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 differentially private Bayesian inference mechanism.
- Improve accuracy by calibrating noise to the sensitivity of a metric over distributions (e.g. Hellinger distance ($\mathcal{H}$), f-divergences, etc…).

An example of Bayesian inference: the Beta-Binomial model

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

Differentially private Bayesian inference

- Baseline approach:
  - Release $\mathbf{beta}(\alpha + \lfloor\widetilde{\Delta\alpha}\rfloor_0^n, \beta + n - \lfloor\widetilde{\Delta\alpha}\rfloor_0^n)$.
  - $\widetilde{\Delta\alpha} \sim \mathcal{L}(\Delta\alpha, \frac{\Delta\mathbf{BI}}{\epsilon})$.
  - $\Delta\mathbf{BI} \equiv \max_{\mathbf{x},\mathbf{x}' \in \{0,1\}^n, ||\mathbf{x}-\mathbf{x}'||_1 \le 1} ||\mathbf{BI}(\mathbf{x}) - \mathbf{BI}(\mathbf{x}')||_1$.
  - Measure accuracy with a metric over distributions. E.g. $\mathcal{H}(f,g)^2 \equiv 1 - \int(\sqrt{f(x)g(x)}\,\mathrm{d}x)$ (f, g densities).

But $\Delta\mathbf{BI}$ grows linearly with the dimension: too noisy when we generalize to Dirichlet-Multinomial ($\mathbf{DL}(\cdot)$) model.

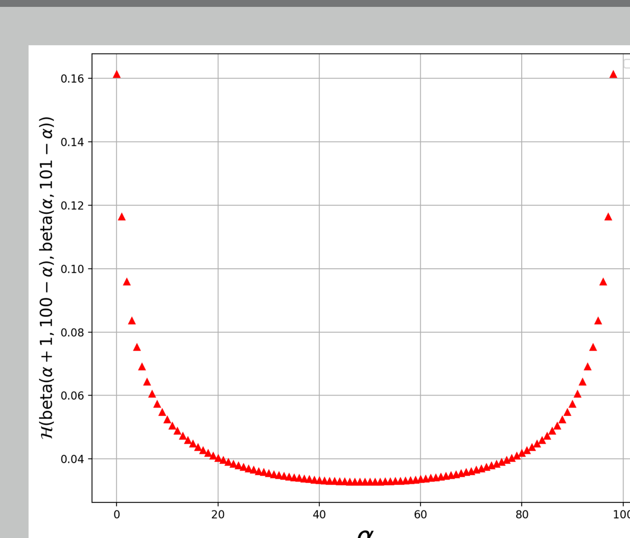


Figure 1: Sensitivity of $\mathcal{H}$. There is a gap between Global and Local sensitivity.

- Another approach:
  - Calibrate noise w.r.t *global* sensitivity of $\mathcal{H}$: but global sensitivity is still too big.
  - Fig. 1 shows that there is a gap between global and local sensitivity of $\mathcal{H}$.
- A different approach:
  - Calibrate noise w.r.t. the *smooth* sensitivity of $\mathcal{H}$.

Our approach: smoothed Hellinger distance based exponential mechanism

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

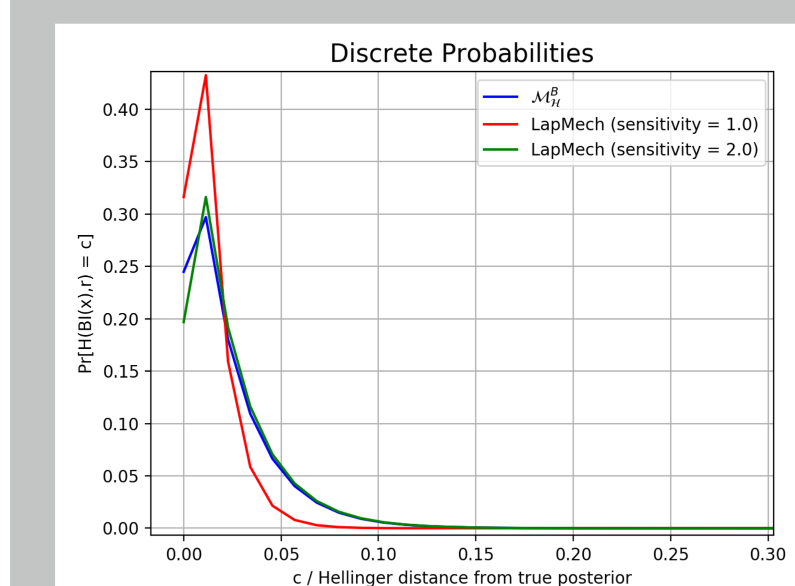
$$\mathbb{P}_{r \sim \mathcal{M}_\mathcal{H}} = \frac{\exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathbf{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}{\sum_{r \in \mathcal{R}_{\text{post}}} \exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathbf{BI}(\mathbf{x}), r)}{2 \cdot S(\mathbf{x})}\right)}$$

