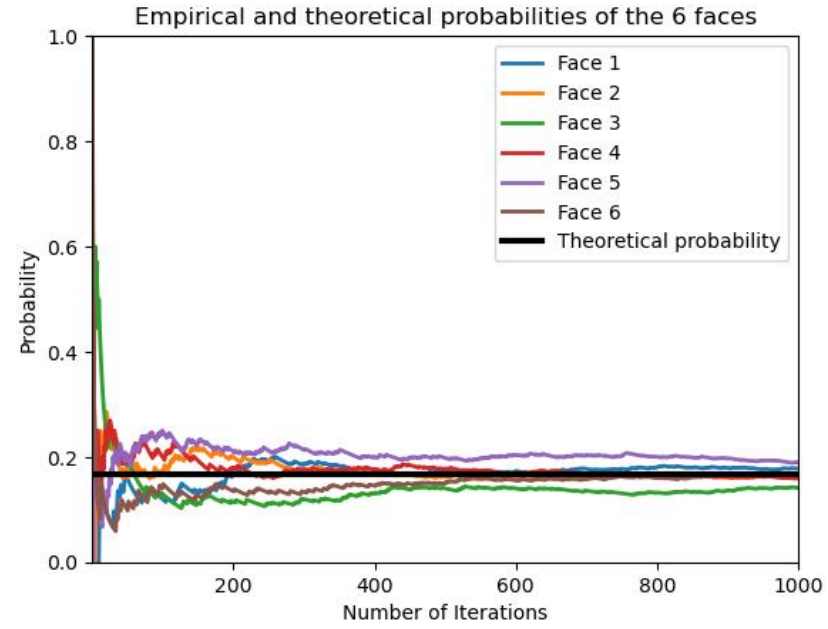
```
P_score15_given_T_p = P_T_given_score15 * P_score15_p.loc["High"] / P_studytime
```

Figure (14) Process for Calculating the Proportion of High-Scoring Students among Different Study Durations.

Experimental Results and Analysis:

1)Dice Rolls:

Experiment Results: By simulating the rolling of dice multiple times, the results of each roll were recorded.



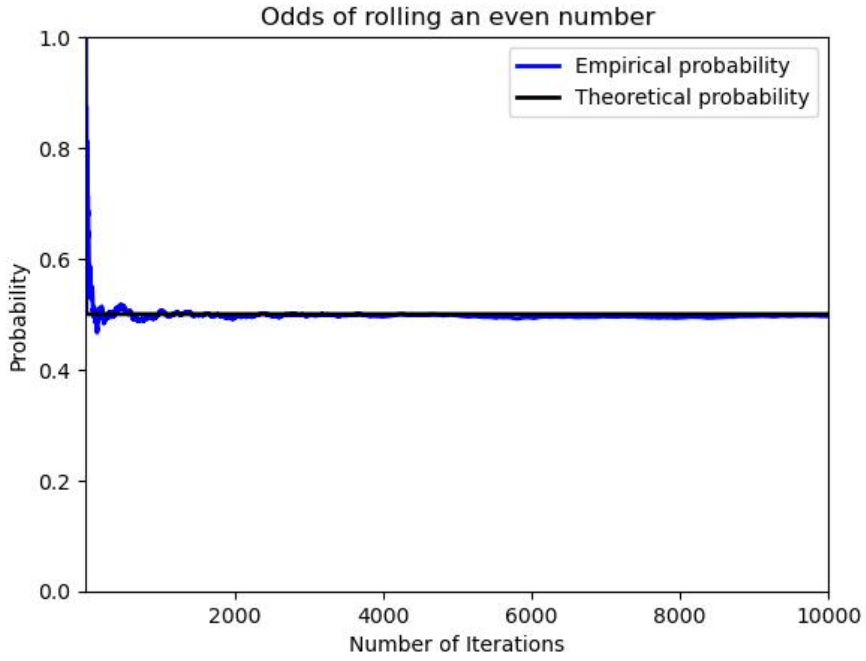


Figure (15) Experiment Results for Dice Rolls.(n=1000)

Analysis: By analyzing the experimental data, it is possible to calculate the frequency and probability distribution of each possible outcome. The experimental results show that for a fair six-sided die, the probability of each outcome is uniform, being 1/6.

2)Conditional Probability and Baye's Rule:

Experiment Results: The experiment results are numerical values of conditional probabilities.

```
rA, wA, rB, wB = 1., 2., 2., 1.
conditional_probability(rA, wA, rB, wB)
```

Output

```
0.6666666666666666
```

Figure (16) Experiment Results for Conditional Probability and Bayes' Rule.

Analysis: The results indicate that, given certain known conditions, Bayes' theorem can be used to calculate conditional probabilities.

3)Conditional probability analysis for math scores:

Experiment Results: The experiment results involve the conditional probability relationship between students' mathematical scores and study time. Specifically, the experiment results include the probability of achieving high scores (greater than or equal to 15 points)

within different study time intervals, as well as the conditional probability between students' study time and achieving high scores.

