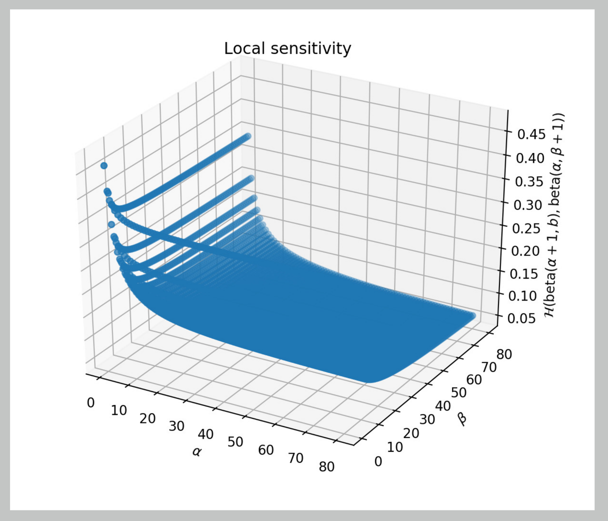
Tailoring Differentially Private Bayesian Inference to Distance Between Distributions

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

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

Objectives

1. Designing a differentially private Bayesian inference mechanism.
2. Measuring accuracy with a metric over distributions (Hellinger distance).
3. Calibrating the noise w.r.t sensitivity of the (Hellinger distance).
4. Applying smooth sensitivity in mechanism to achieve better accuracy. (Fig. 1 shows that local sensitivity of Hellinger distance is very steep and much higher in the edges but very smooth in central part).



say something about beta...

Figure 1: Local sensitivity of Hellinger

Bayesian inference: Beta-Binomial model

- Prior on θ : $\mathbf{beta}(\alpha, \beta), \alpha, \beta \in \mathbb{R}^{+}$, observed data set $\mathbf{x} = (x_1, \dots x_n) \in \{0, 1\}^n$, $n \in \mathbb{N}$.
- Likelihood function: $\mathbb{L}_{\mathbf{x}|\theta} = \theta^{\Delta\alpha}(1-\theta)^{n-\Delta\alpha}$, where $\Delta\alpha = \sum_{i=1}^{n} x_i$;
- Posterior distribution over theta: $\mathbb{P}_{\theta|\mathbf{x}} = \mathbf{beta}(\alpha + \Delta\alpha, \beta + n - \Delta\alpha)$.

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., $\widetilde{\Delta\alpha} \sim Lap(\Delta\alpha, \frac{2}{\epsilon})$. Explain parameters of baseline approach...

Smoothed Hellinger Distance Based Exponential Mechanism

Our approach defines the mechanism $\mathcal{M}^{B}_{\mathcal{H}}$, which outputs an element r in $\mathcal{R}_{\text{post}}$ with probaiblity:

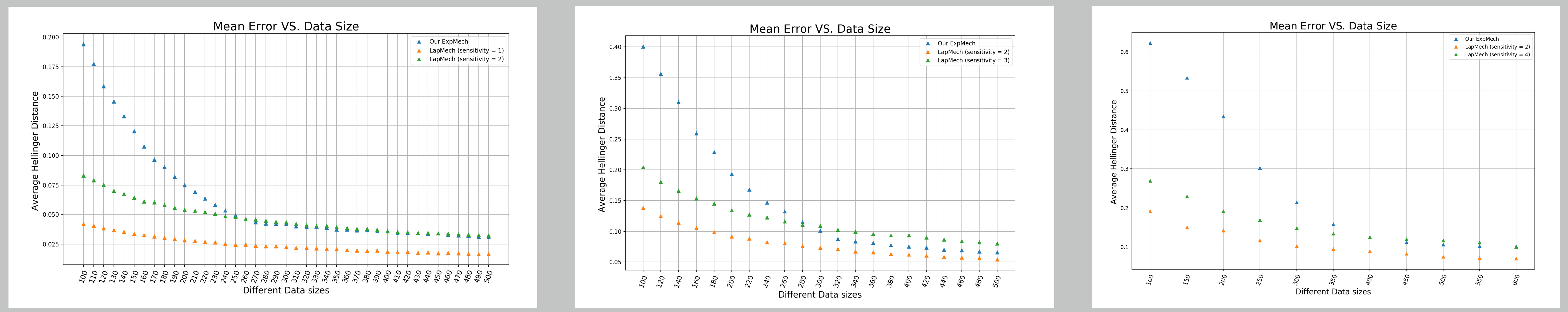
$$\Pr_{z \sim \mathcal{M}^{B}_{\mathcal{H}}}[z = r] = \frac{exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}{\sum_{r \in \mathcal{R}_{\text{post}}} exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}$$