- $\mathcal{R}_{\text{post}} \equiv \{\mathbf{beta}(\alpha', \beta') \mid \alpha' = \alpha + \Delta\alpha, \beta' = \beta + n - \Delta\alpha\}$. With prior distribution $\beta_{\text{prior}} = \mathbf{beta}(\alpha, \beta)$.
- $-\mathcal{H}(\mathbf{BI}(\mathbf{x}), r)$ denotes the scoring function.
- $S(\mathbf{x}) \equiv \max_{\mathbf{x}' \in \{0,1\}^n} \left\{LS(\mathbf{x}') \cdot e^{-\gamma \cdot d(\mathbf{x}, \mathbf{x}')}\right\}$: smooth sensitivity[1], d is the Hamming distance.
- $LS(\mathbf{x}') \equiv \max_{y \in \mathcal{X}^n : \mathrm{adj}(y, \mathbf{x}'), r \in \mathcal{R}} |\mathcal{H}(\mathbf{BI}(y), r) - \mathcal{H}(\mathbf{BI}(\mathbf{x}'), r)|$ is the local sensitivity of $\mathbf{x}'$, $\gamma = \ln(1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})})$.

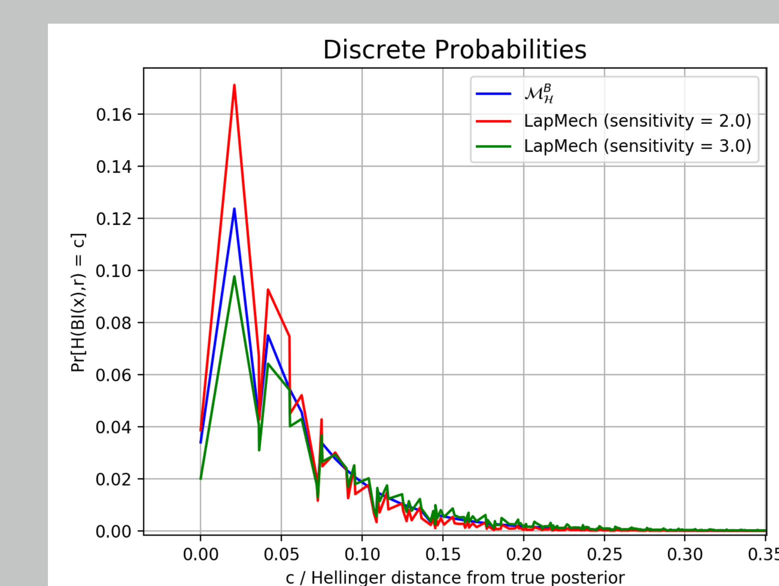
Preliminary experimental results

Experiments are about three mechanisms and plotted as follows:

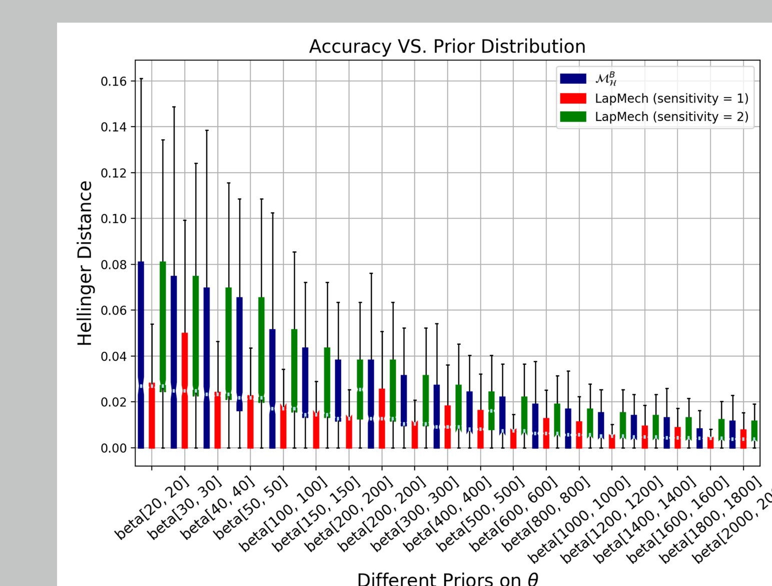
- **Green**: Baseline approach.
- **Red**: Improved approach by using sensitivity 1 in 2 dimensions and 2 in higher dimensions. Indeed: we can see the output of the Bayesian inference as a histogram, and $||\mathbf{BI}(\mathbf{x}) - \mathbf{BI}(\mathbf{x}')||_1 \le 2$.
- **Blue**: $\mathcal{M}_\mathcal{H}$. The fact that there is only one candidate distribution which achieves the highest score and different distributions which achieve a sub-optimal score explains the (highest) peaks in Fig. 2(a) (and Fig. 2(b)).



