Tailoring Differentially Private Bayesian Inference to Distance Between Distributions

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, Dimitrakakis et al. [1], and Wang et al. [6] proved that, under specific conditions, sampling from the posterior distribution is already differentially private. Zhang et al. [8], Foulds et al. [3], 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, χ^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 an approach based on the Laplace mechanism.

KEYWORDS

Differential privacy, Bayesian inference, Hellinger distance

1 INTRODUCTION

2 MOTIVATION

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 ϵ 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 ([7, 8]), it is only measured with respect to the vector of numbers parametrizing the output distribution using, e.g. the ℓ_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. [8]) 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.

3 BAYESIAN INFERENCE BACKGROUND

Given a prior belief $\Pr(\theta)$ on some parameter θ, and an observation $\mathbf{x}$, the posterior distribution on θ 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 θ. Since we consider $\mathbf{x}$ to be fixed, the likelihood is a function of θ. For the same reason $\Pr(\mathbf{x})$ is a constant independent of θ. Usually in statistics the prior distribution $\Pr(\theta)$ is chosen so that it represents the initial belief on θ, 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. In this work we will consider a specific instance of Bayesian inference and one of its generalizations. Specifically, we will consider the situation where θ represents the parameter –informally called *bias*– of a Bernoulli distributed random variable, and its immediate generalization where the parameter $\boldsymbol{\theta}$ represents the vector of parameters of a categorically distributed random variable. In the former case, the prior distribution over $\theta \in [0, 1]$ is going to be a beta distribution, $\mathrm{beta}(\alpha, \beta)$, with parameters $\alpha, \beta \in \mathbb{R}^+$, and with p.d.f:

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

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

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

In the latter case the prior distribution over $\boldsymbol{\theta} \in [0, 1]^k$ is given by a Dirichelet distribution, $\mathrm{DL}(\boldsymbol{\alpha})$, for $k \in \mathbb{N}$, and $\boldsymbol{\alpha} \in (\mathbb{R}^+)^k$, with p.d.f:

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

where $\mathrm{B}(\cdot)$ 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:

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

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

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

where + denotes the componentwise sum of vectors of reals.

4 THE PROBLEM STATEMENT

We are interested in designing a mechanism for privately releasing the full posterior distributions derived in section 3, 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 $\boldsymbol{\alpha}' = (\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha})$. Zhang et al. [8] and Xiao and Xiong [7] 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 $\widetilde{\Delta\alpha} \sim Lap(\Delta\alpha, \frac{2}{\epsilon})$, and where $Lap(\mu, \nu)$ denotes a Laplace random variable with mean μ and scale ν. This mechanism is ϵ-differentially private, and the noise is calibrated w.r.t. to a sensitivity of 2 which is derived by using ℓ_1 norm over the pair of parameters. Indeed, considering two adjacent[1] data observations $\mathbf{x}, \mathbf{x}'$, that, from a unique prior, give rise to two posterior distributions, characterized by the pairs (α', β') and (α'', β'') then $|\alpha' - \alpha''| + |\beta' - \beta''| \leq 2$. This argument extends similarly to the Dirichelet-Multinomial system.
Also, in previous works, the accuracy of the posterior was measured again with respect to ℓ_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. In this work we will use a metric based on a different norm to compute the sensitivity and provide guarantees on the accuracy. In particular we will consider a metric over probability measures and not over the parameters that represent them. Specifically, we will use the Hellinger distance $\mathcal{H}(\cdot, \cdot)$. Our choice to use Hellinger distance is motivated by two facts, first of all it simplifies calculations in the case of the probabilistic models considered here and second of all it also automatically yields bounds on the total variation distance, which represents also the maximum advantage an unbounded adversary can have in distingishing two distributions. Given two beta distributions $\boldsymbol{\beta}_1 = \mathrm{beta}(\alpha_1, \beta_1)$, and $\boldsymbol{\beta}_2 = \mathrm{beta}(\alpha_2, \beta_2)$ the following equality holds

$$\mathcal{H}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sqrt{1 - \frac{\mathrm{B}(\frac{\alpha_1+\alpha_2}{2}, \frac{\beta_1+\beta_2}{2})}{\sqrt{\mathrm{B}(\alpha_1, \beta_1)\mathrm{B}(\alpha_2, \beta_2)}}}$$

The same change of metric will be applied to the experimental accuracy guarantees.

[1] Given $\mathbf{x}, \mathbf{x}'$ we say that $\mathbf{x}$ and $\mathbf{x}'$ are adjacent and we write, $\mathbf{adj}(\mathbf{x}, \mathbf{x}')$, iff $\sum_{i}^{n} [x_i = x_i'] \leq 1$.