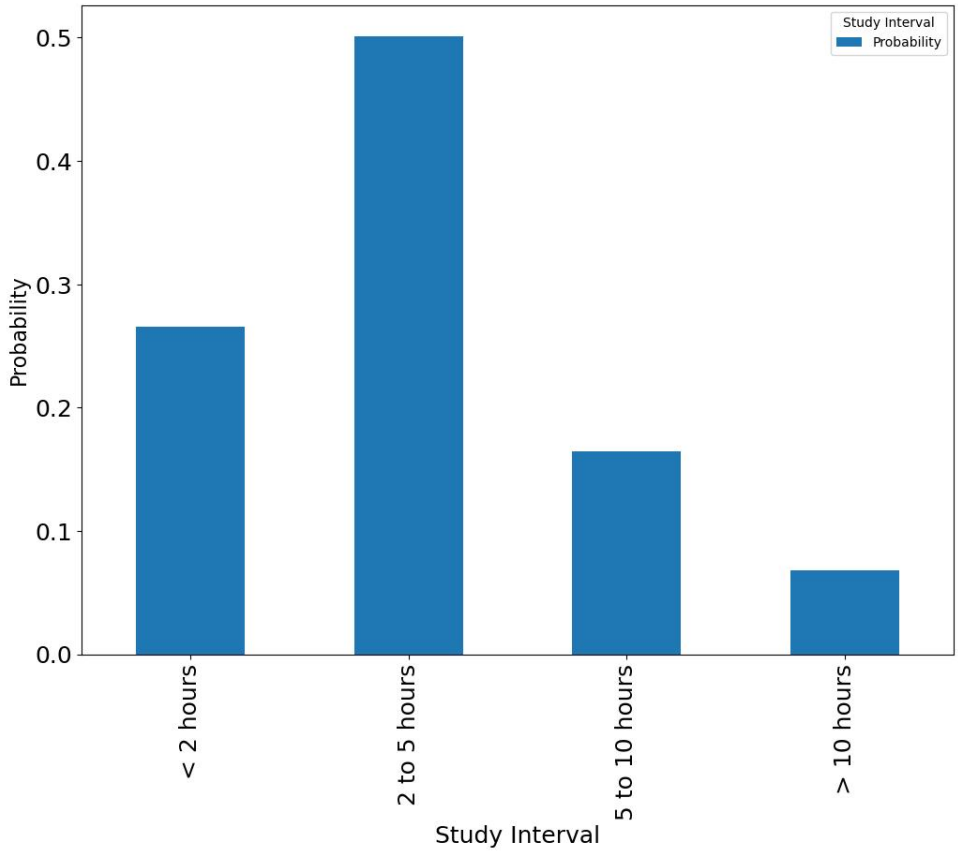


Figure (17) Proportion of Students with Different Study Durations to Total Student Population.

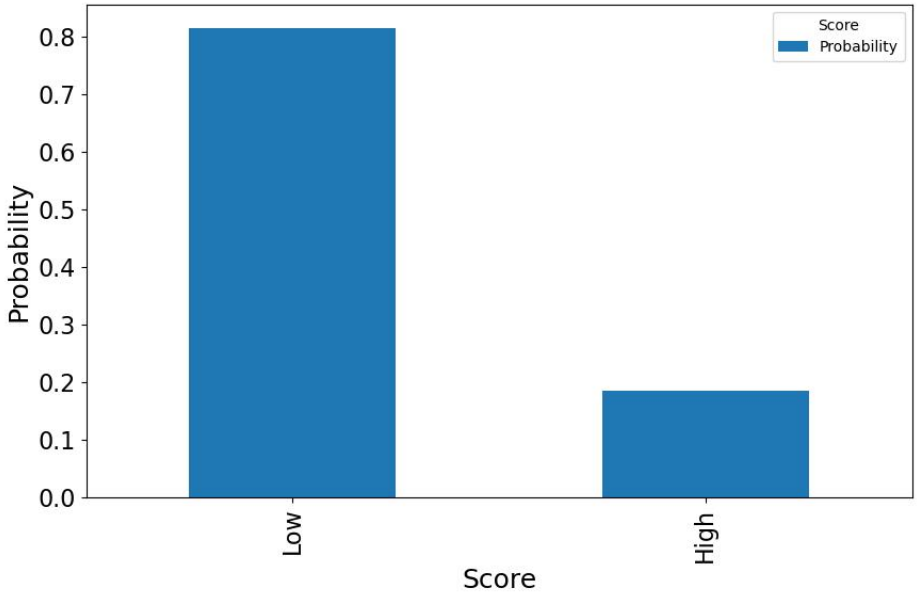


Figure (18) Proportion of High-Scoring and Low-Scoring Students to Total Student Population.

