FINAL PROJECT – PART 1

Probability and Bayes' Theorem Questions:

ראשית נבהיר כי בתרגילים הבאים יש שימוש בחוק בייס שעל פיו מתקיים:

$$\underbrace{P(A|B)}_{\text{Posterior}} = \frac{\overbrace{P(B|A)}^{\text{Likelihood}}\overbrace{P(A)}^{\text{Prior}}}{\underbrace{P(B)}_{\text{Normalizing constant}}}$$

$$P(B) = \sum_{Y} P(B|A)P(A)$$

1א- תרגיל: בערך 1/300 מהלידות היא של תאומים זהים ו-1/125 מהלידות היא של תאומים לא זהים. לאלביס היה אח תאום שמת בלידה. מה ההסתברות שהיה לו אח תאום זהה? (ניתן להניח שההסתברות להולדת בן ובת שווה ל-0.5)

פתרון: ההסתברות לתאומים זהים וגם לתאומים שהם אחים הוא 1/300 X ½ = 1/600. ההסתברות לתאומים לא זהים וגם לתאומים שהם אחים הוא 1/125 X ¼ = 1/500. לכן, ע"פ חוק בייס:
P(Identical twins| Twin brother) = ההסתברות שיהיה לאלביס אח תאום זהה היא 5/11
= (1/600) / (1/500+1/600).

1ב- תרגיל: יש שתי קערות של עוגיות. בקערה 1 יש 10 עוגיות שקדים ו-30 עוגיות שוקולד. בקערה 2 יש 20 עוגיות שקדים ו-20 עוגיות שוקולד. אריק בחר קערה באקראי ובחר ממנה עוגיה באקראי. העוגיה שנבחרה היא שוקולד. מה ההסתברות שאריק בחר את קערה 1?

פתרון: ההסתברות לבחירת קערת שקדים מקערה אחת היא 1/4 X ½ = 1/8. ההסתברות לבחירת עוגיית שוקולד מקערה אחת היא ¾ X ½ = 3/8. הסיכוי לבחירת עוגיית שקדים מקערה 2 היא ½ X ½ = ¼ וההסתברות לבחירת עוגיית שוקולד מקערה 2 זהה. לכן, ע"פ חוק בייס:
P(Bowl 1 | Chocolate cookie) = ההסתברות שאריק בחר עוגיית שוקולד מקערה 1 היא 0.6
= (3/8) / (1/4+3/8).

2- תרגיל: בשנת 1994 חברת M&M הוסיפה את הצבע הכחול. לפני השנה הזו, התפלגות הצבעים בשקית M&M הייתה נראית כך:

30% Brown, 20% Yellow, 20% Red, 10% Green, 10% Orange, 10% Tan

החל משנת 1995, ההתפלגות נראית כך:

24% Blue, 20% Green, 16% Orange, 14% Yellow, 13% Red, 13% Brown

לחבר שלכם יש 2 שקיות M&M, אחת משנת 1994 ואחת משנת 1995 והוא לא מוכן לגלות לכם איזו שקית שייכת לאיזו שנה. אבל הוא נותן לכם סוכריה אחת מכל שקית. סוכריה אחת היא צהובה ואחת היא ירוקה. מה הסיכוי שהסוכריה הצהובה הגיעה מהשקית של 1994?

פתרון: ההסתברות לסוכריה ירוקה משנת 1994 ירוקה היא 0.1. ההסתברות לסוכריה משנת 1994 צהובה היא 0.2. ההסתברות לסוכריה ירוקה משנת 1995 היא 0.2, ואילו לצהובה מאותה שנה היא 0.14.
בנוסף, ההסתברות להוצאת סוכריה צהובה משקית אחת ומהשקית האחרת סוכריה ירוקה היא 0.1X0.14+0.2X0.2 = 0.054 ומושפעת מכך שהצהובה יכולה להיות משקית משנת 1994 ובו בזמן הירוקה משקית שנת 1995, או להפך.
מאחר ויצאה סוכריה צהובה מהשקית של 1994 ברור כי במצב זה יצאה סוכריה ירוקה מהשקית משנת 1995, וההסתברות לכך היא 0.2X0.2 = 0.04. לכן, ע"פ חוק בייס:
ההסתברות שהסוכריה הצהובה הגיעה = P(Yellow from 1994 | Yellow & Green from 2 different bags) מהשקית של 1994 היא בקירוב 0.74 = (0.1X0.14+0.2X0.2) / (0.2X0.2).

3א- תרגיל: הלכת לדוקטור בעקבות ציפורן חודרנית. הדוקטור בחר בך באקראי לבצע בדיקת דם הבודקת שפעת חזירים. ידוע סטטיסטית ששפעת זו פוגעת ב-1 מתוך 10,000 אנשים באוכלוסייה. הבדיקה מדויקת ב-99 אחוז במובן שההסתברות ל false positive היא 1%. הווה אומר שהבדיקה סיווגה בטעות אדם בריא כאדם חולה היא 1 אחוז.
ההסתברות ל false negative היא 0 – אין סיכוי שהבדיקה תגיד על אדם החולה בשפעת חזירים שהוא בריא.
בבדיקה יצאת חיובי (יש לך שפעת). מה ההסתברות שיש לך שפעת חזירים?

