

# Tailoring Differentially Private Bayesian Inference to Distance Between Distributions

Jiawen Liu<sup>\*</sup>, Mark Bun<sup>\*\*</sup>, Gian Pietro Farina<sup>\*</sup>, and Marco Gaboardi<sup>\*</sup>

<sup>\*</sup>Department of Computer Science and Engineering, University at Buffalo, SUNY. {jliu223,gianpiet,gaboardi}@buffalo.edu

<sup>\*\*</sup>Department of Computer Science, Princeton University.  
mbun@cs.princeton.edu

## Abstract

Bayesian inference is a statistical method which allows one to derive a *posterior* distribution, starting from a *prior* distribution and observed data. Several approaches have been explored in order to make this process differentially private. For example, [6], and [13] proved that, under specific conditions, sampling from the posterior distribution is already differentially private. [17], [9], designed differentially private mechanisms that output a representation of the full posterior distribution.

When the output of a differentially private mechanism is a probability distribution, accuracy is naturally measured by means of *probabilistic distances* measuring how far this distribution is from the original one. Some classical examples are total variation distance, Hellinger distance,  $\chi^2$ -distance, KL-divergence, etc.

In this work, we design a mechanism for bayesian inference exploring the idea of calibrating noise using the same probabilistic distance we want to measure accuracy with. We focus on two discrete models, the Beta-Binomial and the Dirichlet-multinomial models, and one probability distance, Hellinger distance. Our mechanism can be understood as a version of the exponential mechanism where the noise is calibrated to the smooth sensitivity of the utility function, rather than to its global sensitivity. In our setting, the utility function is the probability distance we want to use to measure accuracy. To show the usefulness of this mechanism we show an experimental analysis comparing it with mechanisms based on the Laplace mechanism.

## 1 Introduction

Data analysis techniques are broadly used in various applications in different areas to improve their services, including disease-medicine service, financial ser-

vice, location service, social network and so on. In order to provide a better service, large of data are collected from users for analysis. As a consequence, the data privacy came to be a critical issue. Sensitivity information in data can be revealed through the analysis results. The key challenge here is to release a private analysis results, from which adversary cannot observe individual's sensitive information in data.

Plenty of work have been developed to solve this issue, guaranteeing the privacy in specific data analysis algorithms. They achieved the  $\epsilon$ -differential privacy by adopting either Laplace mechanism or achieved the  $(\epsilon, \delta)$ -differential privacy. But they are not giving better accuracy than the differential privacy mechanism itself. Here, we are proposing mechanism with better accuracy.

Our work is conducted under a Bayesian inference scenario, where the posterior distribution is the analysis result we obtained from the data. Publishing the posterior distribution inferred from a sensitive dataset can leak information about the individuals in the dataset. In order to guarantee differential privacy and to protect the individuals' data we can add noise to the posterior before releasing it. The amount of the noise that we need to introduced depends on the privacy parameter  $\epsilon$  and the sensitivity of the inference to small changes in the data set. Sensitivity can be computed in many different ways based on which metric space we consider on the output set of the mechanism. In the literature on private Bayesian inference ([17, 15]), it is only measured with respect to the vector of numbers parametrizing the output distribution using, e.g. the  $\ell_1$  norm. A more natural approach which we explore here, is to measure sensitivity with respect to a metric on the space of inferred probability distributions. A re-loved question is that of how to measure accuracy. Again, this can be answered in different ways based on the metric imposed on the output space, and yet again only in few works in literature (e.g. [17]) distances between probability measures have been used for these purposes.

The question that this work aims at answering is whether an approach based on probability metrics can improve on the accuracy of approaches based on metrics over the numeric parameters of the distributions. We will see that in some cases this can happen.

#### **Main contributions.**

- We designed a differentially private Bayesian inference mechanism based on the standard exponential mechanism.
- The accuracy is improved by two ways: 1) calibrating noise to the sensitivity of a metric over distributions (e.g. Hellinger distance ( $\mathcal{H}$ ),  $f$ -divergences, etc. . . ). 2) Using a smooth upper bound on the local sensitivity and scale the noise to this smooth bound rather than global sensitivity, to improve the mechanism accuracy.
- A full proof on applying exponential mechanism to the smooth sensitivity is  $(\epsilon, \delta)$ -differential privacy is given in paper.
- We implemented the new proposed mechanism and other art-of-state mechanisms, comparing the performance in terms of accuracy and efficiency.

### Related Work.

A plentiful of data analysis algorithms have been studied to preserve differential privacy, including the subspace clustering algorithm [12], the gradient decent algorithm in deep learning [1], logical regression [3], principle component analysis [4], probabilistic inference [14] and convergence in statistic estimation [2], etc.

In Bayesian Inference data analysis, mechanisms are proposed corresponded to maintain their differential privacy, focusing on 3 different goals: 1) Inherited differential privacy property of posterior sampling in Bayesian inference. [6], [17], [18] and [13]. 2) Data sampled and released from posterior distribution of Bayesian is differentially private [16], [7], [9]. 3) The inference process is differentially private and the posterior distribution released should be private itself, in the meantime, with good accuracy. The third topic where our work focus on is still very new.

## 2 Preliminaries

### Bayesian Inference.

Given a prior belief  $\Pr(\theta)$  on some parameter  $\theta$ , and an observation  $\mathbf{x}$ , the posterior distribution on  $\theta$  given  $\mathbf{x}$  is computed as:

$$\Pr(\theta|\mathbf{x}) = \frac{\Pr(\mathbf{x}|\theta) \cdot \Pr(\theta)}{\Pr(\mathbf{x})}$$

where the expression  $\Pr(\mathbf{x}|\theta)$  denotes the *likelihood* of observing  $\mathbf{x}$  under a value of  $\theta$ . Since we consider  $\mathbf{x}$  to be fixed, the likelihood is a function of  $\theta$ . For the same reason  $\Pr(\mathbf{x})$  is a constant independent of  $\theta$ . Usually in statistics the prior distribution  $\Pr(\theta)$  is chosen so that it represents the initial belief on  $\theta$ , that is, when no data has been observed. In practice though, prior distributions and likelihood functions are usually chosen so that the posterior belongs to the same *family* of distributions. In this case we say that the prior is conjugate to the likelihood function. Use of a conjugate prior simplifies calculations and allows for inference to be performed in a recursive fashion over the data.

### Beta-binomial System.

In this work we will consider a specific instance of Bayesian inference and one of its generalizations. specifically, a Beta-binomial mode. We will consider the situation the underlying data is binomial distribution ( $\sim \text{binomial}(\theta)$ ), where  $\theta$  represents the parameter –informally called *bias*– of a Bernoulli distributed random variable. The prior distribution over  $\theta \in [0, 1]$  is going to be a beta distribution,  $\text{beta}(\alpha, \beta)$ , with parameters  $\alpha, \beta \in \mathbb{R}^+$ , and with p.d.f:

$$\Pr(\theta) \equiv \frac{\theta^\alpha (1 - \theta)^\beta}{B(\alpha, \beta)}$$

where  $B(\cdot, \cdot)$  is the beta function. The data  $\mathbf{x}$  will be a sequence of  $n \in \mathbb{N}$  binary values, that is  $\mathbf{x} = (x_1, \dots, x_n)$ ,  $x_i \in \{0, 1\}$ , and the likelihood function is:

$$\Pr(\mathbf{x}|\theta) \equiv \theta^{\Delta\alpha} (1 - \theta)^{n - \Delta\alpha}$$

where  $\Delta\alpha = \sum_{i=1}^n x_i$ . From this it can easily be derived that the posterior distribution is:

$$\Pr(\theta|\mathbf{x}) = \text{beta}(\alpha + \Delta\alpha, \beta + n - \Delta\alpha)$$