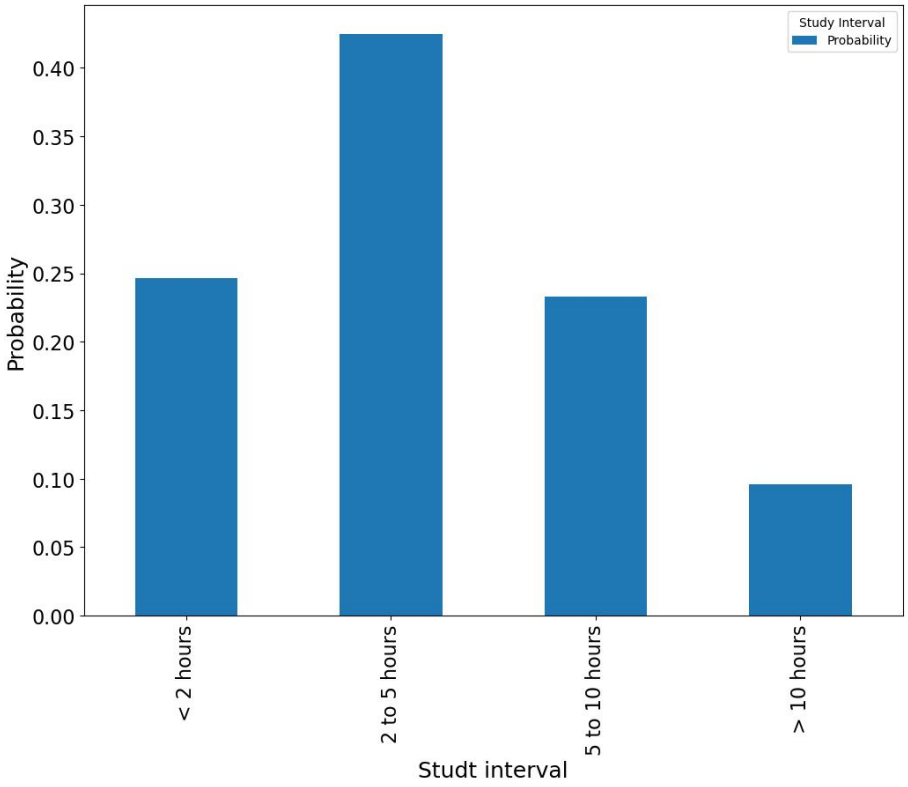


Figure (19) Proportion of Students with Different Study Durations to High-Scoring Students.

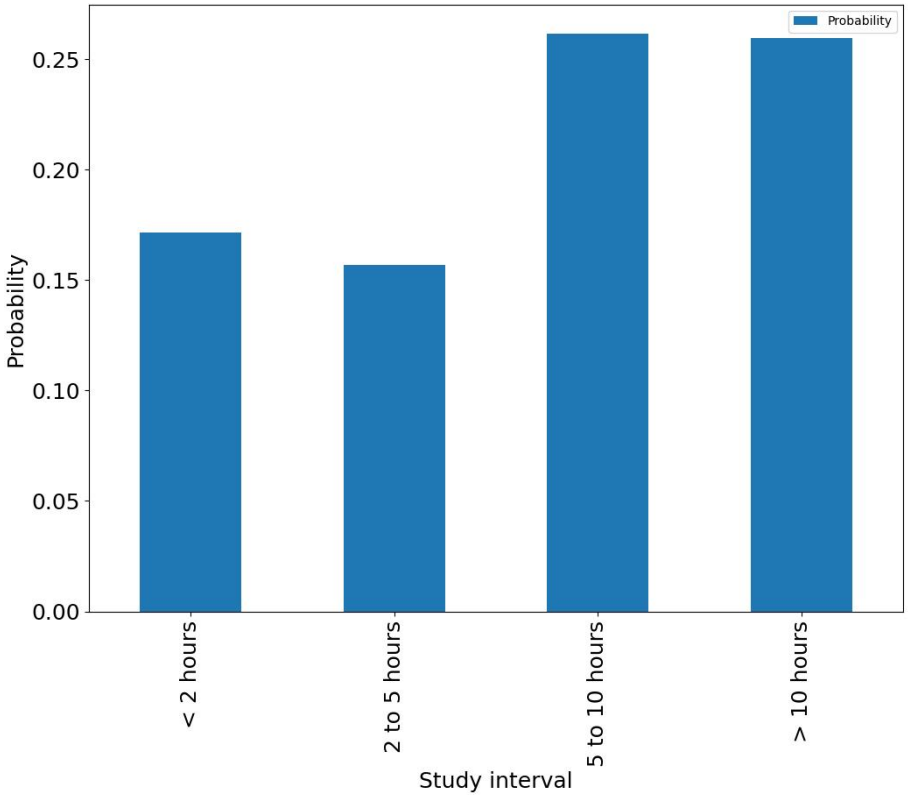


Figure (20) Proportion of High-Scoring Students to Students with Different Study Durations.

Analysis: Through the analysis of real data, it is possible to understand the probability relationship between students' study time and achieving high scores. The analysis results may suggest a positive correlation between longer study times and achieving high scores.

指导教师批阅意见：

成绩评定：

指导教师签字：

年　　月　　日

备注：