CS 513 Knowledge Discovery & Data Mining
Vidya Sagar Polaki
CWID 10430970

Homework 1.1

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.

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

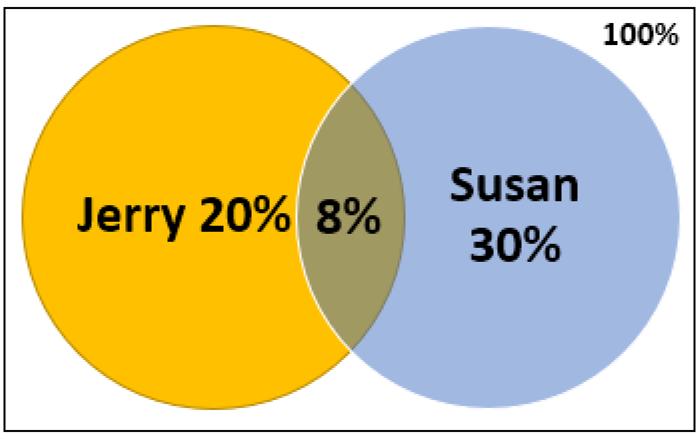
b. Last Friday, Susan wasn't at the bank. What's the probability that Jerry was there?

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

Solution:

Given, 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.

We can draw the following Venn diagram from the above information



	Susan at Bank	Susan not at Bank
Jerry at Bank	8%	12%
Jerry not at Bank	22%	58%

From the above table,

P (Jerry ∩ Susan) = 8%

P (Jerry — Susan) = P (Jerry ∩ Susan′)= 12%

P (Susan — Jerry) = 22%

P (Jerry ∪ Susan) = P (Jerry) + P (Susan) - P (Jerry ∩ Susan) = 20+30-8 = 42%

a. Probability of Jerry going to bank when Susan is already in bank is

$$\text{P (Jerry | Susan)} = \frac{\text{P (Jerry} \cap \text{Susan)}}{\text{P (Susan)}} = 8/30 = \mathbf{26.66\%.}$$

b. Probability of Jerry going to bank when Susan is not there in bank is

$$\text{P (Jerry | Susan)} = \frac{\text{P (Jerry} \cap \text{Susan}')}{\text{P (Susan}')} = 12/70 = \mathbf{17.14\%}$$

c. The probability that both being Jerry and Susan at bank when at least one of them was at the bank is

$$\text{P (Jerry | Susan)} = \frac{\text{P ((Jerry} \cap \text{Susan)} \cap \text{(Jerry} \cap \text{Susan))}}{\text{P (Jerry} \cup \text{Susan)}} = \frac{\text{P (Jerry} \cap \text{Susan)}}{\text{P (Jerry} \cup \text{Susan)}} = 8/42 = \mathbf{19.04\%}$$

Homework 1.2

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%.

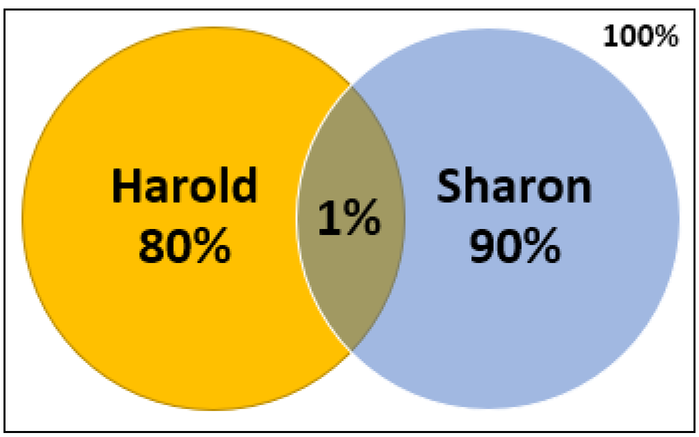
a. What is the probability that only Harold gets a "B"?

b. What is the probability that only Sharon gets a "B"?

c. What is the probability that both won't get a "B"?

Solution:

Given, 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%.