### Dirichlet-multinomial Systems.

The beta-binomial model can be immediately generalized to Dirichlet-multinomial, with underlying data multinomially distributed. The *bias* is represented by parameter  $\theta$ , the vector of parameters of a categorically distributed random variable. The prior distribution over  $\theta \in [0, 1]^k$  is given by a Dirichlet distribution,  $\text{DL}(\alpha)$ , for  $k \in \mathbb{N}$ , and  $\alpha \in (\mathbb{R}^+)^k$ , with p.d.f:

$$\Pr(\theta) \equiv \frac{1}{B(\alpha)} \cdot \prod_{i=1}^k \theta_i^{\alpha_i - 1}$$

where  $B(\cdot)$  is the generalized beta function. The data  $\mathbf{x}$  will be a sequence of  $n \in \mathbb{N}$  values coming from a universe  $\mathcal{X}$ , such that  $|\mathcal{X}| = k$ . The likelihood function will be:

$$\Pr(\mathbf{x}|\theta) \equiv \prod_{a_i \in \mathcal{X}} \theta_i^{\Delta\alpha_i},$$

with  $\Delta\alpha_i = \sum_{j=1}^n [x_j = a_i]$ , where  $[\cdot]$  represents Iverson bracket notation. Denoting by  $\Delta\alpha$  the vector  $(\Delta\alpha_1, \dots, \Delta\alpha_k)$  the posterior distribution over  $\theta$  turns out to be

$$\Pr(\theta|\mathbf{x}) = \text{DL}(\alpha + \Delta\alpha).$$

where  $+$  denotes the componentwise sum of vectors of reals.

### Differential Privacy.

#### Definition 1. $\epsilon$ -differential privacy.

A randomized mechanism  $\mathcal{M} : \mathcal{X} \rightarrow \mathcal{Y}$  is differential privacy, iff for any adjacent<sup>1</sup> input  $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ , a metric  $H$  over  $\mathcal{Y}$  and a  $B \subseteq H(\mathcal{Y})$ ,  $\mathcal{M}$  satisfies:

$$\mathbb{P}[H(\mathcal{M}(\mathbf{x})) \in B] = e^\epsilon \mathbb{P}[H(\mathcal{M}(\mathbf{x}')) \in B].$$

#### Definition 2. $(\epsilon, \delta)$ -differential privacy.

A randomized mechanism  $\mathcal{M} : \mathcal{X} \rightarrow \mathcal{Y}$  is differential privacy, iff for any  $\text{adj}(\mathbf{x}, \mathbf{x}') \in \mathcal{X}$ , a metric  $H$  over  $\mathcal{Y}$  and a  $B \subseteq H(\mathcal{Y})$ ,  $\mathcal{M}$  satisfies:

$$\mathbb{P}[H(\mathcal{M}(\mathbf{x})) \in B] = e^\epsilon \mathbb{P}[H(\mathcal{M}(\mathbf{x}')) \in B] + \delta,$$

where  $\mathbf{x} = (x_i)_{i=1}^n$  and  $\mathbf{x}' = (x'_i)_{i=1}^n$  is adjacent if there is only one  $j$  that  $x_j \neq x'_j$  and  $x_i = x'_i$  for  $i = 1, 2, \dots, n; i \neq j$ .

---

<sup>1</sup>Given  $\mathbf{x}, \mathbf{x}'$  we say that  $\mathbf{x}$  and  $\mathbf{x}'$  are adjacent and we write,  $\text{adj}(\mathbf{x}, \mathbf{x}')$ , iff  $\sum_i^n [x_i = x'_i] \leq 1$ .

### 3 Technical Problem Statement and Motivations

We are interested in designing a mechanism for privately releasing the full posterior distributions derived in section 2, as opposed to just sampling from them. It's worth noticing that the posterior distributions are fully characterized by their parameters, and the family (beta, Dirichlet) they belong to. Hence, in case of the Beta-Binomial model we are interested in releasing a private version of the pair of parameters  $(\alpha', \beta') = (\alpha + \Delta\alpha, \beta + n - \Delta\alpha)$ , and in the case of the Dirichlet-multinomial model we are interested in a private version of  $\alpha' = (\alpha + \Delta\alpha)$ . [17] and [15] have already attacked this problem by adding independent Laplacian noise to the parameters of the posteriors. That is, in the case of the Beta-Binomial system, the value released would be:  $(\tilde{\alpha}, \tilde{\beta}) = (\alpha + \widetilde{\Delta\alpha}, \beta + n - \widetilde{\Delta\alpha})$  where  $\widetilde{\Delta\alpha} \sim \mathcal{L}(\Delta\alpha, \frac{2}{\epsilon})$ , and where  $\mathcal{L}(\mu, \nu)$  denotes a Laplace random variable with mean  $\mu$  and scale  $\nu$ . This mechanism is  $\epsilon$ -differentially private, and the noise is calibrated w.r.t. to a sensitivity of 2 which is derived by using  $\ell_1$  norm over the pair of parameters. Indeed, considering two adjacent data observations  $\mathbf{x}, \mathbf{x}'$ , that, from a unique prior, give rise to two posterior distributions, characterized by the pairs  $(\alpha', \beta')$  and  $(\alpha'', \beta'')$  then  $|\alpha' - \alpha''| + |\beta' - \beta''| \leq 2$ . This argument extends similarly to the Dirichlet-Multinomial system. Details are introduced in Sec. 4.1.

However, in previous works, the accuracy of the posterior was measured again with respect to  $\ell_1$  norm. That is, an upper bound was given on

$$\Pr[|\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| \geq \gamma]$$

where  $(\alpha, \beta), (\tilde{\alpha}, \tilde{\beta})$  are as defined above. This accuracy metric is meaningless when the results released are distributions rather than numerical values. In contrast, distribution metrics such as  $f$ -divergence, Hellinger distance, etc. come into mind overtly when we are measuring distance between distributions. This gives us motivation on using a different norm (a distribution metric) to compute the sensitivity and provide guarantees on the accuracy.

Specifically, we will use the Hellinger distance  $\mathcal{H}(\cdot, \cdot)$ : Given two beta distributions  $\beta_1 = \text{beta}(\alpha_1, \beta_1)$ , and  $\beta_2 = \text{beta}(\alpha_2, \beta_2)$  the following equality holds

$$\mathcal{H}(\beta_1, \beta_2) = \sqrt{1 - \frac{B(\frac{\alpha_1 + \alpha_2}{2}, \frac{\beta_1 + \beta_2}{2})}{\sqrt{B(\alpha_1, \beta_1)B(\alpha_2, \beta_2)}}}$$

Our choice to use Hellinger distance is motivated by three facts:

- It simplifies calculations in the case of the probabilistic models considered here.
- It also automatically yields bounds on the total variation distance, which represents also the maximum advantage an unbounded adversary can have in distinguishing two distributions.
- The accuracy can be improved by using a smooth bound on Hellinger distance's local sensitivity. As shown in Fig. 1, taking advantage of the

gap between the global sensitivity and local sensitivity, we can improve the accuracy by applying a smooth upper bound on local sensitivity instead of using global sensitivity.

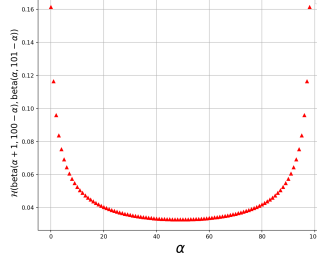


Figure 1: Sensitivity of  $\mathcal{H}$

## 4 Mechanism Proposition

Given a prior distribution  $\beta_{\text{prior}} = \text{beta}(\alpha, \beta)$  and a sequence of  $n$  observations  $\mathbf{x} \in \{0, 1\}^n$ , we define the following set:

$$\mathcal{R}_{\text{post}} \equiv \{\text{beta}(\alpha', \beta') \mid \alpha' = \alpha + \Delta\alpha, \beta' = \beta + n - \Delta\alpha\},$$

where  $\Delta\alpha$  is as defined in Section 2. Notice that  $\mathcal{R}_{\text{post}}$  has  $n + 1$  elements, and the Bayesian Inference process will produce an element from  $\mathcal{R}_{\text{post}}$  that we denote by  $\text{BI}(\mathbf{x})$  – we don't explicitly parametrize the result by the prior, which from now on we consider fixed and we denote it by  $\beta_{\text{prior}}$ .

### 4.1 Baseline Mechanisms

Baseline Mechanisms are introduced in prior to our mechanism:  $\mathcal{M}_{\mathcal{H}}$ .