פתרון: ההסתברות שיש לך שפעת כאשר הבדיקה חיובית היא 1 . ההסתברות שאין לך שפעת היא 9999/10000. ההסתברות שיש לך שפעת היא 1/10000. ההסתברות שתצא חיובי למרות שאין לך שפעת היא 1/100. לכן, ע"פ חוק בייס:
ההסתברות שיש לך שפעת חזירים לאחר שהבדיקה = P(Having swine flu | The test is positive and correct) יצאה חיובית היא בקירוב 0.01 = (0.0001 X 1) / (0.9999 X 0.01 + 0.0001 X 1).

3ב- תרגיל: נניח שחזרת מתאילנד לאחרונה, ואתה יודע ש-1 מתוך 200 אנשים שחזרו לאחרונה מתאילנד, חזרו עם שפעת חזירים. בהינתן אותה סיטואציה כמו בסעיף א, מה ההסתברות (המתוקנת) שיש לך שפעת חזירים?

פתרון : במקרה זה, ההסתברות היחידה שהשתנתה ביחס לסעיף א' היא שיש לך שפעת, וכעת ההסתברות לכך היא 1/200, כלומר ההסתברות לכך שאין לך שפעת היא 199/200, ולכן ע"פ חוק בייס:
ההסתברות שיש לך שפעת חזירים לאחר שהבדיקה יצאה חיובית = P(Having swine flu | The test is positive) היא בקירוב 0.33 = (0.005 X 1) / (0.995 X 0.01 + 0.005 X 1).

4- תרגיל: בערך 1/300 מהלידות היא של תאומים זהים ו-1/125 מהלידות היא של תאומים לא זהים. לנסיך צ'ארלס היה אח תאום שמת בלידה. מה ההסתברות שהיה לו אח תאום זהה? (תאומים זהים חייבים להיות בני אותו מין)

פתרון : תרגיל זה זהה לתרגיל 1 (אך עם שמות שונים), ולכן הפתרון זהה.

Random Variables Questions:

For these questions we will use the following equation:

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

---

1. Roi is playing a dice game with Yael.
Roi will roll 2 six-sided dice, and if the sum of the dice is divisible by 3, he will win 6$. If the sum is not divisible by 3, he will lose 3$.
What is Roi's expected value of playing this game?

Solution: The options of getting a sum divisible by 3 are: {(1,2), (2,1), (1,5), (5,1), (2,4), (4,2), (3,3), (3,6), (6,3), (4,5), (5,4), (6,6)}. We can now see that there are 12 options. Therefore, while rolling 2 six-sided dices, the chance of Roi winning is 12/36 = 1/3. His chances of losing are 1-1/3=2/3.
According to the equation above, the expected value of playing this game is 0 $
= (1/3 X 6)+(2/3 X (-3))

2. Sharon has challenged Alex to a round of Marker Mixup. Marker Mixup is a game where there is a bag of 5 red markers numbered 1 through 5, and another bag with 5 green markers numbered 6 through 10.
Alex will grab 1 marker from each bag, and if the 2 markers add up to more than 12, he will win 5$, 5. If the sum is exactly 12, he will break even, and If the sum is less than 12, he will lose 6$.
What is Alex's expected value of playing Marker Mixup?

Solution:
The options for a sum of 12: {(2,10), (3,9), (4,8), (5,7)}
The options for a sum greater than 12: {(3,10), (4,9), (4,10), (5,10), (5,8), (5,9)}
The options for a sum smaller than 12: {(1,6), (1,7), (1,8), (1,9), (1,10), (2,6), (2,7), (2,8), (2,9), (3,6), (3,7), (3,8), (4,6), (4,7), (5,6)}
Therefore, the probability of grabbing 2 markers from different bags with a sum:
→ greater than 12 is 6/20
→ equal to 12 is 4/20
→ smaller than 12 is 15/20
According to the equation above, the expected value of playing Marker Mixup is -3 $
= (6/20 X 5) + (4/20 X 0) + (15/20 X (-6))

3. A division of a company has 200 employees, 40%, percent of which are male. Each month, the company randomly selects 8 of these employees to have lunch with the CEO.
What are the mean and standard deviation of the number of males selected each month?

Solution: In order to solve this problem, we will use the following equations:

Probability Distribution Formula

Mean ($\bar{x}$) = $\sum [x_i * P(x_i)]$

Standard Deviation (σ) = $\sqrt{\sum (x_i - \bar{x})^2 * P(x_i)}$

The mean of the number of males selected each month is 3.2 = μ = 8 X 0.4.

