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 (**dp**) Bayesian inference mechanism.
2. Measuring accuracy with a metric over distributions (Hellinger distance ($\mathcal{H}$)).

Bayesian inference (BI): Beta-Binomial model

- Prior on θ : $\mathbf{beta}(\alpha, \beta), \alpha, \beta \in \mathbb{R}^+$, observed data set $\mathbf{x} = (x_1, \ldots x_n) \in \{0,1\}^n, n \in \mathbb{N}$.
- 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 and Motivations

Releasing a differentially private posterior $\mathbf{beta}(\tilde{\alpha}, \tilde{\beta}) = (\alpha + \widetilde{\Delta\alpha}, \beta + n - \widetilde{\Delta\alpha})$.

1. Baseline approach samples noise from $\mathbf{Lap}(\Delta\alpha, \frac{1}{\epsilon})$ with mean $\Delta\alpha$ and scale $\frac{1}{\epsilon}$. But noise here is scaled to sensitivity $\mathbf{1}$ over ℓ_1 norm. Motivated by this, we calibrate the noise w.r.t sensitivity of the accuracy metric ($\mathcal{H}$) (v.s. the ℓ_1 norm in baseline approach).
2. Global sensitivity of $\mathcal{H}$ over **beta** distributions takes value at edge ($(\mathbf{0}, \mathbf{0})$ point) as in Fig. 1. But it's very smooth when move away from edge. Motivated by this, we apply smooth sensitivity in mechanism (v.s. global sensitivity) to improve accuracy.

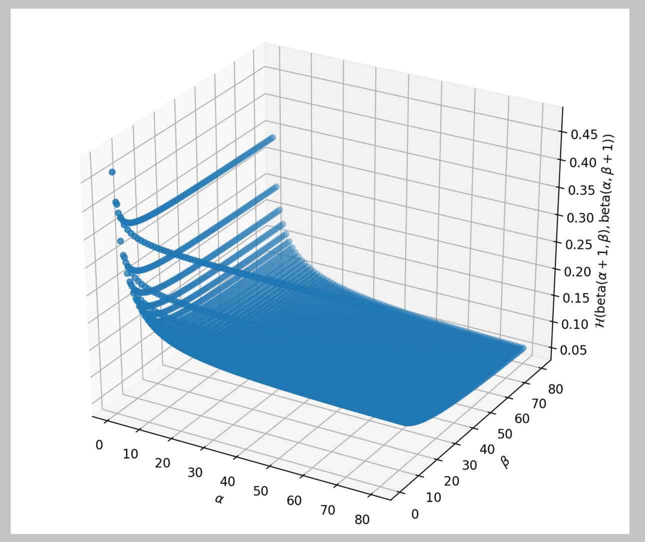


Figure 1: Sensitivities of $\mathcal{H}$ over **beta**

Smoothed Hellinger Distance Based Exponential Mechanism

We define the mechanism $\mathcal{M}^B_{\mathcal{H}}$ which produces an element r in $\mathcal{R}_{\text{post}}$ with probaiblity:

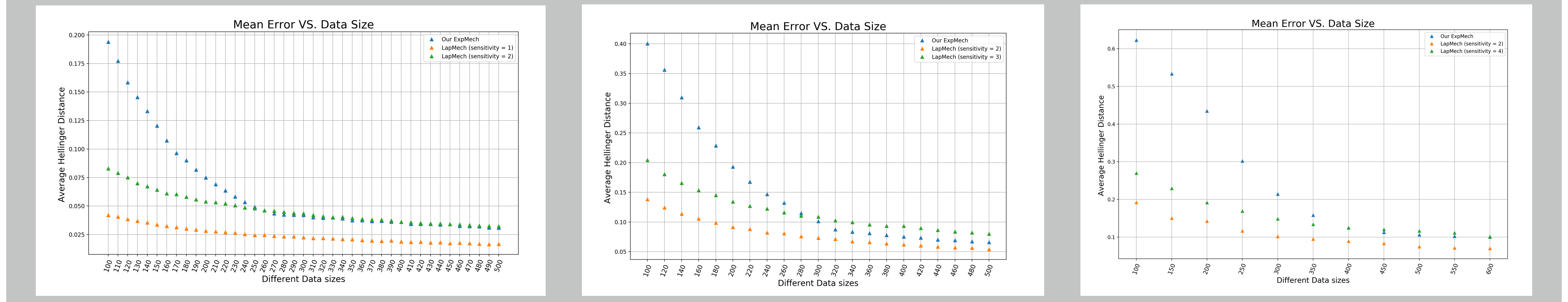
$$\underset{z \sim \mathcal{M}^B_{\mathcal{H}}}{\mathbf{Pr}}[z = r] = \frac{\exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}{\sum_{r \in \mathcal{R}_{\text{post}}} \exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}$$

given in input an observations $\mathbf{x}$, parameters $\epsilon > \mathbf{0}$ and $\delta > \mathbf{0}$, where:

