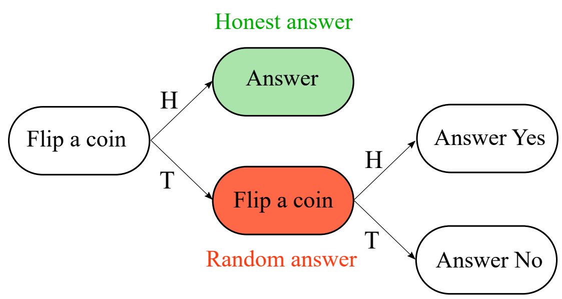
Privacy - Homework

October 26, 2023

1 Plausible deniability example

A sensitive "yes-no" question is asked to the participants of a survey. The following protocol is implemented.



Definition 1 (Hamming's distance) *Let D_1 and D_2 be two datasets of the same size, the Hamming's distance is define as the number of entries on which they differ.*

$$d(D_1, D_2) = \# \{i : (D_1)_i \neq (D_2)_i\}$$

Definition 2 *For all $\epsilon > 0$, a randomized algorithm A is ϵ-differentially private if, for all $S \subset Im(A)$ and for all D_1 and D_2 datasets such as $d(D_1, D_2) = 1$, we have*

$$\frac{\mathbb{P}(A(D_1) \in S)}{\mathbb{P}(A(D_2) \in S)} \leq \mathrm{e}^{\epsilon}$$

1. Let D_1 and D_2 be two sets of answers of size n, differing in their last entry: $(D_1)_n =$ Yes and $(D_1)_n =$ No. Show that the protocol is $\ln(3)$-differentially private.

2. Still in the case of a "yes-no" question, propose an ϵ-differentially private protocol.

2 Mechanisms

Definition 3 (p-sensitivity) $\Delta_p(A) = \max_{d(D_1,D_2)=1} \|A(D_1) - A(D_2)\|_p$

Theorem 1 (Laplace mechanism) *Let $\epsilon > 0$, A be an algorithm with values in $\mathbb{R}^d$ and D a dataset. Then, $\mathcal{M}_{Lap}(A, D, \epsilon) = A(D) + \mathbf{Z}$ with $\mathbf{Z} \sim Lap(\frac{\Delta_1(A)}{\epsilon})^{\otimes d}$ is ϵ-DP.*

Theorem 2 (Gaussian mechanism) *Let $\epsilon, \delta > 0$, A be an algorithm with values in $\mathbb{R}^d$ and D a dataset. Then, $\mathcal{M}_{Gauss}(A, D, \epsilon, \delta) = A(D) + \mathbf{Z}$ with $\mathbf{Z} \sim \mathcal{N}(0, \frac{2 \ln \frac{2}{\delta} \Delta_2(A)^2}{\epsilon^2})^{\otimes d}$ is (ϵ, δ)-DP.*

Definition 4 ((α, β)-accuracy) *A mechanism $\mathcal{M}$ is (α, β)-accurate w.r.t an algorithm A if for all dataset D and with probability at least $1 - \beta$, we have*

$$\|\mathcal{M}(A, D, \epsilon) - A(D)\|_\infty \leq \alpha(\epsilon)$$

1. Recalling that the p.d.f. of a random variable sampled according to Lap(a) is equal to $f(x) = \frac{1}{2a} \mathrm{e}^{-\frac{|x|}{a}}$, prove Theorem 1.
2. Explain why the definition of differential privacy is a "worst case" definition. How can we relax it?
3. Compute $\mathbb{E}[\|\mathcal{M}_{Lap}(A, D, \epsilon) - A(D)\|_1]$
4. For a fixed β, find α such that the Laplace mechanism is (α, β)-accurate.
5. For a fixed β, find α such that the Gaussian mechanism is (α, β)-accurate.
6. When d is big, which mechanism seems to be more appropriate?