CS 513-A: Knowledge Discovery & Data Mining

Schaefer School of Engineering & Science at Stevens Institute of Technology

Spring 2018

Individual Assignment 1: Probability

Name: Balaji Katakam

Date: 02/13/2018

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?

Answer 1.1:

Probability that Jerry goes to the bank is 20% of the days [can be represented as] : $P(J) = 0.2$

Probability that Susan goes to the bank is 30% of the days [can be represented as] : $P(S) = 0.3$

$P(J') = 1 – P(J) = 1 – 0.2 = 0.8$

Probability that Jerry doesn't go to the bank is 80% of the days [can be represented as]: $P(J') = 0.8$

$P(S') = 1 – P(S) = 1 – 0.3 = 0.7$

Probability that Susan doesn't go to the bank is 70% of the days [can be represented as]: $P(S') = 0.7$

Probability that Jerry and Susan are together at the bank is 8% of the days [can be represented as]: $P(J \cap S)=0.08$

a.] The probability that Jerry was there at the bank when Susan was at the bank last Monday is **26.666%.**

$P (J \cap S) / P (S) = (0.08 / 0.3) * 100 = 0.2666 * 100 = 26.66\%$

b.] the probability that Jerry was there at the bank when Susan wasn't there at the bank last Friday is **17.142%**.

$P (J \mid S') = P (J \cap S') / P (S')$

$P(J \cap S') = 0.2 – 0.08 = 0.12$

Hence $P (J \mid S') = 0.12 / 0.7 = 0.17142$

c.] The probability that both Jerry & Susan were there at the bank last Wednesday is **19.048%**

$P (J \cap S) / P(J \cup S) = P (J \cap S) / (P (J) + P (S) – P (J \cap S))$

Therefore, $P (J \cap S) / P (J \cup S) = 0.08/(0.2+0.3-0.08) = 19.048\%$

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"?

Answer 1.2:

Consider two events X & Y.

The probability of Harold getting Grade B is event X.

The probability of Sharon getting Grade B is event Y.

Harold - Probability of event X is [can be represented as]: $P (X) = 0.8$

Sharon - Probability of event Y is [can be represented as]: $P (Y) = 0.9$

Probability of event X & Y both occur is [can be represented as]: $P (X \cup Y) = 0.91$

The sample space for the given problem is being considered equal to 1: Sample Space = S = 1

$P (X \cap Y) = P (X) + P (Y) – P (X \cup Y) = 0.8 + 0.9 - 0.91 = 0.79$

a.] The probability that only Harold gets grade "B" is **1%**

$P (X) – P (X \cap Y) = 0.8 – 0.79 = 0.01$

b.] The probability that only Sharon gets grade "B" is **11%**

$P (Y) – P (X \cap Y) = 0.9 – 0.79 = 0.11$

c.] The probability that none of them gets grade "B" is **9%**

(Sample Space) $S – P (X \cup Y) = 1 - 0.91 = 0.09$

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?

Answer 1.3:

Probability that Jerry goes to the bank is 20% of the days [can be represented as] : $P(J) = 0.2$

Probability that Susan goes to the bank is 30% of the days [can be represented as] : $P(S) = 0.3$

Probability that they are together at the bank [can be represented as] : $P(J \cap S) = 0.08$

Independency test

$P (J \cap S) = P (J) * P (S)$ [if condition satisfied then they are independent]

$0.08 \neq 0.3*0.2$

Since condition is not satisfied they are **dependent events**.

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?

Answer 1.4:

$$\text{The sample space } S \text{ for this experiment is } S = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

a.] Probability that the Sum of the events is 6. P (A) = 5/36

Probability that the second die shows 5 is P (B) = 6/36

Independency test

The probability that both the events occur together is denoted as P (A ∩ B) = 1/36

P (A ∩ B) = P (A) * P (B) [if condition satisfied then they are independent]

1/36 ≠ 5/36 * 6/36

Since condition is not satisfied they are **dependent events**.

b.] Probability that the Sum of the events is 7. P (A) = 6/36

Probability that the first die shows 5 is P (B) = 6/36

Independency test

The probability that both the events occur together is denoted as P (A ∩ B) = 1/36

P (A ∩ B) = P (A) * P (B) [if condition satisfied then they are independent]

1/36 = 6/36 * 6/36

Since condition is satisfied they are **independent events**.

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?

Answer 1.5:

1.] The probability of drilling at State: TX [represented as] : P (TX) = 60%

The probability of drilling at State: AK [represented as] : P (AK) = 30%

The probability of drilling at State: NJ [represented as] : P (NJ) = 10%

The probability of finding Oil at State: TX [represented as] : **P (OTX) = 30 % * 60 % = 18%**

The probability of finding Oil at State: AK [represented as] : P (OAK) = 20 % * 30 % = 6%

The probability of finding Oil at State: NJ [represented as] : P (ONJ) = 10 % * 10 % = 1%

The total probability of finding Oil [represented as] : **P (FO) = P (OTX) + P (OAK) +P (ONJ) = 18% + 6% + 1% = 25%**

2.]

