**CS 513: Knowledge Discovery in Databases**

**Probability Assignment**

Zijing Huang 10414952

**Homework 1.1**

**Juan is playing the following game: he rolls two dice. If they sum up to 7 he loses a dollar. If they sum up to 2, he wins 2 dollars. Otherwise, he doesn’t win nor lose.**

**After playing this game for a long time, what shall happen? why?**

He will lose all his money. There are 6\*6 = 36 possible combinations. There are 6 matching pairs that add up to 7, which is [1,6], [2,5], [3,4], [4,3], [5,2], [6,1]. And there are only one possible pairs that add up to 2, which is [1,1]. Therefore, the possibility of losing 1 dollars would be 6/36 = 1/6 and the possibility of winning 2 dollars would be 1/36. The average win/loss per roll will be 1/6\*(-1)+1/36\*2 = -(1/9) dollars. So, Juan will lose all his money in the end.

**Homework 1.2**

**Jerry and Susan have a joint bank account.**

**Jerry goes to the bank 20% of the days.**

**Susan goes there 30% of the days.**

**Together they are at the bank 8% of the days. ( Implied that two events are dependent. If they are independent events, the possibility that they are both at bank is 6%)**

**Susan was at the bank last Monday. What’s the probability that Jerry was there too?**

8% / 30% = 26.667%

**Last Friday, Susan wasn’t at the bank. What’s the probability that Jerry was there?**

(20%-8%) / (100%-30%) = 17.143%

**Last Wednesday at least one of them was at the bank. What is the probability that both of them were there?**

8% / (100% – 58%) = 19.048%

**Homework 1.3**

**Harold and Sharon are studying for a test.**

**Harold’s chances of getting a “B” are 80%.**

**Sharon’s chances of getting a “B” are 90%.**

**The probability of at least one of them getting a “B” is 91%.( Implied that two events are dependent. If they are independent events, the probability of at least one of them getting a “B” would be 98%)**

The probability of they all getting a B is as following:

80% + 90% - 91% = 79%

1. **What is the probability that only Harold gets a “B”?**

80% - 79% = 1%

1. **What is the probability that only Sharon gets a “B”?**

90% - 79% = 11%

1. **What is the probability that both won’t get a “B”**

1 – 91% = 9%

**Homework 1.4**

**Jerry and Susan have a joint bank account. Jerry goes to the bank 20% of the days. Susan goes there 30% of the days. Together they are at the bank 8% of the days.**

**Are the events “Jerry is at the bank” and “Susan is at the bank” independent?**

NO, because if they are independent events, the probability of together at the bank will be 6%.

**Homework 1.5**

**You roll 2 dice.**

**Are the events “the sum is 6” and “the second die shows 5” independent?**

P(sum=6) = 5/36

P(second\_die=5) = 6/36 = 1/6

P(sum=6 and second\_die=5) = 1/36 ≠ P(sum=6) \* P(second\_die=5)

Therefore, they are dependent events.

**Are the events “the sum is 7” and “the first die shows 5” independent?**

P(sum=7) = 6/36 = 1/6

P(first\_die=5) = 6/36 = 1/6

P(sum=7 and first\_die=5) = 1/36 = P(sum=7) \* P(first\_die=5)

Therefore, they are independent events.

**Homework 1.6**

**An oil company is considering drilling in either TX, AK and NJ. The company may operate in only one state. There is 60% chance the company will choose TX and 10% chance - NJ. There is 30% chance of finding oil in TX, 20% - in AK, and 10% - in NJ.**

1. **What’s the probability of finding oil?**

P(oil and TX) = 30% \* 60% = 18%

P(oil and AK) = 20% \* 30% = 6%

P(oil and NJ) = 10% \* 10% = 1%

P(oil) = P(oil and TX) + P(oil and AK) + P(oil and NJ) = 25%

**2.The company decided to drill and found oil. What is the probability that they drilled in TX?**

P(TX|oil) = P(oil and TX) / P(oil) = 18% / 25% = 72%

**Homework 1.7**

**Using the Titanic dataset“Titanic\_row.csv”in CANVAS answer the following questions.**

**What is the probability that a passenger did not survive?**

817/1316 = 62.082%

**What is the probability that a passenger was staying in the first class?**

325/1316 = 24.696%

**What is the probability that a passenger survived and the passenger was a female?**

324/1316 = 24.620%

**Given that a passenger survived, what is the probability that the passenger was a female?**

324/499 = 64.929%

**Are survival and staying in the first class independent?**

P(Survived) = 711/2201 = 32.303%

P(FirstClass) = 325/2201 = 14.766%

P(FirstClass & Survived) = 203/325 = 62.462%

If they are independent, P(F & S) should equals P(S) \* P(F), but it isn’t. So they are not independent event.

**Given that a passenger survived, what is the probability that the passenger was staying in the first class and the passenger was a female?**

141/499 = 28.257%

**Given that a passenger survived, what is the probability that the passenger was a female?**

324/499 = 64.929%

**Given that a passenger survived, what is the probability that the passenger was staying in the first class?**

203/499 = 40.681%

**Given that a passenger survived, are being female and staying in the first class independent?**

P(Female|S) = 324/499 = 64.929%

P(FirstClass|S) = 203/499 = 40.681%

P(Female & FirstClass|S) = 141/499 = 28.257%

Because P(Female|S)\* P(FirstClass|S)≠ P(Female & FirstClass|S ),

They are dependent .