5 MECHANISMS INTRODUCTION

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

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

where $\Delta\alpha$ is as defined in Section 3. 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 $\mathsf{BI}(\mathbf{x})$ – we don't explicitely parametrize the result by the prior, which from now on we consider fixed and we denote it by $\boldsymbol{\beta}_{\text{prior}}$.

5.1 Baseline Approaches

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

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

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

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

Exponential mechanism is $\epsilon-$differential privacy[2].

5.1.2 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 global sensitivity. This outputs a candidate $r \in \mathcal{R}$ with probability proportional to $exp(\frac{\epsilon u(x,r)}{2LS(x)})$:

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

where local sensitivity is calculated by:

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

The exponential mechanism with local sensitivity is non-differential privacy[2].

5.1.3 Laplace Mechanism.

- Release $\mathrm{beta}(\alpha + \lfloor\widetilde{\Delta\alpha}\rfloor_0^n, \beta + n - \lfloor\widetilde{\Delta\alpha}\rfloor_0^n)$.
- $\widetilde{\Delta\alpha} \sim \mathcal{L}(\Delta\alpha, \frac{\Delta\mathsf{BI}}{\epsilon})$.
- $\Delta\mathsf{BI} \equiv \max_{\mathbf{x},\mathbf{x}' \in \{0,1\}^n, ||\mathbf{x}-\mathbf{x}'||_1 \leq 1} ||\mathsf{BI}(\mathbf{x}) - \mathsf{BI}(\mathbf{x}')||_1$.

5.1.4 Improved Baseline Approach. y 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 $||\mathsf{BI}(\mathbf{x}) - \mathsf{BI}(\mathbf{x'})||_1 \leq 2$.

5.2 Our approach: smoothed Hellinger distance based exponential mechanism

5.2.1 Setting up.

Definition 5.1. The mechanism $\mathcal{M}_{\mathcal{H}}(x)$ outputs a candidate $r \in \mathcal{R}_{\text{post}}$ with probability

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

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

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

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

This mechanism is based on the basic exponential mechanism [4], 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 Nissim et al. [5]– of $\mathcal{H}(\mathsf{BI}(\mathbf{x}),\cdot)$.

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

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

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

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 $\mathcal{M}^D_{\mathcal{H}}$ the mechanism for the Dirichlet-Multinomial system.

By setting the γ as $\ln(1-\frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})})$, $\mathcal{M}_{\mathcal{H}}$ is (ϵ,δ)–differentially private.

6 PRIVACY ANALYSIS

The differential privacy property of $\mathcal{M}_{\mathcal{H}}$ is proved based on the holds of the two properties: *sliding property* and *dilation property.*

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

LEMMA 6.1. *Given $\mathcal{M}_{\mathcal{H}}(x)$ calibrated on the smooth sensitivity. Let $\lambda = f(\epsilon,\delta)$, $\epsilon \geq 0$ and $|\delta| < 1$. Then, the following* sliding property *holds:*

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

PROOF. In what follows, we will use a correspondence between the probability $\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z=r]$ of every $r\in\mathcal{R}_{\text{post}}$ and the probability $\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[\mathcal{H}(\mathsf{BI}(x),z)=\mathcal{H}(\mathsf{BI}(x),r)]$ for the utility score for r. In particular, for every $r\in\mathcal{R}_{\text{post}}$ we have:

$$\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z=r] = \frac{1}{2}\left(\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[\mathcal{H}(\mathsf{BI}(x),z)=\mathcal{H}(\mathsf{BI}(x),r)]\right)$$

To see this, it is enough to notice that: $\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z=r]$ is proportional too $\mathcal{H}(\mathsf{BI}(x),r)$, i.e., $u(x,z)$. We can derive, if $u(r,x)=u(r',x)$ then $\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z=r] = \Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z=r']$. We assume the number of candidates $z\in\mathcal{R}$ that satisfy $u(z,x)=u(r,x)$ is $|r|$, we have $\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[u(z,x)=u(r,x)] = |r|\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z=r]$. Because Hellinger distance $\mathcal{H}(\mathsf{BI}(x),z)$ is axial symmetry, where the $\mathsf{BI}(x)$ is the symmetry axis. It can be infer that $|z|=2$ for any candidates, apart from the true output, i.e., $\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[u(z,x)=u(r,x)] = 2\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z=r]$. This parameter can be eliminate in both sides in proof.