- $\mathcal{R}_{\text{post}}$, 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[1]: $S(\mathbf{x}) = \max_{\mathbf{x}' \in \{0,1\}^n} \left\{ LS(\mathbf{x}') \cdot e^{-\gamma \cdot d(\mathbf{x}, \mathbf{x}')} \right\}$, where:
  - ▷ d: Hamming distance between two data sets,
  - ▷ $LS(\mathbf{x}')$, local sensitivity at $\mathbf{x}'$: $LS(\mathbf{x}) = \max_{\mathbf{x}' \in \mathcal{X}^n : \mathsf{adj}(\mathbf{x}, \mathbf{x}'), r \in \mathcal{R}} |\mathcal{H}(\mathsf{BI}(\mathbf{x}), r) - \mathcal{H}(\mathsf{BI}(\mathbf{x}'), r)|$,
  - ▷ $\gamma = \ln(1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})})$ to ensure the (ϵ, δ)-differentially private.

Preliminary Experimental Results

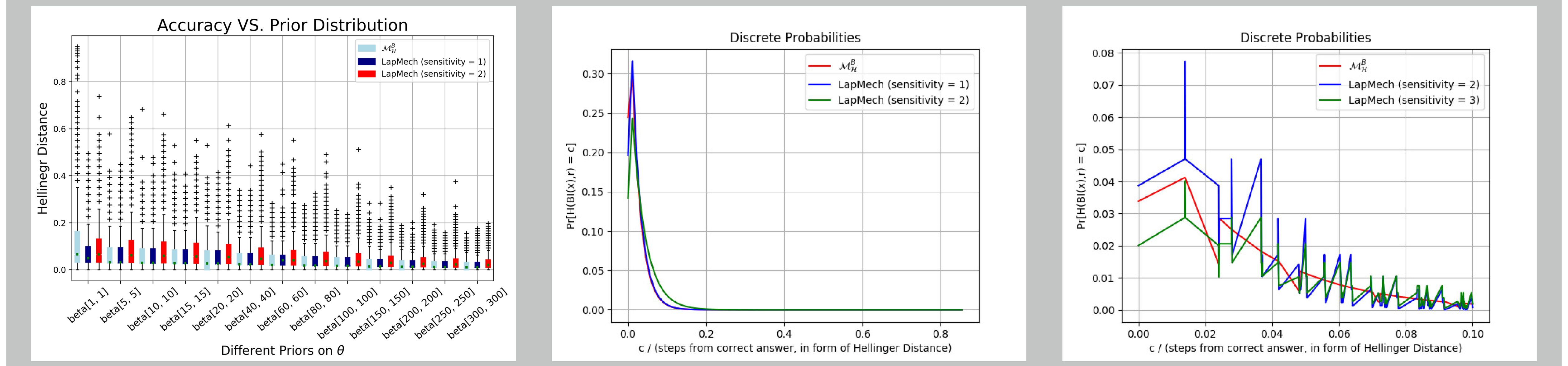
Fig. 2 and Fig. 3(a) gives the average and 4-quantile of Hellinger distance between the sampled results and true posterior, by sampling for $\mathbf{10}k$ times under each data size or prior distribution configuration. In the baseline approach (LapMech with noises being postprocessed to the closest integer), it is enough to add noise with sensitivity $\mathbf{1}$ in 2 dimensions and $\mathbf{2}$ in higher dimensions since it's equivalent to histogram (posteriors of adjacent data sets differ only in two dimensions), giving us the red points in plots. Without knowing equivalence, LapMech adds noise scaled to sensitivity proportional to dimensionality, giving us green points. Points in blue are given by the $\mathcal{M}^B_{\mathcal{H}}$.



(a) 2 dimensions, data size $\in [\mathbf{100}, \mathbf{500}]$ (b) 3 dimensions, data size $\in [\mathbf{100}, \mathbf{500}]$ (c) 4 dimensions, data size $\in [\mathbf{100}, \mathbf{600}]$

Figure 2: Average accuracy by increasing data set size

Fig. 3(b) and 3(c) gives us the concrete probabilities of outputting candidates with certain Hellinegr distance from the correct posterior in 2 and 3 respectively.



(a) 2 dimensions with data size $\mathbf{100}$ (b) 2 dimensions (c) 3 dimensions

Figure 3: The concrete outputting probabilities under different dimensions with data set of size $\mathbf{600}$

Experiments above are with unit prior $\mathbf{beta}(1,1), \mathbf{beta}(1,1,1)$ and $\mathbf{beta}(1,1,1,1)$ (except Fig. 3(a)) , balanced datasets, $\epsilon = \mathbf{1.0}$ and $\delta = \mathbf{10^{-8}}$.

Conclusion

- ▶ The smoothed Hellinger distance based exponential mechanism outperforms asymptotically the baseline approach when the latter uses a sensitivity proportional to dimensionality.
- ▶ From experimental results, the outperformance happens when the data size is greater than $\mathbf{400}$.

References

[1] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84. ACM, 2007.

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