#### 4.1.1 Exponential Mechanism

Exponential mechanism  $\mathcal{M}_E(x, u, \mathcal{R}_{\text{post}})$  samples an element from the candidate set  $\mathcal{R}_{\text{post}} = \{r_1, r_2, \dots, r_n\}$  with probability proportional to  $\exp(\frac{\epsilon u(x, r)}{2GS})$ :

$$\Pr_{z \sim \mathcal{M}_E(x, u, \mathcal{R}_{\text{post}})}[z = r] = \frac{\exp(\frac{\epsilon u(x, r)}{2GS})}{\sum_{r' \in \mathcal{R}} \exp(\frac{\epsilon u(x, r')}{2GS})},$$

where  $u(x, r)$  is the Hellinger scoring function over candidates,  $-\mathcal{H}(\text{BI}(\mathbf{x}), r)$ , and  $GS$  is the global sensitivity calculated by:

$$GS = \max_{\{\mathbf{x}, \mathbf{x}' \mid |\mathbf{x} - \mathbf{x}'| \leq 1\}} \max_{\{\mathbf{x}' \in \mathcal{X}^n\}} \max_{\{r \in \mathcal{R}\}} |\mathcal{H}(\text{BI}(\mathbf{x}), r) - \mathcal{H}(\text{BI}(\mathbf{x}'), r)|$$

Exponential mechanism is  $\epsilon$ -differential privacy [8].

#### 4.1.2 Exponential Mechanism with Local Sensitivity

Exponential mechanism with local sensitivity  $\mathcal{M}_E^{local}(x, u, \mathcal{R}_{\text{post}})$  share the same candidate set and utility function as it with standard exponential mechanism. This outputs a candidate  $r \in \mathcal{R}$  with probability proportional to  $\exp(\frac{\epsilon u(x, r)}{2LS(\mathbf{x})})$ :

$$\Pr_{z \sim \mathcal{M}_E^{local}(x, u, \mathcal{R}_{\text{post}})}[z = r] = \frac{\exp(\frac{\epsilon u(x, r)}{2LS(\mathbf{x})})}{\sum_{r' \in \mathcal{R}} \exp(\frac{\epsilon u(x, r')}{2LS(\mathbf{x})})},$$

where  $LS(\mathbf{x})$  is the local sensitivity calculated by:

$$LS(\mathbf{x}) = \max_{\mathbf{x}' \in \mathcal{X}^n: \text{adj}(\mathbf{x}, \mathbf{x}'), r \in \mathcal{R}} |\mathcal{H}(\text{BI}(\mathbf{x}'), r) - \mathcal{H}(\text{BI}(\mathbf{x}), r)|.$$

The exponential mechanism with local sensitivity is non-differential privacy[8].

#### 4.1.3 Baseline Mechanism - Laplace Mechanism

Adding noise to the posterior distribution parameters directly, through Laplace mechanism ( $\mathcal{L}(\cdot, \cdot)$ ) with post-processing:

$$\text{beta}(\alpha + \lfloor \Delta\alpha + Y \rfloor_0^n, \beta + n - \lfloor \Delta\alpha + Y \rfloor_0^n),$$

where  $Y \sim \mathcal{L}(0, \frac{\Delta \text{BI}}{\epsilon})$  in Beta-binomial model; and

$$\text{DL}(\alpha_1 + \lfloor \Delta\alpha_1 + Y_1 \rfloor_0^n, \dots, \alpha_k + \lfloor n - \sum_{i=1}^{k-1} \lfloor \Delta\alpha_i + Y_i \rfloor_0^n \rfloor_0^n),$$

where  $Y_i \sim \mathcal{L}(0, \frac{\Delta \text{BI}}{\epsilon})$  in Dirichlet-multinomial model.  $\lfloor \cdot \rfloor_0^n$  is taking the floor value and truncating into  $[0, n]$  to make sure the noised posterior is valid.

Then release it as the private posterior distribution.

The sensitivity used in this baseline mechanism is:

$$\Delta \text{BI} \equiv \max_{\mathbf{x}, \mathbf{x}' \in \{0,1\}^n, \|\mathbf{x} - \mathbf{x}'\|_1 \leq 1} \|\text{BI}(\mathbf{x}) - \text{BI}(\mathbf{x}')\|_1,$$

which is proportional to the dimensionality.

#### 4.1.4 Improved Laplace Mechanism

Noise added to posterior distribution parameters are scaled to smaller sensitivity in this improved Laplace mechanism. Because in terms of two adjacent data sets  $\mathbf{x}, \mathbf{x}'$ , their posterior distributions by Bayesian inference –  $\text{BI}(\mathbf{x}), \text{BI}(\mathbf{x}')$  – which parameter differs at most in 2 dimensions even though extended to Dirichlet-multinomial mode, i.e.,  $\Delta \text{BI} \leq 2$ .

Then it is enough to use sensitivity 1 in 2 dimensions and 2 in higher dimensions:

$$\text{beta}(\alpha + \lfloor \Delta\alpha + Y \rfloor_0^n, \beta + n - \lfloor \Delta\alpha + Y \rfloor_0^n),$$

where  $Y \sim \mathcal{L}(0, \frac{1}{\epsilon})$  in Beta-binomial model; and

$$\text{DL}(\alpha_1 + \lfloor \Delta\alpha_1 + Y_1 \rfloor_0^n, \dots, \alpha_k + \lfloor n - \sum_{i=1}^{k-1} \lfloor \Delta\alpha_i + Y_i \rfloor_0^n \rfloor_0^n),$$

where  $Y_i \sim \mathcal{L}(0, \frac{2}{\epsilon})$  in Dirichlet-multinomial model.

Both Laplace mechanism and improved one are  $\epsilon$ -differential privacy[8].

## 4.2 $\mathcal{M}_{\mathcal{H}}$ : Smoothed Hellinger Distance Based Exponential Mechanism

**Definition 3.** The mechanism  $\mathcal{M}_{\mathcal{H}}(\mathbf{x})$  outputs a candidate  $r \in \mathcal{R}_{\text{post}}$  with probability

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

where  $S_{\beta}(\mathbf{x})$  is the smooth sensitivity of  $\mathcal{H}(\text{Bl}(\mathbf{x}), -)$ , calculated by:

$$S(\mathbf{x}) = \max_{\mathbf{x}' \in \{0,1\}^n} \left\{ LS(\mathbf{x}') \cdot e^{-\gamma \cdot d(\mathbf{x}, \mathbf{x}')}\right\}, \quad (1)$$

where  $d$  is the Hamming distance between two datasets, and  $\beta = \beta(\epsilon, \delta)$  is a function of  $\epsilon$  and  $\delta$ .

This mechanism is based on the basic exponential mechanism [10], with  $\mathcal{R}_{\text{post}}$  as the range and  $\mathcal{H}(\cdot, \cdot)$  as the scoring function. The difference is that in this mechanism we don't calibrate the noise w.r.t. to the global sensitivity of the scoring function but w.r.t. to the smooth sensitivity  $S(\mathbf{x})$  – defined by [11]– of  $\mathcal{H}(\text{Bl}(\mathbf{x}), \cdot)$ .

$\gamma = \gamma(\epsilon, \delta)$  is a function of  $\epsilon$  and  $\delta$  to be determined later, and where  $LS(\mathbf{x}')$  denotes the local sensitivity at  $\text{Bl}(\mathbf{x}')$ , or equivalently at  $\mathbf{x}'$ , of the scoring function used in our mechanism.

This mechanism also extends to the Dirichlet-multinomial system  $\text{DL}(\boldsymbol{\alpha})$  by rewriting the Hellinger distance as:

$$\mathcal{H}(\text{DL}(\boldsymbol{\alpha}_1), \text{DL}(\boldsymbol{\alpha}_2)) = \sqrt{1 - \frac{\text{B}(\frac{\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2}{2})}{\sqrt{\text{B}(\boldsymbol{\alpha}_1)\text{B}(\boldsymbol{\alpha}_2)}}},$$

and by replacing the  $\mathcal{R}_{\text{post}}$  with set of posterior Dirichlet distributions candidates. Also, the smooth sensitivity  $S(\mathbf{x})$  in (1) will be computed by letting  $\mathbf{x}'$  range over all the elements in  $\mathcal{X}^n$  adjacent to  $\mathbf{x}$ . Notice that  $\mathcal{R}_{\text{post}}$  has  $\binom{n+1}{m-1}$  elements in this case. We will denote by  $\mathcal{M}_{\mathcal{H}}^D$  the mechanism for the Dirichlet-multinomial system.

