**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 Rule in Python*

**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**

**报告人： 杨烨 学号**： **2022280380**

**班级： 06**

**实验时间： 2023年10月19日**

**实验报告提交时间： 2023年10月26日**

**教务处制**

|  |
| --- |
| Aim of Experiment:   1. Master the basic of Python. Such as cyclic structure, judgment structure. 2. Master the use of some functions in pandas library Numpy, matplotlib library. 3. Develop a deeper understanding of Conditional Probability and Bayes Rule. |
| Experiment Content:  There are three parts in this experiment:  ①The first part is a large number of independent repeated experiments to estimate the probability of dice:  This section contains two modules:  The first module explores the probability of six rolls on a die;  The second module explores the probability of rolling an even number of dice;  To see if dice are fair;  ②The second part is to establish a conditional probability calculation formula for the ball-touching model;  ③The third part is to explore the correlation between students' study time and math achievement |
| Experiment Process：  The first part is a large number of independent repeated experiments to estimate the probability of dice:  The first module explores the probability of six rolls on a die;  def probability\_plot(n):  """  input: n (generate n random die rolls)  output: Count (counting the occurance of each event )  """  # Generate n random die rolls  faces = 6 # 每个骰子具有六个面  rolls = np.random.randint(1, faces+1, size=n) # 返回n个1至6的随机整数组成的一维数组来储存n次骰子的结果  # Count the occurrence of each event  b = 0  Count = np.zeros((6,n+1)) # 设置一个初值全为零的6行，n+1列的二维数组以便储存统计结果    for y in range(n): # 遍历rolls数组  if rolls[b]==1:  Count[0][b+1:n+1]+=1 # 若第一次投掷结果为1，则从第b+1个元素开始往后的每一个元素都加1次  elif rolls[b]==2:  Count[1][b+1:n+1]+=1  elif rolls[b]==3:  Count[2][b+1:n+1]+=1  elif rolls[b]==4:  Count[3][b+1:n+1]+=1  elif rolls[b]==5:  Count[4][b+1:n+1]+=1  elif rolls[b]==6:  Count[5][b+1:n+1]+=1  b += 1 # 变量b每次都加1以遍历  # plot the empirical values  for i in range(6):  Prob = Count[i,1:]/np.arange(1,n+1) #对这个二维数组的第i行除以一个从1递增到n的整数，得到每一次各结果的概率  plt.plot(np.arange(1, n + 1), Prob, linewidth=2.0, label='Face '+str(i+1))  The second module explores the probability of rolling an even number of dice;  def probability\_plot\_B(n):  """  input: n (generate n random die rolls)  output: Prob\_E (calculating the probability of even throw's) # even 偶数  hint: counting the events of even numbers  """  faces = 6 #每个骰子具有六个面  rolls = np.random.randint(1, faces+1, size = n) #返回n个1至6的随机整数组成的一维数组  times\_E = np.zeros(n)  b = 0    for y in range(n):  if rolls[b]%2 == 0: # 判断投掷结果是否为偶数  times\_E[b:n] += 1 # 对结果进行统计  else:  pass # 若为奇数则跳过  b += 1 变量b每次都加1以遍历    Prob\_E = times\_E / np.arange(1,n+1） # 计算每一次投掷之后，对此之前产生的结果为偶数的概率  The second part is to establish a conditional probability calculation formula for the ball-touching model;  def conditional\_\_probability(rA, wA, rB, wB):  # inputs: all of them are of type 'float'  # output: a variable of type 'float'  white\_A\_probability = 0.5\*(wA/(rA + wA))  # 选择到A缸且在A缸中摸到白球的概率  white\_probability = 0.5\*(wA/(rA + wA)) + 0.5\*(wB/(rB + wB))  # 摸到白球的概率  conditional\_probability = white\_A\_probability/white\_probability  # 已知摸到的球是白球，抽到的缸是A缸的概率    return conditional\_probability # 返回所求条件概率  #  The third part is to explore the correlation between students' study time and math achievement  import pandas as pd  import matplotlib.pyplot as plt  data\_mat = pd.read\_csv (r"C:\Users\HP\Desktop\大二上\概率论与数理统计\实验\Experiment 2\dataset\student-mat.csv")  #从名为 "student-mat.csv" 的 CSV 文件中读取数据 并将其存储到名为 data\_math 的 DataFrame  attributes = ["G3","studytime"]  data\_mat = data\_mat[attributes]  # 选择 data\_math DataFrame 中的两列数据，分别是 "G3"（考试分数）和 "studytime"（学习时间），将选取的数据存储到名为 data\_por 的新 DataFrame 中  data\_temp = data\_mat["studytime"].value\_counts() # 计算 "studytime" 列中各个取值的频数  P\_studytime = pd.DataFrame((data\_temp/data\_temp.sum()).sort\_index()) # 将频数转换为概率  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"  P\_studytime.plot.bar(figsize=(12,9),fontsize=18) # 绘制学习时间区间的概率条形图  plt.ylabel("Probability",fontsize=16)  plt.xlabel("Study Interval",fontsize=18)  data\_temp = (data\_mat["G3"]>=15).value\_counts() # 计算考试分数大于等于15的学生人数  P\_score15\_p = pd.DataFrame(data\_temp/data\_temp.sum()) # 将人数转换为概率  P\_score15\_p.index = ["Low","High"]  P\_score15\_p.columns = ["Probability"]  P\_score15\_p.columns.name = "Score"  print(P\_score15\_p)  P\_score15\_p.plot.bar(figsize=(10,6),fontsize=16) # 绘制考试分数的概率条形图  plt.xlabel("Score",fontsize=18)  plt.ylabel("Probability",fontsize=18)  score = 15  data\_temp = data\_mat.loc[data\_mat["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"]  print("Probability of study interval given that the student gets a highscore:")  P\_T\_given\_score15.columns.name="Study Interval"  P\_T\_given\_score15.plot.bar(figsize=(12,9),fontsize=16)  plt.xlabel("Studt interval",fontsize=18)  plt.ylabel("Probability",fontsize=18)  P\_score15\_given\_T\_p = P\_T\_given\_score15 \* P\_score15\_p.loc["High"] / P\_studytime  print("Probability of high score given study interval :")  pd.DataFrame(P\_score15\_given\_T\_p).plot.bar(figsize=(12,9),fontsize=18).legend(loc="best")  plt.xlabel("Study interval",fontsize=18)  plt.ylabel("Probability",fontsize=18) |
| Data Logging and Processing:  **In the first experiment named “Die Rolls”：**  The first module：  I obtained the frequency of each number of points in n dice throwing experiments. When n approaches infinity, the frequency of each number of points is close to 1/6 of the theoretical value. The figure below shows the statistical results of this experiment：    Figure 1- Probability curve for each point  The second module：  I have obtained the frequency of even number of points in n dice throwing experiments. When n approaches infinity, the frequency of even number of points is close to 1/2 of the theoretical value. The figure below shows the statistical results of this experiment：    Figure 2- The probability curve of even number of dots  **In the second experiment named**  **“Conditional Probability and Baye's Rule”**    Figure 3- Conditional probability calculation results correct window  I successfully calculated the conditional probability value of the given experimental data;  **In the third experiment named**  **“Conditional probability analysis for math scores”:** Based on the survey of students' daily study intervals, I sorted out their daily study time and drew the bar chart as follows:   Figure 4- Bar chart of student learning interval distribution  Among them, the study time of 2 hours to 5 hours is the most, and the study time of more than 10 hours is the least;  At the same time, I also plotted a bar chart of their probability of getting high and low grades:    Figure 5- Student achievement distribution bar chart  The following table calculates the probability that a student will fall in each interval of his study time per day if he achieves high grades:    Figure 6- Bar chart of learning interval distribution for high-achieving students  The table below is a calculation that calculates the probability of a student getting a high grade based on the student's daily study time:    Figure 7- Histogram of the distribution of the number of  high-achieving students for each learning interval |
| Experimental Results and Analysis:  Experimental Results are as below：  In this experiment, I successfully implemented the code for three tasks,  First, we simulate a model that flips a coin n times and calculate the probability by counting the number of flips.  The calculated probability value of each point is close to 1/6 with the increase of experiment times, and the probability value of even points is close to 1/2 with the increase of experiment times.  Secondly, the conditional probability of the given two problems is calculated using Bayesian formula.  Finally, the correlation between student achievement and study time is obtained by using the statistical table of student achievement and study time given.  Analysis are as below：  In this experiment, I learned to simulate a coin toss experiment using Python library functions, and store the results using numpy matrix and perform related calculations. I'm also good at drawing probabilistic images using matplotlib.  Experimental expansion suggestions:  I think this experiment can expand new aspects, such as statistical inference of data:Statistical inference is a method to infer the overall data based on the sample data, including point estimation, interval estimation and hypothesis testing, etc. It can also supplement the calculation of the mean, median, mode and variance of the data |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  年 月 日 |
| 备注： |

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。

2、教师批改学生实验报告时间应在学生提交实验报告时间后10日内。