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

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

By bounding the $\Delta \geq -S(x)$, we can get:

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

□

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

LEMMA 6.2. *for any exponential mechanism $\mathcal{M}_{\mathcal{H}}(x)$, $\lambda < |\beta|$, ϵ, $|\delta| < 1$ and $\beta \leq \ln(1-\frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})})$, the dilation property holds:*

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

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

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

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

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

Next, we proof the dilation property for $\lambda > 0$, The ratio of $\frac{h(c)}{h(e^{\lambda}c)}$ is $\exp(\frac{\epsilon}{2}\cdot\frac{u(\mathcal{M}_{\mathcal{H}}(x))(1-e^{\lambda})}{S(x)})$. Consider the event $G = \{\mathcal{M}_{\mathcal{H}}(x) : u(\mathcal{M}_{\mathcal{H}}(x)) \le \frac{S(x)}{(1-e^{\lambda})}\}$. 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 $Pr[c\cap G] + Pr[\overline{G}] \le e^{\frac{\epsilon}{2}}Pr[e^{\lambda}c\cap G] + \frac{\delta}{2} \le e^{\frac{\epsilon}{2}}Pr[e^{\lambda}c] + \frac{\delta}{2}$.

Detail proof:

By simplification, we get this formula: $u(\mathcal{M}_{\mathcal{H}}(x)) \le \frac{S(x)}{(1-e^{\lambda})}$

- $\lambda < 0$

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

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

  is always true when $\lambda < 0$
- $\lambda > 0$

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

  $$u(\mathcal{M}_{\mathcal{H}}(x)) \le \frac{S(x)}{(1-e^{\lambda})} \tag{2}$$

  Based on the accuracy property of exponential mechanism:

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

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

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

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

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

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

□

6.3 $(\epsilon,\delta)-$Differential Privacy

Lemma 6.3. *$\mathcal{M}_{\mathcal{H}}$ is (ϵ,δ)-differential privacy.*

Proof. of Lemma 6.3: For all neighboring $x, y \in D^n$ and all sets $\mathcal{S}$, we need to show that:

$$\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z\in\mathcal{S}] \le e^{\epsilon}\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(y)}[z\in\mathcal{S}] + \delta.$$

Given that $2\left(\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z\in\mathcal{S}]\right) = \Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[u(x,z)\in\mathcal{U}]$, let $\mathcal{U}_1 = \frac{u(y,z)-u(x,z)}{S(x)}$, $\mathcal{U}_2 = \mathcal{U}+\mathcal{U}_1$ and $\mathcal{U}_3 = \mathcal{U}_2\cdot\frac{S(x)}{S(y)}\cdot\ln(\frac{NL(x)}{NL(y)})$. Then,

$$\begin{aligned} 2\left(\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z\in\mathcal{S}]\right) &= \Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[u(x,z)\in\mathcal{U}] \\ &\le e^{\epsilon/2}\cdot\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[u(x,z)\in\mathcal{U}_2] \\ &\le e^{\epsilon}\cdot\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[u(x,z)\in\mathcal{U}_3] + e^{\epsilon/2}\cdot\frac{\delta'}{2} \\ &= e^{\epsilon}\cdot\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(y)}[u(y,z)\in\mathcal{U}] + \delta = 2\left(e^{\epsilon}\cdot\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z\in\mathcal{S}]\right) \end{aligned}$$

The first inequality holds by the sliding property, since the $\mathcal{U}_1 \ge -S(x)$. The second inequality holds by the dilation property, since $\frac{S(x)}{S(y)}\cdot\ln(\frac{NL(x)}{NL(y)}) \le 1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})}$.

□

7 ACCURACY ANALYSIS

7.1 Accuracy Bound for Laplace Mechanism

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

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

where $Y \sim Lap(b)$, $b = \frac{GS}{\epsilon}$

After post-processing: $Pr[\lfloor Y\rfloor = t] = Pr[t-1 \le Y < t] = \frac{1}{2}(e^{-\frac{t-1}{b}} - e^{-\frac{t}{b}})$.

7.2 Accuracy Bound for Exponential Mechanism

The accuracy bound of exponential mechanism is provided in [2] as:

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

where $|R|$ is the size of the candidate set, OPT is the optimal candidates, $|R_{OPT}|$ is the number of optimal results.

7.3 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.

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

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

$$\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[\mathcal{H}(\mathsf{BI}(x),z) \ge c] \le \frac{|R|\exp(\frac{-\epsilon c}{S(x)})}{NL_x}.$$

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

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

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. 7.2, we can derive a loose bound:

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

which has been used in the dilation property proof.

