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.

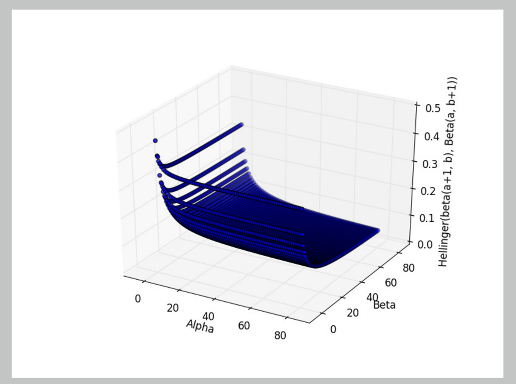


Figure 1: Hellinger Sensitivity

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 $\mathcal{M}^{B}_{\mathcal{H}}$:

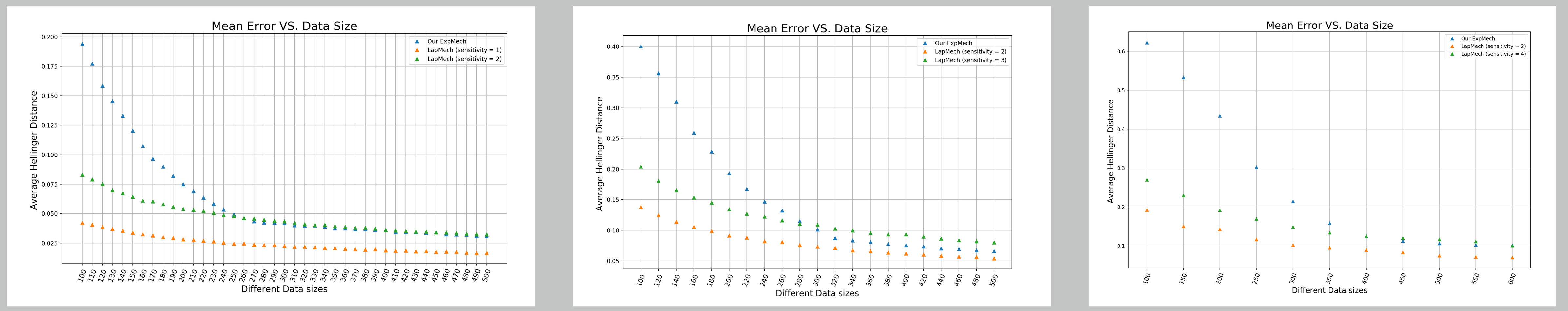
Producing an element r in $\mathcal{R}_{\text{post}}$ with: $\underset{z \sim \mathcal{M}^{B}_{\mathcal{H}}}{\mathbf{Pr}}[z = r] = \dfrac{exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}{\sum_{r \in \mathcal{R}_{\text{post}}} exp\left(\frac{-\epsilon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

(given in input an observations $\mathbf{x}$, parameters $\epsilon > 0$ and $\delta > 0$).

- $\mathcal{R}_{\text{post}}$, the candidates set defined as $\{\mathbf{beta}(\alpha', \beta') \mid \alpha' = \alpha + \Delta\alpha, \beta' = \beta + n - \Delta\alpha\}$, given the prior distribution $\beta_{\text{prior}} = \mathbf{beta}(\alpha, \beta)$ and observed data set size n.
- $-\mathcal{H}(\mathsf{BI}(\mathbf{x}), r)$, the scoring function instantiated by Hellinegr distance.
- $S(\mathbf{x})$, the smooth sensitivity: $S(\mathbf{x}) = \max_{\mathbf{x}' \in \{0,1\}^n} \left\{ LS(\mathbf{x}') \cdot e^{-\gamma \cdot d(\mathbf{x}, \mathbf{x}')} \right\}$, where
  - ▷ d: Hamming distance between two datasets,
  - ▷ $LS(\mathbf{x}')$, local sensitivity at $\mathbf{x}'$: $LS(\mathbf{x}) = \max_{\mathbf{x}' \in \mathcal{X}^n : \mathrm{adj}(\mathbf{x}, \mathbf{x}'), r \in \mathcal{R}} |\mathcal{H}(\mathsf{BI}(\mathbf{x}), r) - \mathcal{H}(\mathsf{BI}(\mathbf{x}'), r)|$,
  - ▷ $\gamma = \ln(1 - \frac{\epsilon}{2 \ln(\frac{\delta}{2(n+1)})})$ to ensure the (ϵ, δ)-differentially private.

Some Experimental Results

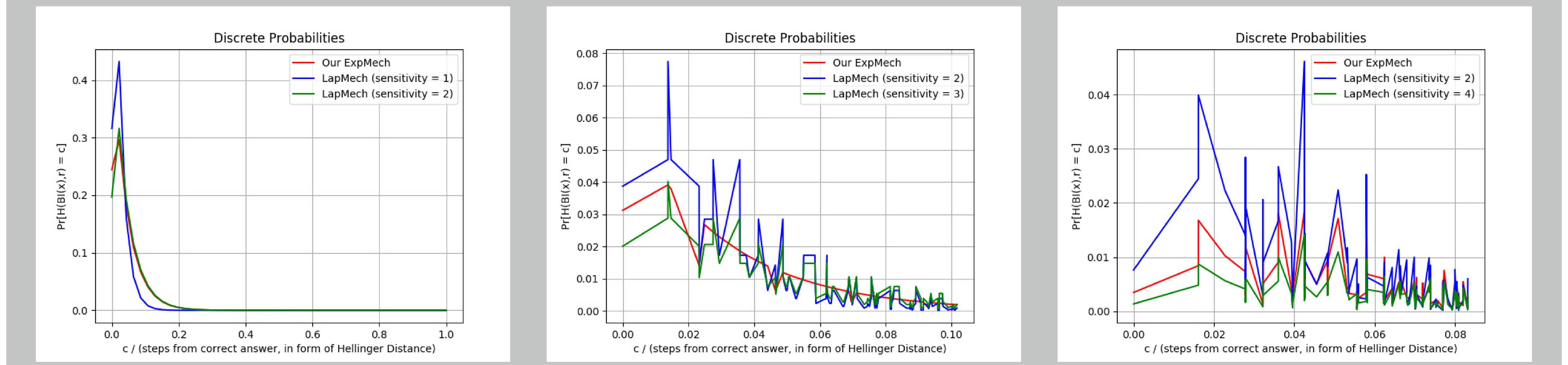
Two groups of experimental results both with unit prior $\mathbf{beta}(1, 1), \mathbf{beta}(1, 1, 1)$ and $\mathbf{beta}(1, 1, 1, 1)$, balanced datasets and parameters $\epsilon = 1.0$ and $\delta = 10^{-8}$.
Fig. 2 gives the average Hellinger distance between the sampled results and true posterior, by sampling for $10k$ times under each data size configuration. In baseline approach (i.e., Laplace mechanism), it is enough to add noise with sensitivity 1 in 2-dimensional and 2 in higher dimensional by equivalent to histogram case, which gives us the red points in plots. Without the knowledge of equivalence, Laplace usually add noise with sensitivity scale to dimensions, giving us green points.



(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

Fig. 3 give us the concrete probability of outputting candidates with certain distance.



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

Figure 3: The concrete outputting probabilities under different dimensions with data set of size 600

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