Tailoring Differentially Private Bayesian Inference to Distance Between Distributions

Mark Bun[†], Gian Pietro Farina*, Marco Gaboardi*, Jiawen Liu*

[†]Princeton University, *University at Buffalo, SUNY

Objectives

Design a mechanism that achieve differential privacy by scaling to a metric between distribution.

1. A differentially private bayesian mechanism,
2. Calibrating mechanism noise by the same probabilistic distance we want to measure accuracy with.
3. Applying smooth sensitivity in mechanism to achieve better accuracy.

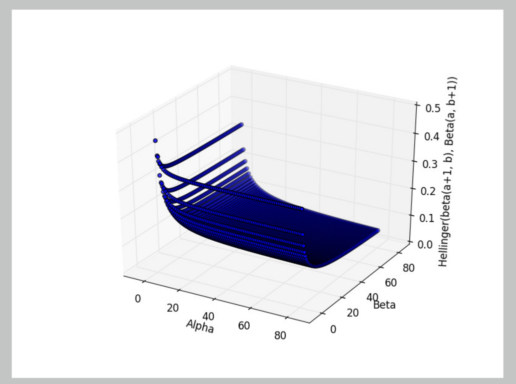


Figure 1: Hellinger Sensitivity

Bayesian Inference Background

Conjugate prior distribution, $\mathbf{beta}(\alpha, \beta)$, with hyper parameters $\alpha, \beta \in \mathbb{R}^+$;
Observed data set $\mathbf{x}$: $\mathbf{x} = (x_1, \dots x_n), x_i \in \{0, 1\}$, $n \in \mathbb{N}$

Bernoulli likelihood function: $\mathbf{Pr}(\mathbf{x}|\theta) \equiv \theta^{\Delta\alpha}(1-\theta)^{n-\Delta\alpha}$ where $\Delta\alpha = \sum_{i=1}^{n} x_i$;

Posterior distribution derived: $\mathbf{Pr}(\theta|\mathbf{x}) = \mathbf{beta}(\alpha + \Delta\alpha, \beta + n - \Delta\alpha)$

Differentially private Bayesian inference

Release a private version of posterior distribution $(\tilde{\alpha}, \tilde{\beta}) = (\alpha + \widetilde{\Delta\alpha}, \beta + n - \widetilde{\Delta\alpha})$.
In a baseline approach, we sample noise from $Lap(\mu, \nu)$ mechanism, i.e., $\widetilde{\Delta\alpha} \sim Lap(\Delta\alpha, \frac{2}{\epsilon})$,

Smoothed Hellinger Distance based Exponential Mechanism

Our approach defines the mechanism $\mathcal{M}^B_{\mathcal{H}}$: given in input an observations $\mathbf{x}$, parameters $\epsilon > 0$ and $\delta > 0$, produces an element r in $\mathcal{R}_{\text{post}}$ with probability:

$$\underset{z \sim \mathcal{M}^B_{\mathcal{H}}}{\mathbf{Pr}}[z = r] = \frac{exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}{\sum_{r \in \mathcal{R}_{\text{post}}} exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}.$$

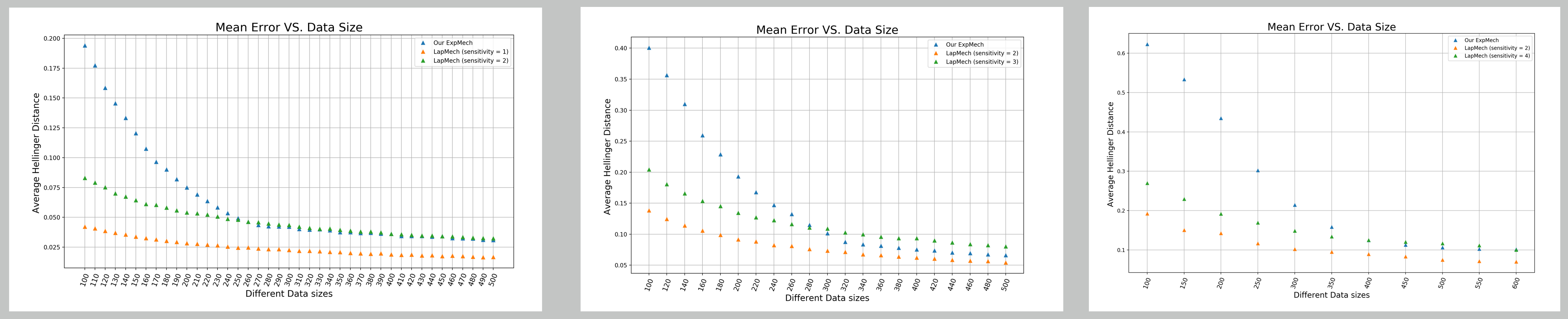
The smooth sensitivity is computed as follows:

$$S(\mathbf{x}) = \max_{\mathbf{x}' \in \{0,1\}^n} \left\{ \Delta_l \left(\mathcal{H}(\mathsf{BI}(\mathbf{x}'), \cdot) \right) \cdot e^{-\gamma \cdot d(C(\mathbf{x}), C(\mathbf{x}'))} \right\},$$

where d is the Hamming distance between two datasets, $\gamma = \gamma(\epsilon, \delta)$ is a function of ϵ and δ to be determined later, and where $\Delta_l \left(\mathcal{H}(\mathsf{BI}(\mathbf{x}'), \cdot) \right)$ denotes the local sensitivity at $\mathsf{BI}(\mathbf{x}')$, or equivalently at $\mathbf{x}'$, of the scoring function used in our mechanism. That is:

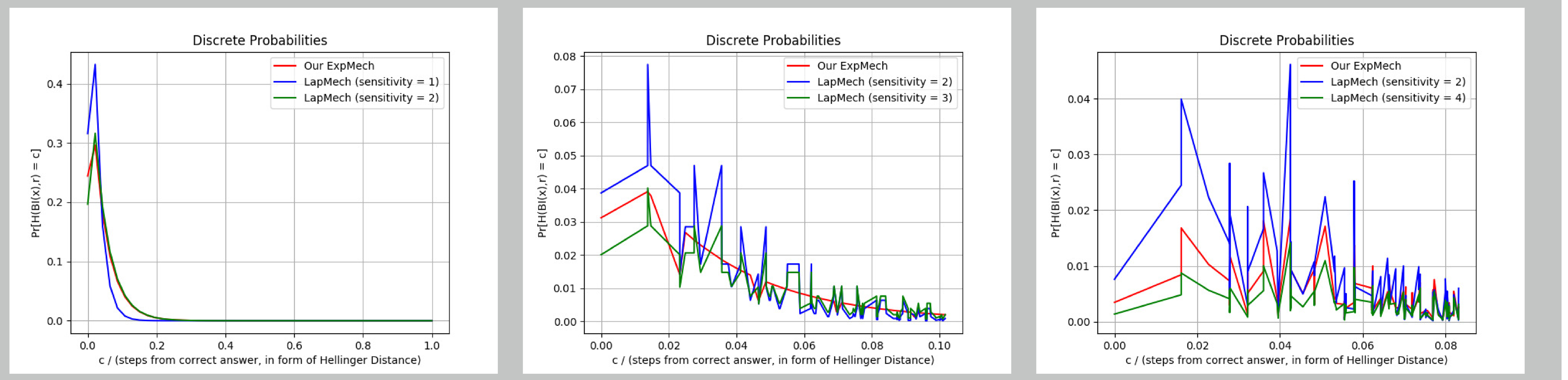
$$\Delta_l \left(\mathcal{H}(\mathsf{BI}(\mathbf{x}'), \cdot) \right) = \max_{\mathbf{x}'' \in \mathcal{X}^n : \mathsf{adj}(\mathbf{x}', \mathbf{x}''), r \in \mathcal{R}_{\text{post}}} |\mathcal{H}(\mathsf{BI}(\mathbf{x}'), r) - \mathcal{H}(\mathsf{BI}(\mathbf{x}''), r)|.$$

Some Experimental Results



(a) 2-dimensional, data size $\in [100, 500]$ (b) 3-dimensional, data size $\in [100, 500]$ (c) 4-dimensional, data size $\in [100, 600]$

Figure 2: Increasing data size with unit prior $\mathbf{beta}(1,1), \mathbf{beta}(1,1,1)$ and $\mathbf{beta}(1,1,1,1)$, balanced datasets and parameters $\epsilon = 0.8$ and $\delta = 10^{-8}$



(a) 2-dimensional (b) 3-dimensional (c) 4-dimensional

Figure 3: The concrete outputting probabilities under different dimensions with data set of size 600 ,unit prior $\mathbf{beta}(1,1), \mathbf{beta}(1,1,1)$ and $\mathbf{beta}(1,1,1,1)$, balanced datasets and parameters $\epsilon = 0.8$ and $\delta = 10^{-8}$

Conclusion

- Our the probabiliy measure approach outperforms the ℓ_1-norm approach when the Laplace noise cannot recognize the data to be protected is histogram and data size grow large.
-

References

 mbun@cs.princeton.edu, gaboardi@buffalo.edu