8 EXPERIMENTAL EVALUATIONS

8.1 Computation Efficiency

The formula for computing the local sensitivity presented in Sec. 5.1.2: $\max_{\{|x,y'| \leq 1; y' \in \mathcal{X}^n\}} \max_{\{r \in R\}} \{\mathcal{H}(\mathsf{BI}(x), r) - \mathcal{H}(\mathsf{BI}(y'), r)|\}$ can be reduced to $\max_{\{|x,y'| \leq 1; y' \in \mathcal{X}^n\}} \mathcal{H}(\mathsf{BI}(x), \mathsf{BI}(y'))|$ by applying the distance triangle property. i.e., the maximum value over $\max_{r \in R}$ always achieves at $r = \mathsf{BI}(x)$ itself, where $\Delta_l u(x) = \max_{\{|x,y'| \leq 1; y' \in \mathcal{X}^n\}} \{\mathcal{H}(\mathsf{BI}(x), \mathsf{BI}(x)) - \mathcal{H}(\mathsf{BI}(y'), \mathsf{BI}(x))|\} = \max_{\{|x,y'| \leq 1; y' \in \mathcal{X}^n\}} \{\mathcal{H}(\mathsf{BI}(y'), \mathsf{BI}(x))|\}$. We also have some experiments for validating our proposal as in Fig. **??**, where we calculate the $\max_{\{|x,y'| \leq 1; y' \in \mathcal{X}^n\}}$ value for every candidate $r \in R$. It is shown that maximum value taken when $r = \mathsf{BI}(x)$.

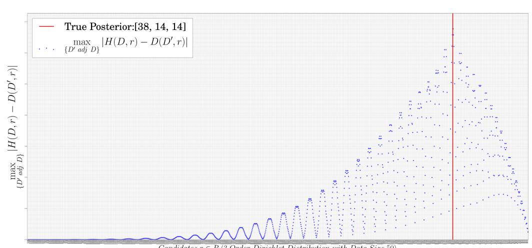
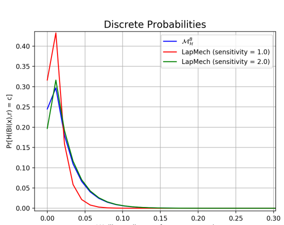
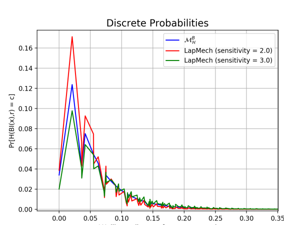


Figure 1: Experimental Results for Finding the Local Sensitivity Efficiently

8.2 Theoretical Results



(a) 2 dimensions with data size 600



(b) 3 dimensions with data size 600

8.3 Experimental Results

In this section, we evaluate the accuracy of the mechanisms defined in Section (5) w.r.t. four variables, including data size, dimensions, data variance, prior 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 datasize 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 (5.1) with the same settings: that is ϵ, δ, 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.

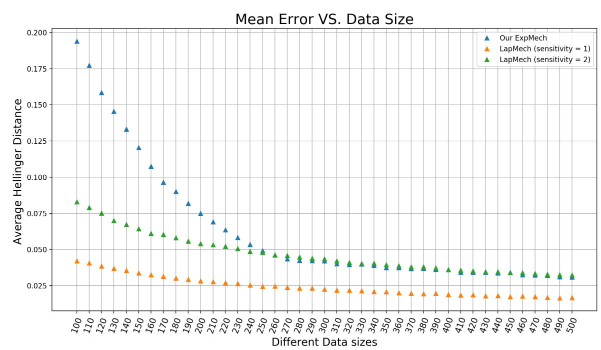


Figure 2: Increasing data size with prior $\text{beta}(1,1)$**, balanced datasets and parameters** $\epsilon = 1.0$ **and** $\delta = 10^{-8}$

Increasing data size with balanced datasets. In Figures 2, 3 and 4 we consider *balanced* data sets of observations. This means that in the Beta-Binomial setting the datasets will consist of 50% 1s and the rest 0s, while for the Dirichelet-Multinomial (Figure 3 and 4) the data will be split in the $k = 3$ bins with perecentages of: 33%, 33% and 34% in 3 dimensionality and 25%, 25%, 25% and 25% in 4 dimensionality. 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 2 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 baseline

Figure 3: Increasing data size with $\mathrm{DL}(1,1,1)$ **prior distribution, balanced datasets and parameters** $\epsilon = 1.0$ **and** $\delta = 10^{-8}$