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.
- $-\mathcal{H}(\mathsf{BI}(\mathbf{x}), r)$ denotes the scoring function based on the Hellinger 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 $\mathbf{x}'$: $LS(\mathbf{x}) = \max_{\mathbf{x}' \in \mathcal{X}^n : \mathbf{adj}(\mathbf{x}, \mathbf{x}'), r \in \mathcal{R}} |\mathcal{H}(\mathsf{BI}(\mathbf{x}), r) - \mathcal{H}(\mathsf{BI}(\mathbf{x}'), r)|$,
  - ▷ $\gamma = \ln(1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})})$ to ensure the (ϵ, δ)-differentially private.

Preliminary Experimental Results

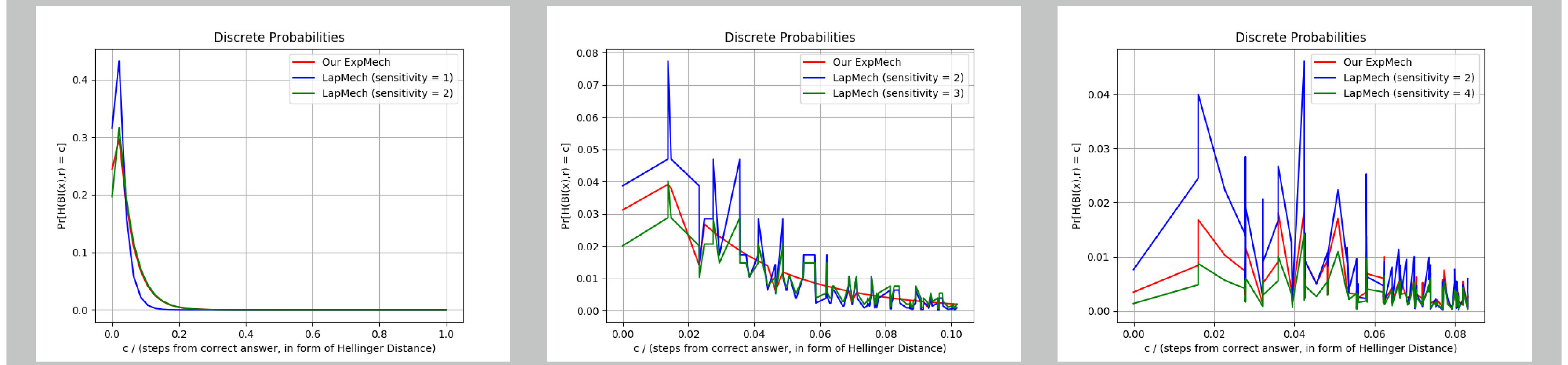
Fig. 2 gives the average Hellinger distance between the sampled results and true posterior, by sampling for $10k$ times under each data size configuration. In the baseline approach (i.e., Laplace mechanism), it is enough to add noise with sensitivity 1 in 2 dimensions and 2 in higher dimensions since it's equivalent to histogram (Should explain better). giving us the red points in plots. Without the knowledge of equivalence, Laplace usually add noise with sensitivity scale to dimensions, giving us green points. Points in blue are given by the M_B...



(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 gives us the concrete probabilities of outputting candidates with certain Hellinegr distance from the correct posterior in 2, 3 and 4 dimensions respectively.



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

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

Two groups of experiments 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 $\delta = 10^{-8}$.

Conclusion

- ▶ The smoothed Hellinger distance based exponential mechanism outperforms the asymptotically the baseline approach when the latter uses a sensitivity proportional to dimensionality. (when can this happen?)

References

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