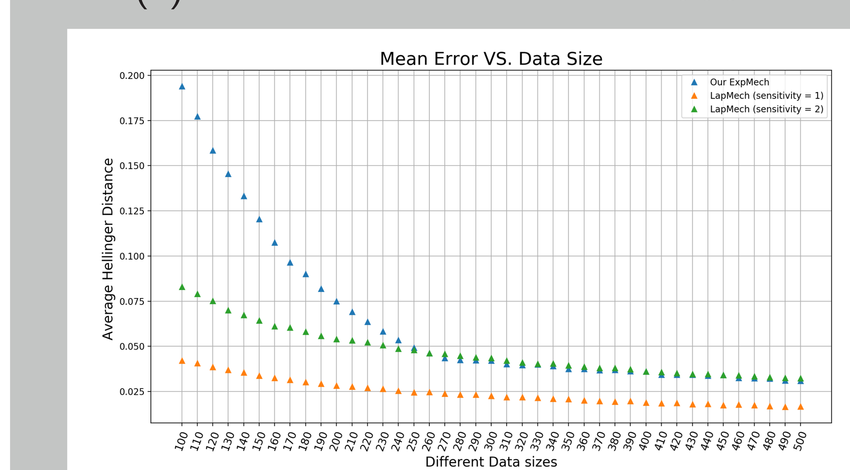
(a) 2 dimensions with data size $\mathbf{600}$



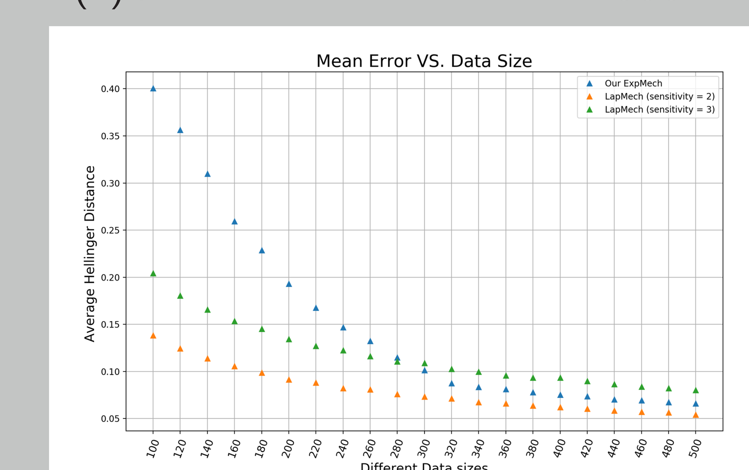
(b) 3 dimensions with data size $\mathbf{600}$



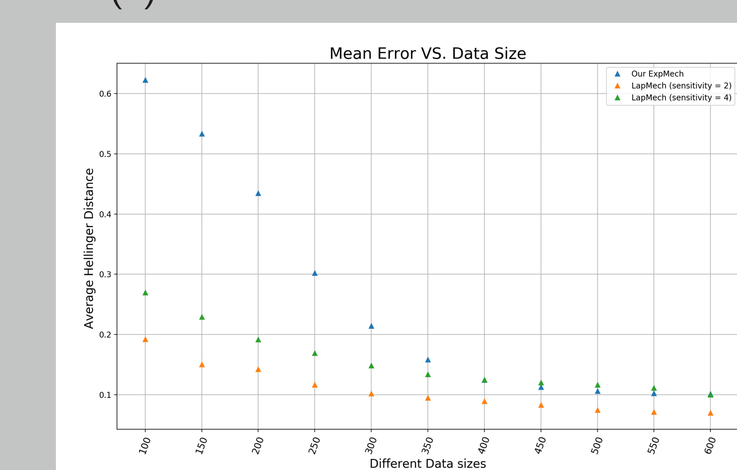
(c) 2 dimensions with data size $\mathbf{100}$



(d) 2 dimensions, data size $\in [100, 500]$



(e) 3 dimensions, data size $\in [100, 500]$



(f) 4 dimensions, data size $\in [100, 600]$

Figure 2: Priors are $\mathbf{beta}(1,1), \mathbf{DL}(1,1,1)$ and $\mathbf{DL}(1,1,1,1)$ (except for Figure 2(c)) , balanced datasets, $\epsilon = \mathbf{1.0}$ and $\delta = \mathbf{10^{-8}}$.

Conclusion

- $\mathcal{M}_\mathcal{H}$ outperforms the baseline approach but not the improved one, for priors with small parameters.
- When the prior parameters increase $\mathcal{M}_\mathcal{H}$ is comparable with the improved baseline approach.

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, {gianpiet,gaboardi,jliu223}@buffalo.edu