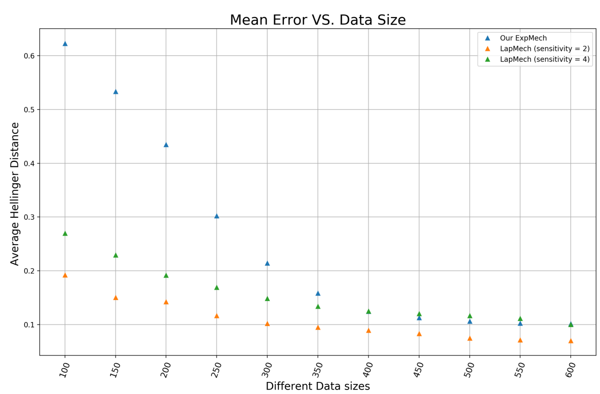


Figure 4: Increasing data size with $\mathrm{DL}(1,1,1,1)$ **prior distribution, Unbalanced datasets and parameters** $\epsilon = 1.0$ **and** $\delta = 10^{-8}$

Laplace mechanism. Similar experimental tendencies were obtained for the Dirichlet-Multinomial system (Figure 3 and 4).

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. The plot shows that in the beginning the Laplace mechanism performs better but it is outperformed after a while.

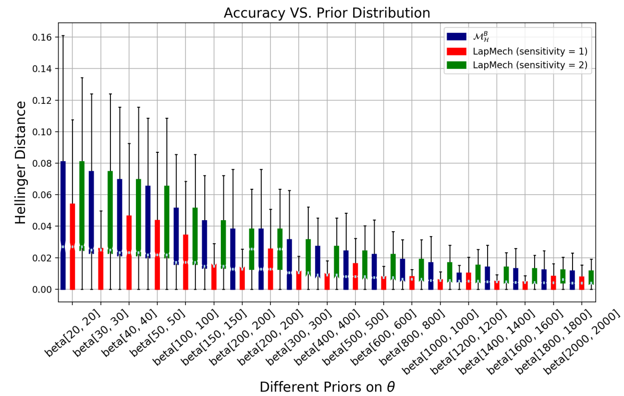


Figure 5: Observed data set is: $(50, 50)$**, varying balanced priors**

8.4 Experiment Evaluations on Privacy Loss

In order to see our privacy behavior, we study the accurate epsilon under concrete cases in this section. The (ϵ, δ) - differential privacy we proved in Sec. 5.2 is just an upper bound, we concrete ϵ should be smaller than upper bound in our exponential mechanism. We calculate the concrete privacy value in following ways wrt. the data size, and obtain plots in Fig. 6.

$\epsilon = 0.8$ is a privacy upper bound, we can observe that the concrete ϵ 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.

REFERENCES

[1] Christos Dimitrakakis, Blaine Nelson, Aikaterini Mitrokotsa, and Benjamin IP Rubinstein. 2014. Robust and private Bayesian inference. In *International Conference on Algorithmic Learning Theory*. Springer, 291–305.

[2] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.

[3] James Foulds, Joseph Geumlek, Max Welling, and Kamalika Chaudhuri. 2016. On the theory and practice of privacy-preserving Bayesian data analysis. *arXiv preprint arXiv:1603.07294* (2016).

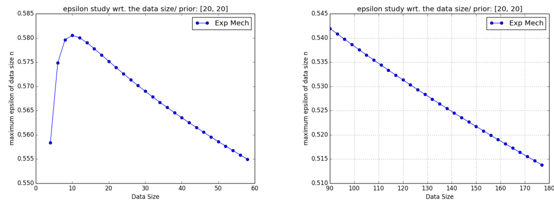
[4] Frank McSherry and Kunal Talwar. 2007. Mechanism Design via Differential Privacy, In Annual IEEE Symposium on Foundations of Computer Science (FOCS). https://www.microsoft.com/en-us/research/publication/mechanism-design-via-differential-privacy/

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

[6] Yu-Xiang Wang, Stephen Fienberg, and Alex Smola. 2015. Privacy for free: Posterior sampling and stochastic gradient monte carlo. In *International Conference on Machine Learning*. 2493–2502.

[7] Yonghui Xiao and Li Xiong. 2012. Bayesian inference under differential privacy. *arXiv preprint arXiv:1203.0617* (2012).

[8] Zuhe Zhang, Benjamin IP Rubinstein, Christos Dimitrakakis, et al. 2016. On the Differential Privacy of Bayesian Inference.. In *AAAI*. 2365–2371.



(a) data size range from 90 to 180 (b) data size range from 90 to 180

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