P (Harold) = 80%

P (Sharon) = 90%

P (Harold ∪ Sharon) = 91%

P (Harold ∩ Sharon) = P (Harold) + P (Sharon) - P (Harold ∪ Sharon) = 80+90-91 = 79%

a. Probability that only Harold gets a "B" is
   P (only Harold) = P (Harold) - P (Harold ∩ Sharon) = 80 – 79 = 1 %.
b. Probability that only Sharon gets a "B" is
   P (only Sharon) = P (Sharon) - P (Harold ∩ Sharon) = 90 – 79 = 11 %.
c. Probability that both wont gets a "B" is
   P ((Harold ∩ Sharon)′) = 100 - P (Harold ∩ Sharon) = 100 – 79 = 21 %.

CS 513 Knowledge Discovery & Data Mining
Vidya Sagar Polaki
CWID 10430970

Homework 1.3

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?

Solution:

The events "Jerry is at the Bank" and "Susan is at the Bank" are **not Independent.** As both go together to the bank at 8% of days. If independent, probability of them going together to bank should be product of their individual probabilities which is 20*30 = 6%.

Hence events are not independent.

Homework 1.4

You roll 2 dice.

a. Are the events "the sum is 6" and "the second die shows 5" independent?

b. Are the events "the sum is 7" and "the first die shows 5" independent?

Solution:

a. The outcomes of 2 dices can be shown as below

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$\mathrm{P}\ (\mathrm{Sum} = 6)$ = 5/36 = 13.89%

$\mathrm{P}\ (\text{Second die shows } 5)$ = 6/36 = 16.67%

$\mathrm{P}\ (\text{Second die shows 5 and sum} = 6)$ = 1/36 = 2.78%

If independent,

$\mathrm{P\ (Second\ die\ shows\ 5\ and\ sum} = 6)$ should equal $\mathrm{P\ (Sum} = 6) * \mathrm{P\ (Second\ die\ shows\ 5)}$

= 13.89 * 16.67 = 2.31%. As both are not equal, events are **not independent.**

b. The outcomes of 2 dices can be shown as below

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	9	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$\mathrm{P\ (Sum} = 7)$ = 6/36 = 16.67%

$\mathrm{P\ (First\ die\ shows\ 5)}$ = 6/36 = 16.67%

$\mathrm{P\ (First\ die\ shows\ 5\ and\ sum} = 7)$ = 1/36 = 2.78%

If independent,

$\mathrm{P\ (First\ die\ shows\ 5\ and\ sum} = 6)$ should equal $\mathrm{P\ (Sum} = 6) * \mathrm{P\ (First\ die\ shows\ 5)}$

= 16.67 * 16.67 = 2.78%. As both are equal, events are **independent.**

Homework 1.5

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?

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

Solution:

	TX	AK	NJ	
Oil	18%	6%	1%	25%
No Oil	42%	24%	9%	75%
	60%	30%	10%	100%

Probability of finding oil in TX (P (Oil | TX) = 30%) and Probability of choosing TX (P (TX) = 60%),
Therefore, $\mathrm{P}\ (\mathrm{Oil} \cap\ \mathrm{TX})$ = P (Oil | TX) * P (TX) = 30 % * 60% = 18 %

Similarly,

$\mathrm{P}\ (\mathrm{Oil} \cap\ \mathrm{AK})$ = P (Oil | AK) * P (AK) = 20 % * 30% = 6%

$\mathrm{P}\ (\mathrm{Oil} \cap\ \mathrm{NJ})$ = P (Oil | NJ) * P (NJ) = 10 % * 10% = 1 %

a. Therefore, probability of finding oil = 18% + 6% + 1% = 25%
b. Probability of drilling in TX if company found oil is $\mathrm{P}\ (\mathrm{TX} \mid \mathrm{Oil}) = \frac{P(\mathrm{TX} \cap \mathrm{Oil})}{\mathrm{P}\ (\mathrm{Oil})}$ = 18/25 = 72%

Homework 1.6

The following slide shows the survival status of individual passengers on the Titanic. Use this information to answer the following questions

- What is the probability that a passenger did not survive?
- What is the probability that a passenger was staying in the first class?
- Given that a passenger survived, what is the probability that the passenger was staying in the first class?
- Are survival and staying in the first class independent?
- Given that a passenger survived, what is the probability that the passenger was staying in the first class and the passenger was a child?
- Given that a passenger survived, what is the probability that the passenger was an adult?
- Given that a passenger survived, are age and staying in the first class independent?

