CYBER 759 – Final Exam
Gunnar Yonker

Part 1: Mathematical Reasoning

1.

Using Bayes' theorem:

A as the event that a subscriber watches action movies.

B as the event that a subscriber binge-watches.

P(A) = Probability that a subscriber watches action movies = 0.40

P(¬A) = Probability that a subscriber does not watch action movies (and thus watches romantic or other genres) = 1 - 0.40 = 0.60

P(B|A) = Probability that a subscriber binge-watches given that they watch action movies = 0.80

P(B|¬A) = Probability that a subscriber binge-watches given that they watch romantic or other genres = 0.40

P(A|B) = $\frac{P(B|A)\times P(A)}{P(B|A)xP(A)+P(B|\neg\mathrm{A})\mathrm{xP}(\neg\mathrm{A})}$

Plug in:

P(A|B) = $\frac{0.80x0.40}{0.80x0.40+0.40x0.60} = \frac{0.32}{0.56} = 0.5714$

The probability that they watched an action movie given that they binge-watched is approximately 0.5714 or 57.14%.

2.

Yes, we can deduce Lisa's salary form this data.

Before Lisa was hired there were 72 employees, and the mean salary was $60,000.

72 * 60,000 = 4,320,000

After Lisa was hired there were 73 employees, and the mean salary was $59,900.

73 * 59,900 = 4,372,700

To calculate Lisa's salary, we subtract the total salary before she was hired from the total salary after she was hired.

4,372,700 – 4,320,000 = 52,700

Thus, Lisa's salary is $52,700.

3.

The number of uploads per week follows a Poisson distribution with an average rate $\lambda=3$ videos per week.

The probability of more than 4 uploads is given by:

$P(X>4) = 1-P(X\leq4)$

Where $P(X\leq4)$ is the cumulative probability for 0, 1, 2, 3, and 4 uploads.

Using the above formula, we can calculate $P(X\leq4)$:

$P(X\leq4) = P(X = 0) + P(X = 1) + P(X = 2)+P(X = 3) + P(X = 4)$

$P(X = 0) = (1e^{-3})/1$

$P(X = 1) = (3e^{-3})/1$

$P(X = 2) = (9e^{-3})/2$

$P(X = 3) = (27e^{-3})/6$

$P(X = 4) = (81e^{-3})/24$

Then:

$P(X\leq4)= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

Calculate:

$P(X>4) = 1-P(X\leq4)$

Then:

$P(X = 0) = 0.0498$

$P(X = 1) = 0.1494$

$P(X = 2) = 0.2241$

$P(X = 3) = 0.2241$

$P(X = 4) = 0.1681$

Summing them up:

$P(X\leq4)= 0.0498 + 0.1494 + 0.2241 + 0.2241 + 0.1681 = 0.8155$

Finally:

$P(X>4)= 1 - 0.8155 = 0.1845$

So, there is about an 18.45% chance that the content creator uploads more than 4 videos in a given week.

4.

The probability that coin A lands tails is $P(A_{tail}) = 0.20$

The probability that coin B lands tails is $P(B_{tail}) = 0.70$

The combined probability that both coins show tails in one flip is:

$P(A_{tail} \text{ and } B_{tail}) = P(A_{tail}) \text{ x } P(B_{tail}) = 0.20 \text{ x } 0.70 = 0.14$

The probability that both coins show tails in three consecutive flips is:

$(P(A_{tail} \text{ and } B_{tail}))^3 = (0.14)^3 = 0.002744$

P(at least one head) = 1 – P(no heads in any flip)

= 1 – 0.002744

= 0.997256

So, the probability of getting at least one head (in either of the coins) when both coins are flipped simultaneously three times in a row is approximately 0.997256 or 99.73%.

5.

P(Review|Electronics) = 0.4

P(Review|Clothing) = 0.2

P(Review|Groceries) = 0.1

Given the overall shopping behavior:

P(Electronics) = 0.3

P(Clothing) = 0.5

P(Groceries) = 0.2

Using Bayes' Theorem:

$P(A|B) = \frac{P(B|A) x P(A)}{P(B)}$

Where:

P(A|B) is the probability that a product is from category A given that a review was written.