The probability that they drilled in TX and found oil [represented as] : P (DTXFO) = P (OTX) / P (FO) = 18 %/25 % = **72%**

Homework 1.6: the survival status of individual passengers on the Titanic. Use this information to answer the following questions

1. What is the probability that a passenger did not survive?
2. What is the probability that a passenger was staying in the first class?
3. Given that a passenger survived, what is the probability that the passenger was staying in the first class?
4. Are survival and staying in the first class independent?
5. Given that a passenger survived, what is the probability that the passenger was staying in the first class and the passenger was a child?
6. Given that a passenger survived, what is the probability that the passenger was an adult?
7. Given that a passenger survived, are age and staying in the first class independent?

Survived

Cabin

Age		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

Cabin

Age		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

Cabin

Age		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

Answer 1.6:

[Considering crew is not a passenger]

1. The total number of passengers = 2201 – 885 = 1316

The total number of passengers who did not survive = 1490 – 673 = 817

The probability that a passenger did not survive = (817/1316)*100 = **62.082%**

2. The total number of passengers in first class = 325

The probability that a passenger was staying in the first class = 325/1316 = **24.696%**

3. The total number of passengers who survived in first class = 203

The total number of passengers who survived = 499

the probability that the passenger was staying in the first class and survived = 203/499 = **40.681%**

4. The probability of survival = P (S) = 711/2201 = 32.3%

The probability of staying in first class = P (F) = 325/2201 = 14.766% = 14.77%

The probability of survival and staying in the first class is : P (S ∩ F) = 203/325 = 62.46%

Independency test

P (S ∩ F) = P (S) * P (F) [if condition satisfied then they are independent]

0.6246 ≠ 0.323*0.1477

Since condition is not satisfied they are **dependent events.**

5. The total number of passengers who survived = 499

The total number of child passengers staying in first class = 6

the probability that the passenger was staying in the first class and the passenger was a child P (FC) = 6/499 = **1.202%**

6. The total number of adult passengers who survived = 442

The total number of passengers who survived = 499

the probability that the passenger was an adult given that a passenger survived | P (APS) = 442/499 = **88.577%**

7. The probability of age and survived passengers = P(Age|Survived)=(P(Adult|Survived) + P(Child|Survived)) 499/499 =1

The probability of first class passengers & given survived = P(First class |Survived) = 203/499 = 40.681%

Probability of first class and age and they survived = P(First class|Survived ∩ Age|Survived) = 203/499 = 40.681%

To check if they are Independent events, then

P(First class|Survived ∩ Age|Survived)= P(Age|Survived)*P(First class|Survived)

0.40681 = 1 * 0.40681

so they are **independent events**.

Homework 1.7:

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

Total

Cabin

Age		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

Cabin

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

Not Survived

Cabin

Age		1st	2nd	3rd	Crew	Sub Total
	Adult	?	?	?	?	1,438
	Child	?	?	?	?	52
	Sub Total	122	167	528	673	1,490

Answer 1.7:

P(A∩B/C)= P(A/C) * P(B/C)

Or Can also be referred to as

Z=P(B/C)

Y=P(A/C)

P(Y and Z) = P(Y)*P(Z)

The calculations have been given below and the values have been listed below

			Total			
				Cabin		
		1st	2nd	3rd	Crew	Grand Total
	Adult	309	271	671	841	2092
Age	Child	16	14	35	44	109
	Grand Total	325	285	706	885	2201

			Not Survived			
				Cabin		
		1st	2nd	3rd	Crew	Grand Total
	Adult	187	108	164	195	654
Age	Child	16	10	14	17	57
	Grand Total	203	118	178	212	711

			Survived			
			Total			
				Cabin		
		1st	2nd	3rd	Crew	Grand Total
	Adult	118	161	510	649	1438
Age	Child	4	6	18	24	52
	Grand Total	122	167	528	673	1490

For conditional independence for survived

For Number of adults in 1st class

P(A∩B/C) = P(A/C) * P(B/C)

C=person survived

A=Person is adult

B=Person stayed in 1st class

P(A/C)=654/711

P(B/C)=203/711

P(A∩B/C)=X/711

X/711=(654/711)*(203/711)

X=187 = Number of adults in 1st class.

For Number of adults in 2nd class

P(A∩B/C) = P(A/C) * P(B/C)

C=person survived

A=Person is adult

B=Person stayed in 2^{nd} class

P(A/C)=654/711

P(B/C)=118/711

P(A∩B/C)=X/711

X/711=(654/711)*(118/711)

X=108.54=109 = Number of adults in 2^{nd}class.

For Number of adults in 3^{rd} class

P(A∩B/C) = P(A/C) * P(B/C)

C=person survived

A=Person is adult

B=Person stayed in 3^{rd} class

P(A/C)=654/711

P(B/C)=178/711

P(A∩B/C)=X/711

X/711=(654/711)*(178/711)

X=163.729=164 = Number of adults in 3^{rd} class.