By setting the  $\gamma$  as  $\ln(1 - \frac{\epsilon}{2 \ln(\frac{\delta'}{2(n+1)})})$ ,  $\mathcal{M}_{\mathcal{H}}$  is  $(\epsilon, \delta)$ -differentially private, where  $\delta = \frac{\epsilon^{\frac{5}{2}} \delta'}{2}$ . (or  $\gamma = \ln(1 - \frac{\epsilon}{2 \ln(\frac{\delta'}{2|\mathcal{R}_{\text{post}}|})})$  when generalized to Dirichlet-multinomial System)



## 5 Privacy Analysis

### 5.1 Privacy Analysis for Baseline Mechanisms

In baseline mechanisms, *exponential mechanism*, *Laplace mechanism*, *improved Laplace mechanism* are  $\epsilon$ -differential privacy provided by [8]. The *exponential mechanism with local sensitivity* is non-differential privacy, also from [8].

### 5.2 Privacy Analysis for $\mathcal{M}_{\mathcal{H}}$

The differential privacy property of  $\mathcal{M}_{\mathcal{H}}$  is proved based on following two lemmas: Lemma 5.1 and Lemma 5.6.

#### Sliding Property of $\mathcal{M}_{\mathcal{H}}$

**Lemma 5.1.** *Given  $\mathcal{M}_{\mathcal{H}}(\mathbf{x})$  calibrated on the smooth sensitivity. Let  $\lambda = f(\epsilon, \delta')$ ,  $\epsilon \geq 0$  and  $|\delta'| < 1$ . Then, the following sliding property holds:*

$$\Pr_{r \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(r, \mathbf{x}) = \hat{s}] \leq e^{\frac{\epsilon}{2}} \Pr_{r \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(r, \mathbf{x}) = (\delta' + \hat{s})] + \frac{\delta'}{2},$$

*Proof.* In what follows, we will use a correspondence between the probability  $\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r]$  of every  $r \in \mathcal{R}_{\text{post}}$  and the probability  $\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[\mathcal{H}(\text{BI}(\mathbf{x}), z) = \mathcal{H}(\text{BI}(\mathbf{x}), r)]$  for the utility score of  $r \in \mathcal{R}_{\text{post}}$ . In Beta-binomial system, because of the property of beta function, the order of parameters in Beta distribution doesn't matter when computing the Hellinegr distance.

**Theorem 5.2.** *for 3 different beta distributions  $\text{beta}(\alpha_0, \beta_0)$ ,  $\text{beta}(\alpha_1, \beta_1)$  and  $\text{beta}(\alpha_2, \beta_2)$ ,  $\mathcal{H}(\text{beta}(\alpha_0, \beta_0), \text{beta}(\alpha_1, \beta_1)) = \mathcal{H}(\text{beta}(\alpha_0, \beta_0), \text{beta}(\alpha_2, \beta_2))$  iff:*

$$\alpha_2 = \beta_1, \alpha_1 = \beta_2, \alpha_0 + \alpha_1 = \beta_0 + \beta_2 \alpha_0 + \alpha_2 = \beta_0 + \beta_1.$$

**Theorem 5.3.** *In Dirichlet-multinomial system, for 3 different Dirichlet distributions  $\text{DL}(\alpha)$ ,  $\text{DL}(\beta_1)$  and  $\text{DL}(\beta_2)$ ,  $\mathcal{H}(\text{DL}(\alpha), \text{DL}(\beta_1)) = \mathcal{H}(\text{DL}(\alpha), \text{DL}(\beta_2))$  iff:*

$$\text{set}(\beta_1) = \text{set}(\beta_2) \text{ and } \text{set}(\alpha + \beta_1) = \text{set}(\alpha + \beta_2),$$

where  $\text{set}(\cdot)$  means we only consider the value of each dimension in the vector without the order.

**Theorem 5.4.** *There are at most two candidates in  $\mathcal{R}_{\text{post}}$  can have the same score, i.e., for every  $r \in \mathcal{R}_{\text{post}}$ ,  $\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[\mathcal{H}(\text{BI}(\mathbf{x}), z) = \mathcal{H}(\text{BI}(\mathbf{x}), r)]$  either*

$$= 2 \left( \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r] \right)$$

or

$$= \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r].$$

*Proof.* of Theorem 5.4.

Let  $|z|$  be the number of candidates  $z \in \mathcal{R}$  that satisfy  $\mathcal{H}(\text{Bl}(\mathbf{x}), z) = \mathcal{H}(\text{Bl}(\mathbf{x}), r)$ , i.e.  $u(z, \mathbf{x}) = u(r, \mathbf{x})$ , then  $\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(z, \mathbf{x}) = u(r, \mathbf{x})] = |z| \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r]$ .

From Theorem 5.2, It can be infer that  $|z| = 2$  or  $|z| = 1$ , i.e.,  $\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(z, x) = u(r, x)] = 2 \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r]$  or  $\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r]$ .  $\square$

**Theorem 5.5.** *In Dirichlet-multinomial system with  $k$ -dimensional Dirichlet distribution, for every  $r \in \mathcal{R}_{\text{post}}$ :*

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[\mathcal{H}(\text{Bl}(\mathbf{x}), z) = \mathcal{H}(\text{Bl}(\mathbf{x}), r)] = |z| \left( \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r] \right),$$

and  $|z|$  can have  $\frac{k!k!}{|a_1|!|a_2|!\dots|a_{k_a}|!|b_1|!|b_2|!\dots|b_{k_b}|!}$  possible values, where  $\sum a_i = \sum b_i = k$ .

*Proof.* of Theorem 5.5.

Let  $\text{Bl}(\mathbf{x}) = \text{DL}(\boldsymbol{\alpha})$ , and  $\text{DL}(\boldsymbol{\beta})$  be a candidate from  $\mathcal{R}_{\text{post}}$ ,  $a_i$  is the distinct value in  $\boldsymbol{\alpha}$  and  $b_i$  the same in  $\boldsymbol{\beta}$ ,  $|a_i|$  denotes the number of values in  $\boldsymbol{\alpha}$  having the value  $a_i$  and  $|b_i|$  the same in  $\boldsymbol{\beta}$ . Then, by combination and permutation, there are  $\frac{k!k!}{|a_1|!|a_2|!\dots|a_{k_a}|!|b_1|!|b_2|!\dots|b_{k_b}|!}$  possible combination that satisfy the condition in Theorem 5.3.  $\square$

In Beta-binomial system, let  $R_1, R_2 \subset \mathcal{R}_{\text{post}}$  be a partition of  $\mathcal{R}_{\text{post}}$ , where every  $z \in R_1$  has a distinct score, i.e.,  $|z| = 1$  and  $z \in R_2$  has another  $z' \in R_2$  with the same score, i.e.,  $|z| = 2$ . The proofs are given by parts:

- for  $r \in R_1$ :

We denote the normalizer of the probability mass in  $\mathcal{M}_{\mathcal{H}}(\mathbf{x})$ :  $\sum_{r' \in \mathcal{R}} \exp(\frac{\epsilon u(r', x)}{2S(\mathbf{x})})$  as  $NL(\mathbf{x})$ :

$$\begin{aligned} LHS &= \Pr_{r \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(r, x) = \hat{s}] = \frac{\exp(\frac{\epsilon \hat{s}}{2S(\mathbf{x})})}{NL(\mathbf{x})} \\ &= \frac{\exp(\frac{\epsilon(\hat{s} + \delta' - \delta')}{2S(\mathbf{x})})}{NL(\mathbf{x})} \\ &= \frac{\exp(\frac{\epsilon(\hat{s} + \delta')}{2S(\mathbf{x})} + \frac{-\epsilon \delta'}{2S(\mathbf{x})})}{NL(\mathbf{x})} \\ &= \frac{\exp(\frac{\epsilon(\hat{s} + \delta')}{2S(\mathbf{x})})}{NL(\mathbf{x})} \cdot e^{\frac{-\epsilon \delta'}{2S(\mathbf{x})}}. \end{aligned}$$