Survived table

			Survived		
			Cabin		
	1st	**2nd**	**3rd**	**Crew**	**Sub Total**
Adult	197	94	151	212	654
Child	6	24	27	-	57
Sub Total	**203**	**118**	**178**	**212**	**711**

Not survived table

			Not Survived		
			Cabin		
	1st	**2nd**	**3rd**	**Crew**	**Sub Total**

Adult	122	167	476	673	1,438
Child	-	-	52	-	52
Sub Total	**122**	**167**	**528**	**673**	**1,490**

Total Table

Total					
Cabin					
	1st	2nd	3rd	Crew	Grand Total
Adult	319	261	627	885	2,092
Child	6	24	79	-	109
Grand Total	325	285	706	885	2,201

Solution:

From the above tables, total number of passengers survived = 711-212= 499

Total passengers = 2201- 885 = 1316

- Probability that a passenger did not survive P (NS) is (1490-673)/ (2201-885) = 817/1316 = 62.08%
  Probability of passenger survived P (S)= 100 – P (NS) = 100 – 62.08 = 37.92%
- Probability that a passenger was staying in the first-class P (F) = 325/1316 = 24.69%
- Probability that the passenger was staying in the first class, given passenger survived P (S ∩ F) = 203 / 499 = 40.68%
- Survival and staying in first class are independent if probability of staying in first class and surviving is equal to their individual probabilities, Therefore
  24.69 * 37.92 = 9.36 %
  Clearly, Survival and staying in first class are **not independent.**
- Given that a passenger survived, probability that the passenger was staying in the first class and the passenger was a child = 6 / 499 = 1.2%
- Given that a passenger survived, probability that the passenger was an adult = 442 / 499 = 88.57%
- Probability of age given passenger survived = P(A|S) +P(C|S) = 442/499 + 57/499 = 1.
  Given passenger survived, probability of staying in first class = 40.68%
  Probability of age and staying in first class = 40.68%
  For age and staying in first class to be independent, probability of age and first class must be equal to the product of their individual probabilities, clearly both are equal, and the events are conditional **Independent.**

CS 513 Knowledge Discovery & Data Mining
Vidya Sagar Polaki
CWID 10430970

Homework 1.7

Replace the missing values below (?), assuming independence between age and cabin class

Total Table

Total					
Cabin					
	1st	2nd	3rd	Crew	Grand Total
Adult	?	?	?	?	2,092
Child	?	?	?	?	109
Grand Total	325	285	706	885	2,201

Replace the missing values below (?), assuming independence between age and cabin class given survival status (conditional independence)

Survived table

Survived					
Cabin					
	1st	**2nd**	**3rd**	**Crew**	**Sub Total**
Adult	?	?	?	?	654
Child	?	?	?	?	57
Sub Total	**203**	**118**	**178**	**212**	**711**

Not survived table

Not Survived					
Cabin					
	1st	**2nd**	**3rd**	**Crew**	**Sub Total**
Adult	?	?	?	?	1438
Child	?	?	?	?	52
Sub Total	**122**	**167**	**528**	**673**	**1,490**

Solution: As mentioned in the problem statement, assuming independence between age and cabin class given survival status (conditional independence), we can get values from the previous problem as it satisfies the required conditions. Therefore

Survived table

Survived					
Cabin					
	1st	**2nd**	**3rd**	**Crew**	**Sub Total**
Adult	197	94	151	212	654
Child	6	24	27	-	57
Sub Total	**203**	**118**	**178**	**212**	**711**

Not survived table

Not Survived					
Cabin					
	1st	**2nd**	**3rd**	**Crew**	**Sub Total**
Adult	122	167	476	673	1,438
Child	-	-	52	-	52
Sub Total	**122**	**167**	**528**	**673**	**1,490**

Total Table

Total					
Cabin					
	1st	2nd	3rd	Crew	Grand Total
Adult	319	261	627	885	2,092
Child	6	24	79	-	109
Grand Total	325	285	706	885	2,201

******* THE END *******