P(B|A) is the probability that a review was written given that the product is from category A.

CYBER 759 – Final Exam
Gunnar Yonker

P(A) is the probability that a product is from category A.

P(B) is the total probability that a review was written.

P(Review) = P(Electronics) × P(Review|Electronics) + P(Clothing) × P(Review|Clothing) + P(Groceries)× P(Review|Groceries)

P(Review) = 0.3 × 0.4 + 0.5 × 0.2 + 0.2 × 0.1

P(Review) = 0.24

Calculate probability for each category given a review:

P(Electronics|Review) = $\frac{0.4*0.3}{0.24}$

P(Electronics|Review) = 0.5

P(Clothing|Review) = $\frac{0.2*0.5}{0.24}$

P(Clothing|Review) = 0.4167

P(Groceries|Review) = $\frac{0.1*0.2}{0.24}$

P(Groceries|Review) = 0.0833

From the probabilities calculated, if you come across a customer who has left a review for a product, there's a:

50% chance they bought from Electronics

41.67% chance they bought from Clothing

8.33% chance they bought from Groceries

Given this, the best guess would be that the product they bought is from the Electronics category since it has the highest probability. The probability that my guess is correct is 50%.

6.

$$\frac{P(M(D) = O)}{P(M(D') = O)} \leq e^{\varepsilon}$$

M is the mechanism.

D and D’ are the neighboring databases.

P(M(D)=O) is the probability that the mechanism, when executed on database *D*, produces output O.

E is the base of the natural logarithm and ε is the differential privacy parameter.

Given ε = 0.85, the ratio of probabilities for any output is:

$E^{0.85}$ = 2.34

Bayesian Perspective:

- P(Mark is involved) is the prior probability that Mark is involved in a security incident (before the query).
- P(O|Mark is involved) is the probability of observing output O given that Mark is involved.
- P(O|Mark is not involved) is the probability of observing output O given that Mark is not involved.

$$P(Mark\ is\ involved|O) = \frac{P(O|Mark\ is\ involved)xP(Mark\ is\ involved)}{P(O)}$$

Where:

P(O) = P(O|Mark is involved) × P(Mark is involved) + P(O|Mark is not involved) × P(Mark is not involved)

Given the differential privacy guarantee:

$$\frac{P(O|Mark\ is\ invovled)}{P(O|Mark\ is\ not\ invovled)} \leq 2.34$$

After observing the output O, the evidence can at most increase the analyst's suspicion about Mark being involved by a factor of 2.34, compared to the original belief. Conversely, it can decrease the analyst's suspicion by the same factor.

Therefore:

Upper bound on suspicion: P(Mark is involved|O) ≤ 2.34 x P(Mark is involved)

Lower bound on suspicion: P(Mark is involved|O) ≥ 1/2.34 x P(Mark is involved)

7.

If the coin lands heads (with a probability of 45%), students answer truthfully.

If it lands tails (with a probability of 55%), they respond Yes regardless of whether they've cheated or not.

Denote p as the proportion of students who have actually cheated. Then, the expected proportion of Yes answers is:

$0.45p + 0.55$ = Expected proportion of Yes answers

Expected proportion of Yes answers if none of the students had cheated:

$0.45(0) + 0.55 = 0.55$

Out of 100 students, 40 said Yes. So, the observed proportion of Yes answers is 0.40 (or 40%).

Equating the expected proportion of Yes answers to the observed proportion:

$0.45p + 0.55 = 0.40$

Solve for p:

$0.45p = 0.40 - 0.55$

$0.45p = -0.15$

$p = -0.15/0.45$

$p = -0.33$

This answer does not make sense because a proportion cannot be negative. It implies that the observed number of Yes responses is lower than what we'd expect if none of the students had cheated. Only 40% were Yes answers which is less than the baseline expected, which lead to a negative proportion when trying to back out the actual rate of cheating. Considering the setup, the results indicate that it's statistically unlikely that any of the students have cheated. The negative value indicates a contradiction in the data. If the survey was properly conducted and the student properly followed instructions, then this could suggest that there was no cheating. However, it is possible that there was a bias in coin flipping, misunderstandings about the technique among the students, students worried the survey wasn't truly anonymous, or other external factors that influenced the results.

