CYBER 759 – Assignment 4
Gunnar Yonker

1.

There are a few different reasons that this definition is impractical:

Loss of Symmetry: Traditional differential privacy ensures that for any outcome, the probabilities between the two datasets are nearly equal (up to the privacy parameter). This means that neither dataset has an advantage in generating a particular outcome over the other. The proposed definition breaks this symmetry by only constraining outcomes in one direction.

Erosion of Privacy: This one-sided constraint could lead to scenarios where the output probabilities for D and D' are vastly different, yet the mechanism still qualifies as "differentially private" under the new definition. This could result in revealing more information about one dataset than the other.

Doesn't Capture Intuition Behind Differential Privacy: The essence of differential privacy is to ensure that the presence or absence of a single individual in the dataset does not significantly affect the outcome of a computation. The proposed definition does not guarantee this, as there could still be a significant difference in probabilities, just so long as D doesn't have a higher probability than D' for a given outcome.

Lack of Utility: If strictly adhered to, this definition could seriously constrain the utility of the data. For example, if D has a naturally higher occurrence of a certain feature or result than D', the mechanism would need to suppress this difference to ensure compliance. This could lead to outcomes that are less reflective of the actual data.

In summary, the modified definition of differential privacy doesn't effectively ensure privacy in a symmetric or robust manner, potentially erodes the privacy and utility of the data, and would be challenging to implement in practice.

2.

Local Differential Privacy (LDP): In local differential privacy, each individual's data is randomized before being sent to the database or collector. That is, noise is added at the individual data point level. As a result, the data collector never sees the raw, obfuscated data.

Advantages: Users' raw data is never shared, providing strong privacy guarantees at the data source.

Disadvantages: Often requires more noise to be added compared to global differential privacy, leading to less accurate aggregate results.

Some example scenarios are user surveys and telemetry data collection. If a company wants to collect user feedback without knowing specific users' answers, they can apply LDP. For example, when asking if users like a new feature, noise is added to individual responses before they are collected. Modern operating systems or apps might collect telemetry data to improve their services. Using LDP ensures that individual user data remains private.

Global Differential Privacy (GDP): In global differential privacy, raw data is collected by a trusted curator without noise addition. However, noise is added when queries are made to the dataset, ensuring that the results of these queries do not compromise individual privacy.

Advantages: Typically allows for more accurate statistical results than LDP since noise is added once during the query phase rather than at every data point.

Disadvantages: Requires trust in the data collector or curator, as they have access to raw, obfuscated data.

Some example scenarios are medical research and census data. If a trusted research institution collects medical data from patients for a study, they might use GDP to provide researchers with aggregate insights without revealing individual patient information. Governments might use GDP to publish statistical information about populations without risking the identification of individual citizens.

3.

P(Yes) = ½(1/2p + ½(1-p)) + ½ * ½ = 1/4p + ¼ -1/4p + 1/4 = ½

P(No) = ½ (works out the same way as P(Yes)

P(Yes) is ½ regardless of whether the true answer is “Yes” or “No”

Ratio of probabilities:

$$\frac{P(Yes|Yes)}{P(Yes|No)} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1$$

Taking the natural logarithm of both sides:

ln(1) = 0

Thus,

ϵ=0

The smallest value of ϵ for which the mechanism satisfies differential privacy is 0, which means that it offers perfect differential privacy.

4.

The probability of saying “Yes” because they lied: P(Lie) = 0.7.

The probability of saying “Yes” because they were telling the truth and got a head on the second flip: P(Truth) x P(Yes | Truth) = 0.3 * 0.7 = 0.21.

Therefore, the total probability of saying “Yes” = 0.7 + 0.21 = 0.91.

Given 1000 participants:

Expected number saying “Yes” if none were cheaters: 1000 * 0.91 = 910.

Given 1000 participants:

CYBER 759 – Assignment 4
Gunnar Yonker

Expected number saying “Yes” if all were truthful 1000 * 0.91 = 910

Expected number saying “No” if all were truthful 1000 * 0.09 = 90

Since 400 out of 1000 said “Yes”, this 510 fewer than expected if none were cheaters. Therefore, we can estimate that 510 participants are cheaters.

In conclusion:

510 individuals are estimated to cheat on their spouse.

490 individuals are estimated not to cheat on their spouse.