Exploring the Design Space of Differentially Private Bayesian Inference under Hellinger Distance

Anonymous Authors[1]

Abstract

We explore the design space of differentially private Bayesian inference mechanisms with accuracy measured in terms of the Hellinger distance between posterior distributions. We focus on two discrete models for parametric Bayesian inference: the Beta-Binomial and the Dirichlet-Multinomial models. We study two mechanisms based on Laplace perturbation of the parameters of the posterior distribution under the ℓ_1 norm, and compare them with a discrete mechanism calibrating noise to a smooth upper bound on the Hellinger distance. Accuracy is measured through the Hellinger distance between the posterior distribution released by a given mechanism and the one obtained via non-private inference. We compare the accuracy of our mechanisms theoretically and experimentally.

1. Introduction

Modern data analysis techniques have enabled significant advances in a variety of applications in medicine, finance, social science, and transportation. In order to provide better services, these applications need large amounts of users' data, putting at risk the privacy of individual users contributing their data. Differential privacy was proposed a decade ago to address privacy concerns in these situations, and is now the standard for privacy-preserving data analysis.

Many statistical applications are driven by Bayesian inference. This is a standard statistical tool in which a prior distribution is combined with a sample of observed data in order to estimate a new posterior distribution.

In this work we consider the design space of parametric Bayesian inference mechanisms guaranteeing differential privacy. Our work is motivated by recent developments in the area of probabilistic programming where several tools have been proposed to perform parametric and non-parametric Bayesian inference in an efficient way.

We focus on inference tasks based on two classical statistical models: the Beta-Binomial and the Dirichlet-Multinomial models. In the Beta-Binomial model, we are given a dataset consisting of independent Bernoulli trials with bias p. Given an initial prior belief that p follows a Beta distribution, we may use the dataset to infer a posterior belief that p follows a Beta distribution with updated parameters. This model generalizes to the Dirichlet-Multinomial model, in which a multinomial distribution is used to describe categorial data with more than two possible outcomes. Similarly, the goal is infer a belief about the proportions of each outcome. Given a prior belief that these proportions come from a Dirichlet distribution, the posterior belief given multinomial data will also follow a Dirichlet distribution.

In our work, we consider mechanisms which output a complete description of the posterior distribution inferred using private data. While the prior distribution associated to a given inference task is considered public information, the inferred posterior distribution may leak information about the individuals in the dataset. In order to guarantee differential privacy, we introduce random noise to the posterior before releasing it. Our goal is to design mechanisms which are accurate, in the sense that this noisy posterior is close to the posterior one would have inferred without concern for privacy. The most natural way to measure closeness is via a distance between probability distributions – nevertheless, only a few works in the literature (e.g., (Zhang et al., 2016)) actually state accuracy guarantees in this way.

The amount of the noise that we need to introduce depends on the differential privacy parameter ϵ and the sensitivity of the inference to small changes in the input dataset. As with accuracy, sensitivity may be measured in many different ways, with each choice giving rise to a different method of introducing noise for differential privacy. In the existing literature on private Bayesian inference (e.g. (Zhang et al., 2016; Xiao & Xiong, 2012)), sensitivity is measured via a norm (e.g., the ℓ_1 norm) on the the vector of numbers parametrizing the output distribution. But when the accuracy of a mechanism is determined by a distance on dis-

[1] Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. **AUTHORERR: Missing \icmlcorrespondingauthor.**

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

tributions themselves, this may not be such a natural choice. The distance between two parameterized distributions may behave wildly as a function of the distance between the vectors of parameters characterizing them.

A more natural approach, which we propose here, is to also directly measure sensitivity with respect to a metric on the space of inferred probability distributions. Our goal is to identify cases where jointly measuring accuracy and sensitivity in terms of a single metric on probability distributions can yield improved accuracy.

In summary, the main contributions of our work are:

- We give a new differentially private Bayesian inference mechanism based on the standard exponential mechanism.
- We propose two ways of improving the accuracy of this basic mechanism: 1) We calibrate noise to a measure of sensitivity determined by a metric between distributions (e.g. Hellinger distance ($\mathcal{H}$), f-divergences, etc...). 2) We scale noise according to a a smooth upper bound on the local sensitivity, rather than according to global sensitivity.
- We implement the new proposed mechanism and other art-of-state mechanisms, comparing their performances in terms of accuracy and privacy.