By bounding the  $\delta' \geq -S(\mathbf{x})$ , we can get:

$$\begin{aligned}
\frac{\exp(\frac{\epsilon(\hat{s}+\delta')}{2S(\mathbf{x})})}{NL(\mathbf{x})} \cdot e^{\frac{-\epsilon\delta'}{2S(\mathbf{x})}} &\leq \frac{\exp(\frac{\epsilon(\hat{s}+\delta')}{2S(\mathbf{x})})}{NL(\mathbf{x})} \cdot e^{\frac{\epsilon}{2}} \\
&= e^{\frac{\epsilon}{2}} \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(r, x) = (\delta' + \hat{s})] \leq RHS
\end{aligned}$$

- for  $r \in R_2$ : Proof actually is exactly the same as above by eliminating the parameter  $|z|$  in both sides.

The same in Dirichlet-multinomial distribution, partitioning the  $\mathcal{R}_{\text{post}}$  by the possible value of  $|z|$ , the proof can be derived in the same way as above by part.

□

### Dilation Property of $\mathcal{M}_{\mathcal{H}}$

**Lemma 5.6.** *for any exponential mechanism  $\mathcal{M}_{\mathcal{H}}(\mathbf{x})$ ,  $\lambda < |\beta|$ ,  $\epsilon$ ,  $|\delta'| < 1$  and  $\beta \leq \ln(1 - \frac{\epsilon}{2 \ln(\frac{\epsilon}{\delta'})})$ , the dilation property holds:*

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(z, \mathbf{x}) = c] \leq e^{\frac{\epsilon}{2}} \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(z, \mathbf{x}) = e^{\lambda}c] + \frac{\delta'}{2},$$

where the sensitivity in mechanism is still smooth sensitivity as above.

More specifically, in Beta-binomial system it is enough to take  $\beta \leq \ln(1 - \frac{\epsilon}{2 \ln(\frac{\epsilon}{\delta'(n+1)})})$ .

*Proof.* Based on Theorem 5.2 and 5.4, we use the same partition of  $\mathcal{R}_{\text{post}}$ ,  $R_1, R_2$ . It is enough to proof under just one partition. Without loss of generalization, we take  $R_1$ .

The sensitivity is always greater than 0, and our utility function  $-\mathcal{H}(\text{BI}(\mathbf{x}), z)$  is smaller than zero, i.e.,  $u(z, \mathbf{x}) \leq 0$ , we need to consider two cases where  $\lambda < 0$ , and  $\lambda > 0$ :

We set the  $h(c) = \Pr[u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) = c] = \frac{\exp(\frac{-\epsilon c}{2S(\mathbf{x})})}{NL(\mathbf{x})}$ .

We first consider  $\lambda < 0$ . In this case,  $1 < e^{\lambda}$ , so the ratio  $\frac{h(c)}{h(e^{\lambda}c)} = \frac{\exp(\frac{-\epsilon c}{2S(\mathbf{x})})}{\exp(\frac{-\epsilon(c \cdot e^{\lambda})}{2S(\mathbf{x})})}$  is at most  $e^{\frac{\epsilon}{2}}$ .

Next, we proof the dilation property for  $\lambda > 0$ , The ratio of  $\frac{h(c)}{h(e^{\lambda}c)}$  is  $\exp(\frac{\epsilon}{2} \cdot \frac{u(\mathcal{M}_{\mathcal{H}}(\mathbf{x}))(1-e^{\lambda})}{S(\mathbf{x})})$ . Consider the event  $G = \{\mathcal{M}_{\mathcal{H}}(\mathbf{x}) : u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) \leq \frac{S(\mathbf{x})}{(1-e^{\lambda})}\}$ . Under this event, the log-ratio above is at most  $\frac{\epsilon}{2}$ . The probability of  $G$  under density  $h(c)$  is  $1 - \frac{\delta'}{2}$ . Thus, the probability of a given event  $z$  is at most  $\Pr[c \cap G] + \Pr[\bar{G}] \leq e^{\frac{\epsilon}{2}} \Pr[e^{\lambda}c \cap G] + \frac{\delta'}{2} \leq e^{\frac{\epsilon}{2}} \Pr[e^{\lambda}c] + \frac{\delta'}{2}$ .

**Detail proof:** To show:

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(z, \mathbf{x}) = c] \leq e^{\frac{\epsilon}{2}} \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(z, x) = e^{\lambda}c] + \frac{\delta'}{2},$$

for  $z \in R_1$ .

Let  $Pr[u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) = c] = \frac{\exp(\frac{\epsilon c}{2S(\mathbf{x})})}{NL(\mathbf{x})}$  and  $Pr[u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) = e^\lambda c] = \frac{\exp(\frac{\epsilon e^\lambda c}{2S(\mathbf{x})})}{NL(\mathbf{x})}$  by definition.

After simplification, we need to show:  $u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) \leq \frac{S(\mathbf{x})}{(1-e^\lambda)}$ . Because the sensitivity is always greater than 0, and our utility function  $-\mathcal{H}(\mathbf{B}(\mathbf{x}), z)$  is smaller than zero, i.e.,  $u(z, x) \leq 0$ , we need to consider two cases where  $\lambda < 0$ , and  $\lambda > 0$ :

- $\lambda < 0$

The left hand side will always be smaller than 0 and the right hand side greater than 0. This will always holds, i.e.

$$u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) \leq \frac{S(\mathbf{x})}{(1-e^\lambda)}$$

is always true when  $\lambda < 0$

- $\lambda > 0$

Because  $\hat{s} = u(r)$  where  $r \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})$ , we can substitute  $\hat{s}$  with  $u(\mathcal{M}_{\mathcal{H}}(\mathbf{x}))$ . Then, what we need to proof under the case  $\lambda > 0$  is:

$$u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) \leq \frac{S(\mathbf{x})}{(1-e^\lambda)}. \quad (2)$$

Based on the accuracy property of exponential mechanism:

$$Pr[u(\mathcal{M}_E(x, u, \mathcal{R}_{\text{post}})) \leq c] \leq \frac{|\mathcal{R}| \exp(\frac{\epsilon c}{2GS})}{|\mathcal{R}_{OPT}| \exp(\frac{\epsilon OPT_{u(\mathbf{x})}}{2GS})},$$

we derived the accuracy bound for  $\mathcal{M}_{\mathcal{H}}$ :

$$Pr[u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) \leq c] \leq |\mathcal{R}_{\text{post}}| \exp(\frac{\epsilon c}{2S(\mathbf{x})}).$$

In Beta-binomial system,  $|\mathcal{R}_{\text{post}}| = n + 1$ , apply this bound to eq. 2:

$$\begin{aligned} Pr[u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) \leq \frac{S(\mathbf{x})}{(1-e^\lambda)}] &= (n+1) \exp(\frac{\epsilon S(\mathbf{x})}{(1-e^\lambda)2S(\mathbf{x})}) \\ &= (n+1) \exp(\frac{\epsilon}{2(1-e^\lambda)}). \end{aligned}$$

When we set  $\lambda \leq \ln(1 - \frac{\epsilon}{2 \ln(\frac{\delta'}{2(n+1)})})$ , it is easily to derive that  $Pr[u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) \leq \frac{S(\mathbf{x})}{(1-e^\lambda)}] \leq \frac{\delta'}{2}$ .

In Dirichlet-multinomial system,  $\lambda$  is set as  $\leq \ln(1 - \frac{\epsilon}{2 \ln(\frac{\delta'}{2|\mathcal{R}_{\text{post}}|})})$  since  $|\mathcal{R}_{\text{post}}| \neq n + 1$  any more.