For Number of adults in Crew

P(A∩B/C) = P(A/C) * P(B/C)

C=person survived

A=Person is adult

B=Person was a crew

P(A/C)=654/711

P(B/C)=212/711

P(A∩B/C)=X/711

X/711=(654/711)*(212/711)

X=195 = Number of adults in Crew.

For Number of children in 1st class

P(A∩B/C) = P(A/C) * P(B/C)

C=children survived

A=Person is child

B=Person stayed in 1st class

P(A/C)=57/711

P(B/C)=203/711

P(A∩B/C)=X/711

X/711=(57/711)*(203/711)

X=16.274=16 = Number of children in 1st class.

For Number of children in 2nd class

P(A∩B/C) = P(A/C) * P(B/C)

C=children survived

A=Person is child

B=Person stayed in 2nd class

P(A/C)=57/711

P(B/C)=118/711

P(A∩B/C)=X/711

X/711=(57/711)*(118/711)

X=9.4599=9 = Number of children in 2nd class.

For Number of children in 3rd class

P(A∩B/C) = P(A/C) * P(B/C)

C=children survived

A=Person is child

B=Person stayed in 3rd class

P(A/C)=57/711

P(B/C)=178/711

P(A∩B/C)=X/711

X/711=(57/711)*(178/711)

X=14.27=14 = Number of children in 3rd class.

For Number of children in crew

P(A∩B/C) = P(A/C) * P(B/C)

C=children survived

A=Person is child

B=Person was a crew

P(A/C)=57/711

P(B/C)=212/711

P(A∩B/C)=X/711

X/711=(57/711)*(212/711)

X=16.995=17 = Number of children in crew.

Not Survived

Cabin

Age

	1st	2nd	3rd	Crew	Sub Total
Adult	187	109	164	195	654
Child	16	9	14	18	57
Sub Total	203	118	178	212	711

For not Survived

For Number of adults in 1st class

P(A∩B/C) = P(A/C) * P(B/C)

C=person survived

A=Person is adult

B=Person stayed in 1st class

P(A/C)=1438/1490

P(B/C)=122/1490

P(A∩B/C)=X/1490

X/1490=(1438/1490)*(122/1490)

X=117.742=118 = Number of adults in 1st class.

For Number of adults in 2nd class

P(A∩B/C) = P(A/C) * P(B/C)

C=person survived

A=Person is adult

B=Person stayed in 2nd class

P(A/C)=1438/1490

P(B/C)=167/1490

P(A∩B/C)=X/1490

X/1490=(1438/1490)*(167/1490)

X=161.171=161 = Number of adults in 2ndclass.

For Number of adults in 3rd class

P(A∩B/C) = P(A/C) * P(B/C)

C=person survived

A=Person is adult

B=Person stayed in 3rd class

P(A/C)=1438/1490

P(B/C)=528/1490

P(A∩B/C)=X/1490

X/1490=(1438/1490)*(528/1490)

X=509.573=510 = Number of adults in 3rd class.

For Number of adults in Crew

P(A∩B/C) = P(A/C) * P(B/C)

C=person survived

A=Person is adult

B=Person was a crew

P(A/C)=1438/1490

P(B/C)=673/1490

P(A∩B/C)=X/1490

X/1490=(1438/1490)*(673/1490)

X=649.512 = 650= Number of adults in Crew.

For Number of children in 1^{st} class

P(A∩B/C) = P(A/C) * P(B/C)

C=children survived

A=Person is child

B=Person stayed in 1^{st} class

P(A/C)=52/1490

P(B/C)=122/1490

P(A∩B/C)=X/1490

X/1490=(52/1490)*(122/1490)

X=4.257=4= Number of children in 1^{st} class.

For Number of children in 2^{nd} class

P(A∩B/C) = P(A/C) * P(B/C)

C=children survived

A=Person is child

B=Person stayed in 2^{nd} class

P(A/C)=52/1490

$P(B/C)=167/1490$

$P(A\cap B/C)=X/1490$

$X/1490=(52/1490)*(167/1490)$

X=5.828=6 = Number of children in 2^{nd} class.

For Number of children in 3^{rd} class

$P(A\cap B/C) = P(A/C) * P(B/C)$

C=children survived

A=Person is child

B=Person stayed in 3^{rd} class

$P(A/C)=52/1490$

$P(B/C)=528/1490$

$P(A\cap B/C)=X/1490$

$X/1490=(52/1490)*(528/1490)$

X=18.426=18 = Number of children in 3^{rd} class.

For Number of children in crew

$P(A\cap B/C) = P(A/C) * P(B/C)$

C=children survived

A=Person is child

B=Person was a crew

$P(A/C)=52/1490$

$P(B/C)=673/1490$

$P(A\cap B/C)=X/1490$

$X/1490=(52/1490)*(673/1490)$

X=23.487=23 = Number of children in crew.

On adding the table of Survived and not Survived we can create the total table