We will now go over all the possibilities of men out of 8 employees.
We will describe one option thoroughly and write a rough draft for the rest:

Theoretical Probability

- The general formula is

$$P(E) = \frac{\text{Number of ways } E \text{ can occur}}{\text{Number of possible outcomes}}$$

- If we have an experiment where
  - There are n equally likely outcomes (i.e. $N(S) = n$)
  - The event E consists of m of them (i.e. $N(E) = m$)

  then
$$P(E) = \frac{m}{n} = \frac{N(E)}{N(S)}$$

A division company with 80 = 200X0.4 male and 120 = 200X0.6 female employees selects 8 of them for lunch with the CEO.
8 out of 200 people can be selected in
(200!) / (8! X (200-8)!) ways.
There are (80!) / (3! X (80-3)!) different ways of picking 3 men out of 80.
There are (120!) / (5! X (120-5)!) different ways of picking 5 women out of 120.
So, the total number of ways of choosing 3 men out of 80 men and 5 women out of 120 women simultaneously are
(120! X 80!) / (3! X (80-3)! X 5! X (120-5)!) ways.
Hence the probability of selecting 3 men and 5 women for an 8-member group from a group of 80 men and 120 women is = [(120! X 80!) / (3! X (80-3)! X 5! X (120-5)!)] / [(200!) / (8! X (200-8)!)] = 0.28417.

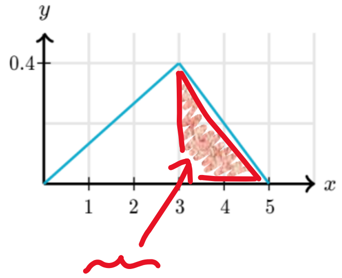
Choosing no man = P(x0)= 0.0153
Choosing 1 man = P(x1)= 0.0864
Choosing 2 men = P(x2)= 0.2095
Choosing 3 men = P(x3)= 0.2842
Choosing 4 men = P(x4)= 0.2358
Choosing 5 men = P(x5)= 0.1225
Choosing 6 men = P(x6)= 0.0389
Choosing 7 men = P(x7)= 0.0069
Choosing 8 men = P(x8)= 0.0005
We will define x as the number of males who got chosen.

Therefore, according to the following equation: $\sqrt{\sum_{i=1}^{N} p_i(x_i-\mu)^2}$, the Std Dev is 1.36.

4. Different dealers may sell the same car for different prices. The sale prices for a particular car are normally distributed with a mean and standard deviation of 26,000$ and 2,000$, respectively. Suppose we select one of these cars at random. Let X = the sale price (in thousands of dollars) for the selected car.
Find P(26<X<30).

Solution: This probability can be calculated using the normal distribution formula: z=(x-μ)/ʊ .
The Std Dev is ʊ = 2 , the Mean is μ = 26. Therefore, P(26<X<30) = P((26-26)/2 < z < (30-20)/2) = P(0 < z < 2) = P(z<2) – P(z<0) = 0.97725-0.5 (from the Standard Normal Table) ≈ 0.48 =
48% = P(26< X < 30) .

5. Given the following distribution, what is P(x>3)?



Solution: P(x>3) = 0.4 = (0.4 X (5-3) / 2 .

6. A company has 500 employees, and 60% of them have children. Suppose that we randomly select 4 of these employees.
What is the probability that exactly 3 of the 4 employees selected have children?

Solution: For this problem we will use the following formula:

Binomial Distribution Formula

$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

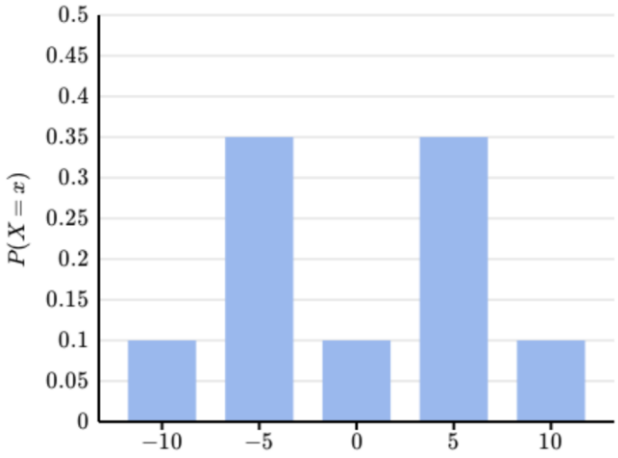
where
n = the number of trials (or the number being sampled)
x = the number of successes desired
p = probability of getting a success in one trial
$q = 1 - p$ = the probability of getting a failure in one trial

In accordance with the formula:

- ➔ n=4=The number of employees selected
- ➔ x=3=The desired number of employees with children
- ➔ p=0.6=There is a 60% chance of having a child out of the companies' employees
- ➔ q=1-p=0.4

Therefore, the probability that exactly 3 of the 4 employees selected have children is 0.1152 =(4/3) X (0.6)^3 X 0.4 .

7. Look at the next Graph. What is the expected value of X?



Solution: The expected value of x is 0 because (0.1X(-10))+(0.35X(-5))+(0.1X0)+(0.35X5)+(0.1X10)=0.

Pandas Coding Questions:

1. Write a Python Program that converts a number from base 10 and prints the number in other bases (base 2, 8 and 16).

   Solution:

```
def convert(num, base):

    if base < 2:
        print("False")
    remains = []
    while num > 0:
        remains.append(str(num % base))
        num = num // base
    remains.reverse()
    print(''.join(remains))
```

2. Answer the questions regarding the dataset.

   Solution: The answers are in the “cast.csv” file.