□

**$(\epsilon, \delta)$ -Differential Privacy of  $\mathcal{M}_{\mathcal{H}}$**

**Lemma 5.7.**  $\mathcal{M}_{\mathcal{H}}$  is  $(\epsilon, \delta)$ -differential privacy.

*Proof.* For all  $\mathbf{adj}(\mathbf{x}, \mathbf{x}') \in \mathcal{X}$  and any beta distribution  $r$ , we need to show that:

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r] \leq e^{\epsilon} \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x}')}[z = r] + \delta.$$

Let partition  $R_1, R_2 \subset \mathcal{R}_{\text{post}}$  the same as above. Then we prove by part:

- for  $r \in R_1$ :

Given that  $\Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r] = \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(\mathbf{x}, z) = u(\mathbf{x}, r)]$ , let  $U = u(\mathbf{x}, r)$ ,  $U_1 = u(\mathbf{x}', z) - u(\mathbf{x}, z)$ ,  $U_2 = U + U_1$  and  $U_3 = U_2 \cdot \frac{S(\mathbf{x})}{S(\mathbf{x}')} \cdot \ln(\frac{NL(\mathbf{x})}{NL(\mathbf{x}')}).$  Then,

$$\begin{aligned} \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r] &= \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(x, z) = U] \\ &\leq e^{\epsilon/2} \cdot \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(x, z) = U_2] \\ &\leq e^{\epsilon} \cdot \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(x, z) = U_3] + e^{\epsilon/2} \cdot \frac{\delta'}{2} \\ &= e^{\epsilon} \cdot \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x}')} [u(\mathbf{x}', z) = U] + \frac{e^{\epsilon/2} \delta'}{2} \\ &= e^{\epsilon} \cdot \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r] + \delta, \end{aligned}$$

where  $\delta = \frac{e^{\epsilon/2} \delta'}{2}$ . The first inequality holds by the sliding property, since the  $U_1 \geq -S(\mathbf{x})$ . The second inequality holds by the dilation property, since  $\frac{S(\mathbf{x})}{S(\mathbf{x}')} \cdot \ln(\frac{NL(\mathbf{x})}{NL(\mathbf{x}')} ) \leq 1 - \frac{\epsilon}{2 \ln(\frac{\delta'}{2(n+1)})}$ .

- for  $r \in R_2$ :

Given that  $2 \left( \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r] \right) = \Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(\mathbf{x}, z) = u(\mathbf{x}, r)]$ , let  $U = u(\mathbf{x}, r)$ ,  $U_1 = u(y, z) - u(\mathbf{x}, z)$ ,  $U_2 = U + U_1$  and  $U_3 = U_2 \cdot \frac{S(\mathbf{x})}{S(\mathbf{x}')} \cdot \ln(\frac{NL(\mathbf{x})}{NL(\mathbf{x}')} )$ . Then,

$$\begin{aligned}
Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r] &= \frac{1}{2} \left( Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(x, z) = U] \right) \\
&\leq \frac{1}{2} \left( e^{\epsilon/2} \cdot Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(x, z) = U_2] \right) \\
&\leq \frac{1}{2} \left( e^{\epsilon} \cdot Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[u(x, z) = U_3] + e^{\epsilon/2} \cdot \frac{\delta'}{2} \right) \\
&= \frac{1}{2} \left( e^{\epsilon} \cdot Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x}')} [u(y, z) = U] + \frac{e^{\epsilon/2} \delta'}{2} \right) \\
&= e^{\epsilon} \cdot Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[z = r] + \delta,
\end{aligned}$$

where the  $\delta = \frac{e^{\epsilon/2} \delta'}{4}$ .

The first inequality holds by the sliding property, since the  $U_1 \geq -S(\mathbf{x})$ .

The second inequality holds by the dilation property, since  $\frac{S(\mathbf{x})}{S(\mathbf{x}')} \cdot \ln\left(\frac{NL(\mathbf{x})}{NL(\mathbf{x}')}\right) \leq 1 - \frac{\epsilon}{2 \ln(\frac{\delta'}{2(n+1)})}$ .

Based on two cases above,  $\mathcal{M}_{\mathcal{H}}$  is  $(\epsilon, \delta)$ -differential privacy, where  $\delta$  takes the maximum value from the two cases, i.e.,  $\frac{e^{\frac{\epsilon}{2}} \delta'}{2}$ .

Proof in Dirichlet-multinomial system can be derived in the same way by proving in parts and taking the maximum  $\delta$  value of each part.  $\square$

## 6 Accuracy Analysis

### 6.1 Accuracy Bound for Baseline Mechanisms

#### 6.1.1 Accuracy Bound for Exponential Mechanism

The accuracy bound of exponential mechanism is provided in [8] as:

$$Pr[u(\mathcal{M}_E(x, u, \mathcal{R}_{\text{post}})) \leq c] \leq \frac{|\mathcal{R}| \exp(\frac{\epsilon c}{2GS})}{|\mathcal{R}_{OPT}| \exp(\frac{\epsilon OPT_{u(\mathbf{x})}}{2GS})},$$

where  $|\mathcal{R}|$  is the size of the candidate set,  $OPT$  is the optimal candidates,  $|\mathcal{R}_{OPT}|$  is the number of optimal candidates.

#### 6.1.2 Accuracy Bound for Laplace Mechanism

Accuracy bound for Laplace mechanism is provided by its probability density function:

$$Pr[|Y| \geq t] = e^{-\frac{t}{b}},$$

where  $Y \sim Lap(b)$ ,  $b = \frac{\Delta \mathbf{BI}}{\epsilon}$  in our case.

After post-processing, Laplace noise is discretized. Then the accuracy bound for Laplace mechanism is obtained by:

$$Pr[\lfloor Y \rfloor = t] = Pr[t-1 \leq Y < t] = \frac{1}{2} (e^{-\frac{\epsilon(t-1)}{\Delta \mathbf{BI}}} - e^{-\frac{\epsilon t}{\Delta \mathbf{BI}}}).$$

### 6.1.3 Accuracy Bound for Improved Laplace Mechanism

Accuracy bound for improved Laplace mechanism is obtained from the standard Laplace Mechanism by replacing the sensitivity of  $\Delta \text{BI}$  with 2.

$$Pr[\lfloor Y \rfloor = t] = Pr[t - 1 \leq Y < t] = \frac{1}{2}(e^{-\frac{\epsilon(t-1)}{2}} - e^{-\frac{\epsilon t}{2}}).$$

## 6.2 Accuracy Bound for $\mathcal{M}_{\mathcal{H}}$

We explored three accuracy bounds for our exponential mechanism with smooth sensitivity.

First is the tight bound with very accurate calculation.

$$Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[\mathcal{H}(\text{BI}(\mathbf{x}), z) \geq c] = \sum_{\{z | \mathcal{H}(\text{BI}(\mathbf{x}), z) \geq c\}} \frac{e^{-\frac{\epsilon \mathcal{H}(\text{BI}(\mathbf{x}), z)}{S(\mathbf{x})}}}{NL_x}.$$

In order to be more efficient, we designed the second accuracy bound which is slightly looser than the first one:

$$Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[\mathcal{H}(\text{BI}(\mathbf{x}), z) \geq c] \leq \frac{|R| \exp\left(\frac{-\epsilon c}{S(\mathbf{x})}\right)}{NL_x}.$$

In the second bound, we still need to calculate the normaliser every time. So we want make further improvements on efficiency like follows:

$$Pr_{z \sim \mathcal{M}_{\mathcal{H}}(\mathbf{x})}[\mathcal{H}(\text{BI}(\mathbf{x}), z) \geq c] \leq \frac{|R| \exp\left(\frac{-\epsilon c}{S(\mathbf{x})}\right)}{N(n)},$$

where we replace the  $NL_x$  with a value only related to the size of the data. However, we haven't figured out the formula of this  $N(n)$ .

Moreover, based on the accuracy bound in Sec. 6.1.1, we can derive a loose bound:

$$Pr[u(\mathcal{M}_{\mathcal{H}}(\mathbf{x})) \leq c] \leq |\mathcal{R}_{\text{post}}| \exp\left(\frac{\epsilon c}{2S(\mathbf{x})}\right),$$

which has been used in the dilation property proof.