CYBER 759 – Final Exam
Gunnar Yonker

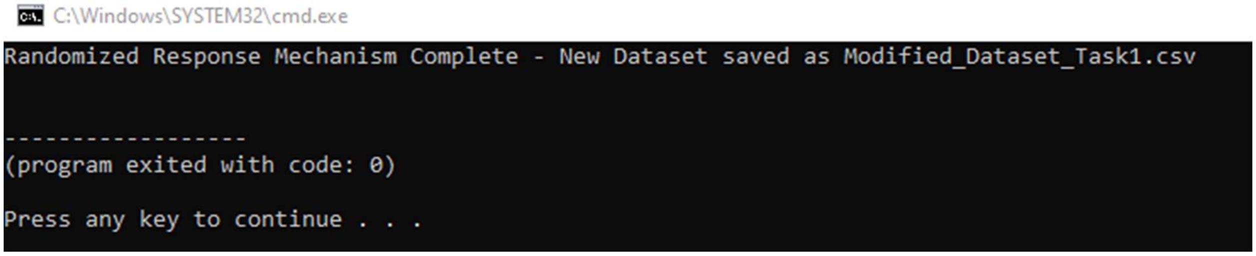
Part 2: Implementation

Task 1 – Randomized Response

task1.py

Dataset_Task1.csv

When running the code, it should produce the new dataset with the modified responses in a csv called Modified_Dataset_Task1.csv. This was tested successfully on Geany and contains comments that explain each step. Here is an example screenshot of the code running:



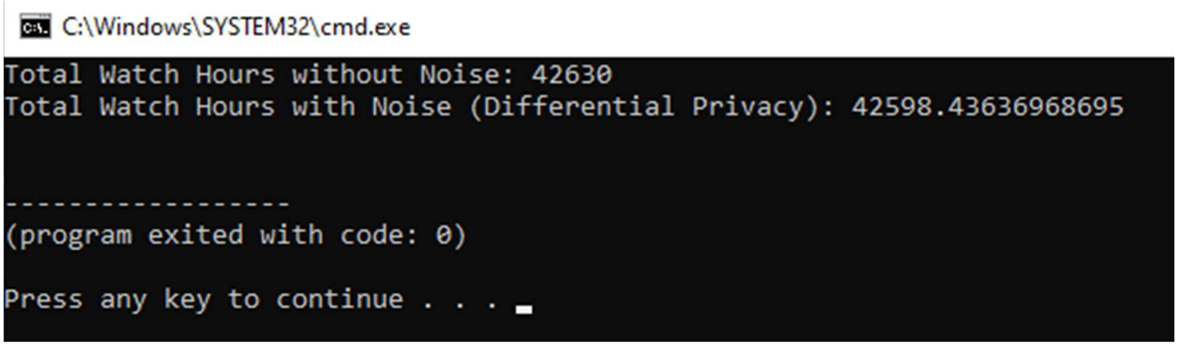
```
Randomized Response Mechanism Complete - New Dataset saved as Modified_Dataset_Task1.csv


------------------
(program exited with code: 0)

Press any key to continue . . .
```

Task 2 – The Laplace Mechanism

task2.py

Dataset_Task2.csv

When running the code, in the console window the output should show the total watch hours in a week, and then another line showing the total watch hours in a week with the Laplace mechanism applied to ensure differential privacy. This was tested successfully on Geany and contains comments that explain each step. Here is an example screenshot of the code running:



```
Total Watch Hours without Noise: 42630
Total Watch Hours with Noise (Differential Privacy): 42598.43636968695


------------------
(program exited with code: 0)

Press any key to continue . . .
```