## 7 Experimental Evaluations

### 7.1 Computation Efficiency Optimization

The formula for computing the local sensitivity presented in Sec. 4.1:  $LS(\mathbf{x}) = \max_{\mathbf{x}' \in \mathcal{X}^n: \text{adj}(\mathbf{x}, \mathbf{x}'), r \in \mathcal{R}} |\mathcal{H}(\text{BI}(\mathbf{x}'), r) - \mathcal{H}(\text{BI}(\mathbf{x}), r)|$ . can be reduced to  $\max_{\{|\mathbf{x}, \mathbf{x}'| \leq 1; \mathbf{x}' \in \mathcal{X}^n\}} \mathcal{H}(\text{BI}(\mathbf{x}), \text{BI}(\mathbf{x}'))$  by applying the distance triangle property.

Specifically, the maximum value over  $r \in R$  always achieves at  $r = \text{BI}(\mathbf{x})$ :

$$\begin{aligned} LS(\mathbf{x}) &= \max_{\{|\mathbf{x}, \mathbf{x}'| \leq 1; \mathbf{x}' \in \mathcal{X}^n\}} \{\mathcal{H}(\text{BI}(\mathbf{x}), \text{BI}(\mathbf{x})) - \mathcal{H}(\text{BI}(\mathbf{x}'), \text{BI}(\mathbf{x}))\} \\ &= \max_{\{|\mathbf{x}, \mathbf{x}'| \leq 1; \mathbf{x}' \in \mathcal{X}^n\}} \{\mathcal{H}(\text{BI}(\mathbf{x}'), \text{BI}(\mathbf{x}))\}. \end{aligned}$$

This equation is validated by an experimental result shown in Fig. 2. We calculate the  $\max_{\{|\mathbf{x}, \mathbf{x}'| \leq 1; \mathbf{x}' \in \mathcal{X}^n\}}$  value for every candidate  $r \in \mathcal{R}_{\text{post}}$ . It is shown that maximum value taken when  $r = \text{BI}(\mathbf{x})$ .

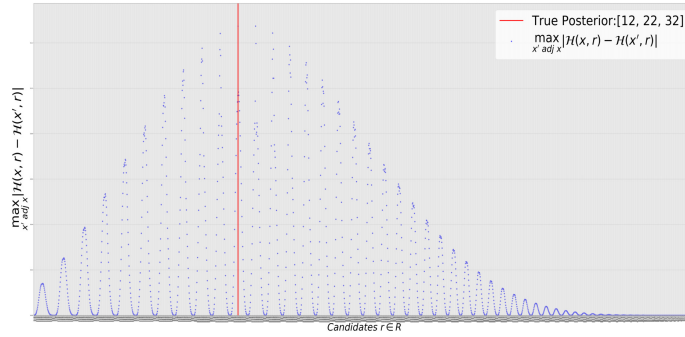


Figure 2: Experimental Results for Finding the Local Sensitivity Efficiently

## 7.2 Theoretical Results

In Fig. 3 we plot on the x-axis the Hellinger distance from the true posterior and on the y-axis the theoretical probabilities of outputting the candidates with that distance under the different mechanisms. We consider *balanced* data sets, which means that in the Beta-Binomial model (Figure 3(a)) the datasets will consist of 50% 1s and the rest 0s, while for the Dirichelet-Multinomial (Figure 3(b)) the data will be split in the  $k = 3$  bins with percentages of: 33%, 33% and 34% in 3 dimensionality. Same concept in 4 dimensionality.

In Fig. 3, candidates of smaller distance from true posterior can be outputted by  $\mathcal{M}_{\mathcal{H}}$  (in blue line) with larger probability than by baseline Laplace mechanism (in green line). This means  $\mathcal{M}_{\mathcal{H}}$  can produce good results with larger probability than baseline mechanism. However, the improved Laplace mechanism represented by red line can produce good results with probability higher than  $\mathcal{M}_{\mathcal{H}}$ . It outperforms  $\mathcal{M}_{\mathcal{H}}$ .

## 7.3 Experimental Results

In this section, we evaluate the accuracy of the mechanisms defined in Section (4) w.r.t. four variables, including data size, dimensions, data variance, prior



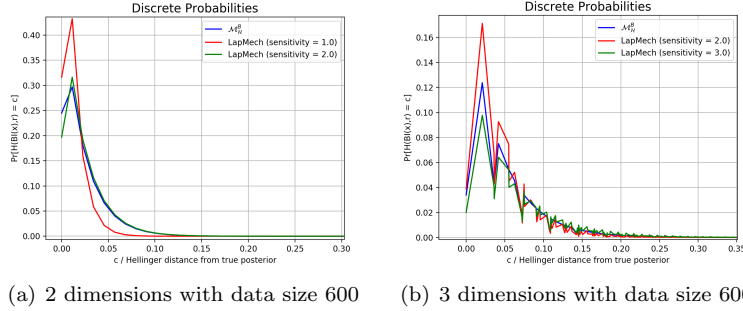


Figure 3: The theory probabilities of candidates in three mechanisms, with balanced data set and parameters  $\epsilon = 1.0$  and  $\delta = 10^{-8}$

distribution, and some combinations thereof. Every plot is an average over 1000 runs. In all the experiments we set  $\epsilon = 1.0$ , and  $\delta = 10^{-8}$ .

In the following some of the plots show mean error as a function of the dataset size while one is a whiskers-plot where the y-axis shows the average accuracy (or equivalently, the error) of the mechanisms, and the x-axis, instead shows different balanced priors used. The boxes extend from the lower to the upper quartile values of the data, with a line at the median. A notch on the box around the median is also drawn to give a rough guide to the significance of difference of medians; The whiskers extend from the box to show the range of the data. A blue box in the plots represents our newly designed exponential mechanism's behavior—where the sensitivity is calibrated w.r.t Hellinger distance—while the yellow box next to it represents the performance of a variation of the basic Laplace mechanism presented in Section (4.1) with the same settings: that is  $\epsilon, \delta$ , data, prior. The variation considered performs a postprocessing on the released parameters so that they are consistent. For instance when the sum of the noised parameters is greater than  $n$  we will truncate them so that they sum up to  $n$ .

**Increasing data size with balanced datasets** In Figures 4, 5(a) and 5(b) we still consider *balanced* data sets of observations. The results show that when the data size increases, the average errors of  $\mathcal{M}_{\mathcal{H}}$ , Laplace mechanism and decrease. For small datasets, i.e with size less 300 in the case of Beta-Binomial models, both the baseline Laplace mechanisms and improved Laplace mechanism outperform  $\mathcal{M}_{\mathcal{H}}$ . But for bigger data sets, that is, bigger than 300, or as in Figure 4 where we considered data sets of the order of 15 thousands elements, the  $\mathcal{M}_{\mathcal{H}}$  outperforms the baseline Laplace mechanism, and asymptotically approaches the improved Laplace mechanism. Similar experimental tendencies were obtained for the Dirichlet-multinomial model (Figure 5(a) and 5(b)).

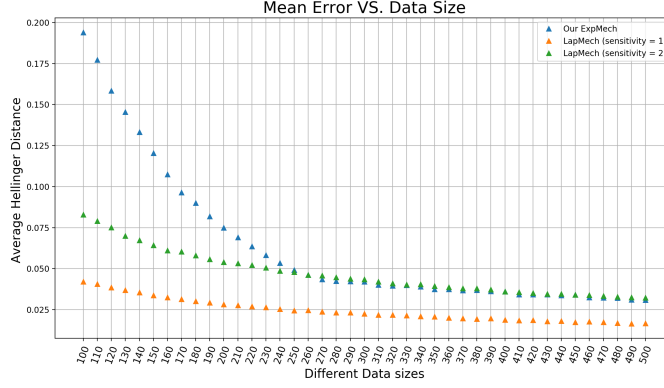
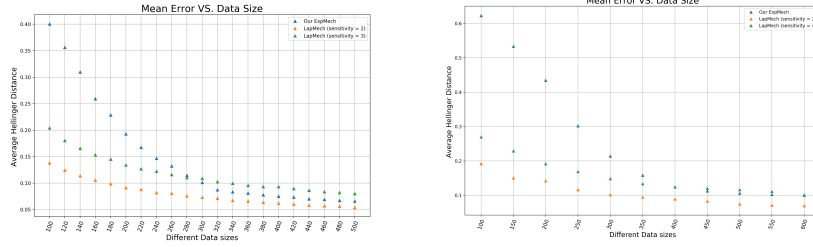


Figure 4: Increasing data size with prior  $\text{beta}(1,1)$ , balanced datasets and parameters  $\epsilon = 1.0$  and  $\delta = 10^{-8}$



(a) Increasing data size with  $\text{DL}(1,1,1)$  prior distribution, balanced datasets and parameters  $\epsilon = 1.0$  and  $\delta = 10^{-8}$  (b) Increasing data size with  $\text{DL}(1,1,1)$  prior distribution, Unbalanced datasets and parameters  $\epsilon = 1.0$  and  $\delta = 10^{-8}$

**Fixed dataset varying balanced priors** In Figure 5, we fix the data set to be  $(50,50)$ , and the parameters the same as before:  $\epsilon = 1.0$  and  $\delta = 10^{-8}$ . We studied the accuracy under different priors, where the priors considered are also balanced. Similar to the plots above, Figure 5 shows that in the beginning the baseline Laplace mechanism and improved Laplace mechanism performs better but the baseline approach is outperformed after a while, and very close to the improved Laplace mechanism.

## 7.4 Experiment Evaluations on Privacy Loss

In order to see our privacy behavior, we study the accurate epsilon under concrete cases in this section. The  $(\epsilon, \delta)$  - differential privacy we proved in Sec. 4.2 is just an upper bound, we concrete  $\epsilon$  should be smaller than upper bound in our exponential mechanism. We calculate the concrete privacy value in following

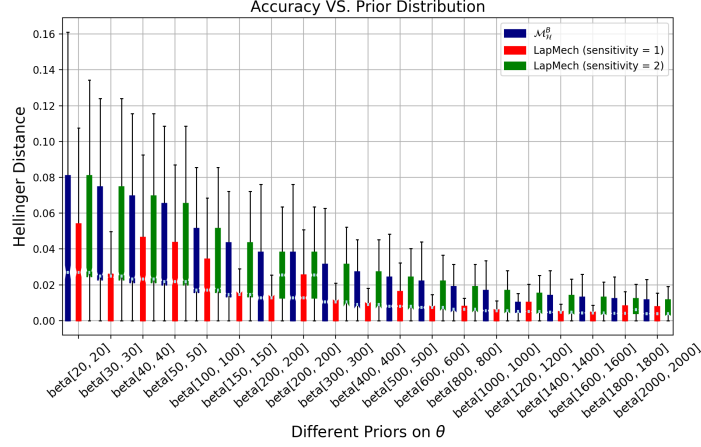


Figure 5: Observed data set is:  $(50, 50)$ , varying balanced priors

ways wrt. the data size, and obtain plots in Fig. 6.

$\epsilon = 1.0$  is a privacy upper bound, we can observe that the concrete  $\epsilon$  values are smaller than the upper bound. That is to say, we achieved a higher privacy level than expected. In next step, we are going to improve the accuracy using this property.

## 8 Conclusion and Future Work

From what we have seen in the previous sections we can obtain some preliminary conclusions. That is, the probability measure approach outperforms the  $\ell_1$ -norm approach in the following cases:

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

These results although very motivating, are still not enough for real world applications. Hence, we will continue our work in the following directions:

1. For now, we just have a intuitive idea on the accuracy behavior of our mechanisms, and not a precise formula or bound on it. When do our mechanisms perform better than the baseline mechanism and when they don't? How much influence will elements in Section 7 have on the accuracy? Are there any other important factors we missed? These are all questions w.r.t. the accuracy that we are going to explore next, and in a more principled and formal way.

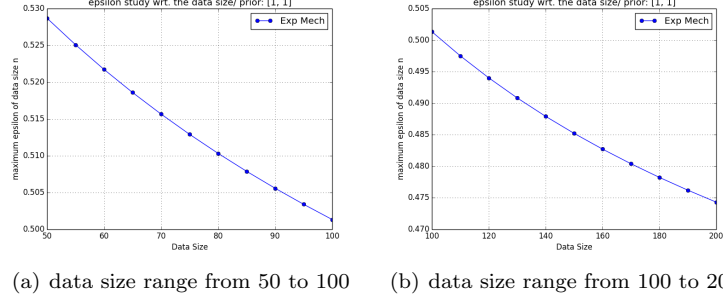


Figure 6: Concrete privacy calculation under settings that: prior distribution:  $[1, 1]$ ,  $\epsilon = 1.0$ ,  $\delta = 0.0005$  and observed data are uniformly distributed

2. Theorem 5.7 provides an upper bound on the privacy loss for  $\mathcal{M}_{\mathcal{H}}$  and  $\mathcal{M}_{\mathcal{H}}^D$  but not necessarily a tight one. Indeed, experiments have shown that the actual privacy loss in the experiments can be smaller than  $\epsilon$ . This means that we could improve accuracy, by adding less noise – that is noise proportional to a higher value of  $\epsilon$  – but still achieve  $(\epsilon, \delta)$ -dp.
3. The choice of the Hellinger distance might seem quite ad-hoc. Hence, it is worth exploring other distances over distributions. An interesting class of probability metrics is the family of  $f$ -divergences [5].

## References

- [1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *CCS 2016*, pages 308–318.
- [2] Kamalika Chaudhuri and Daniel Hsu. Convergence rates for differentially private statistical estimation. In *ICML, 2012*, page 1327.
- [3] Kamalika Chaudhuri and Claire Monteleoni. Privacy-preserving logistic regression. In *NIPS, 2009*, pages 289–296.
- [4] Kamalika Chaudhuri, Anand Sarwate, and Kaushik Sinha. Near-optimal differentially private principal components. In *NIPS, 2012*, pages 989–997.
- [5] I. Csiszár and P.C. Shields. Information theory and statistics: A tutorial. *Foundations and Trends in Communications and Information Theory*, 1(4):417–528, 2004.

- [6] Christos Dimitrakakis, Blaine Nelson, Aikaterini Mitrokotsa, and Benjamin IP Rubinstein. Robust and private bayesian inference. In *ALT, 2014*, pages 291–305.
- [7] Christos Dimitrakakis, Blaine Nelson, Zuhe Zhang, Aikaterini Mitrokotsa, and Benjamin IP Rubinstein. Differential privacy in a bayesian setting through posterior sampling.
- [8] Cynthia Dwork, Aaron Roth, et al. *The algorithmic foundations of differential privacy*. Now Publishers, Inc., 2014.
- [9] James Foulds, Joseph Geumlek, Max Welling, and Kamalika Chaudhuri. On the theory and practice of privacy-preserving bayesian data analysis.
- [10] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *FOCS, 2007*.
- [11] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *STOC, 2007*, pages 75–84.
- [12] Yining Wang, Yu-Xiang Wang, and Aarti Singh. Differentially private subspace clustering. In *NIPS, 2015*, pages 1000–1008.
- [13] Yu-Xiang Wang, Stephen Fienberg, and Alex Smola. Privacy for free: Posterior sampling and stochastic gradient monte carlo. In *ICML, 2015*, pages 2493–2502.
- [14] Oliver Williams and Frank McSherry. Probabilistic inference and differential privacy. In *NIPS, 2010*, pages 2451–2459.
- [15] Yonghui Xiao and Li Xiong. Bayesian inference under differential privacy.
- [16] Jun Zhang, Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava, and Xiaokui Xiao. Privbayes: Private data release via bayesian networks. pages 1423–1434.
- [17] Zuhe Zhang, Benjamin IP Rubinstein, Christos Dimitrakakis, et al. On the differential privacy of bayesian inference. In *AAAI, 2016*, pages 2365–2371.
- [18] Shijie Zheng. The differential privacy of bayesian inference. In *Bachelor’s thesis, Harvard College, 2015*.