2. Related Work

Several works have considered the intersection of Bayesian Inference and differential privacy. Williams & McSherry (2010) explored the use of probabilistic inference to improve the accuracy of differentially private data analyses. A similar approach has been also studied by Xiao & Xiong (2012).

A general trend has been to study robustness conditions for different likelihood functions and different classes of conjugate priors guaranteeing differential privacy for samples taken from a posterior distribution. Dimitrakakis et al. (2014) consider this approach and identify some general robustness technique which can be applied to several models including the one considered in this paper. Similarly, Zheng (2015) considers a similar approach but with different conditions. Zhang et al. (2016) extend the previous approaches to address more general models with multiple random variables. Wang et al. (2015) identifies conditions guaranteeing differential privacy for samples generated using standard montecarlo algorithm. Geumlek et al. (2017) consider the problem of releasing samples from distributions in the exponential family under the constraints imposed by Renyi differential privacy. Zhang et al. (2014) propose a technique for releasing high dimensional data based on the differentially private learning of a Bayesian graphical model.

Few works take an approach similar to the one we consider here. Dimitrakakis et al. (2015) study differentially private posterior release by output perturbation based on Laplace noise for the Beta-Binomial model. However, they measure accuracy in terms of ℓ_1 distance over the parameters and in term of KL-divergence over the distribution. Foulds et al. (2016) generalizes the approach of Dimitrakakis et al. (2015) by analyzing further the impact that Laplace noise has on the inference. They also focus on accuracy measured in term of ℓ_1 distance over the parameters. Barthe et al. (2016) study a verification method for probabilistic programs including constructs for Bayesian inference. In their work they also study an approach based on the use of the exponential mechanism with the Hellinger distance as scoring function. We show here that their approach achieve very poor utility.

3. Preliminaries

Bayesian inference. Given a prior belief $\Pr(\theta)$ on some parameter θ, and an observation $\mathbf{x}$, the posterior distribution on θ given $\mathbf{x}$ is defined by Bayes' Theorem 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. In parametric statistics, prior distributions and likelihood functions are usually chosen so that the posterior belongs to the same *family* of distributions as the prior. 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. This approach is implemented in a precise or approximate way in the compiler of several probabilistic programming languages ().

Dirichlet-multinomial model. In this work we will focus on the Dirichlet-Multinomial model for categorical data and its restriction to binary data: the Beta-Binomial model. That is, we will consider the situation where the underlying categorical data is drawn from a multinomial distribution with parameter the vector $\boldsymbol{\theta} \in [0,1]^k$. The prior distribution over $\boldsymbol{\theta}$ is given by a Dirichlet distribution, $\mathsf{Dir}(\boldsymbol{\alpha})$, for, $\boldsymbol{\alpha} \in (\mathbb{R}^+)^k$ and $k \in \mathbb{N}$, with p.d.f:

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

where $\mathrm{B}(\cdot)$ is the generalized beta function. The data $\mathbf{x}$ is a sequence of $n \in \mathbb{N}$ values coming from a universe $\mathcal{X}$, such that $|\mathcal{X}| = k$. The likelihood function is $\Pr(\mathbf{x}|\boldsymbol{\theta}) = \prod_{a_i \in \mathcal{X}} \theta_i^{\Delta\alpha_i}$, with $\Delta\alpha_i = \sum_{j=1}^{n} \mathbf{1}_{[x_j = a_i]}$. Denoting by $\Delta\boldsymbol{\alpha}$ the vector $(\Delta\alpha_1, \ldots \Delta\alpha_k)$ the posterior distribution over

$\boldsymbol{\theta}$ is $\mathsf{Dir}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha})$, where $+$ denotes the componentwise sum of vectors. In the following we will use the notation $\mathsf{DirP}(()\boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{x})$ to denote the posterior of this process.

The Beta-Binomial model can be obtained from the Dirichlet-Multinomial model by fixing $k = 2$.

Hellinger distance. The *Hellinger* distance $\mathcal{H}(\cdot, \cdot)$ is a statistical divergence measuring the distance between two probability distributions. Given two probability distributions P and Q, the Hellinger distance between them is defined as:

$$\mathcal{H}(P, Q) = \sqrt{\frac{1}{2} \int \left(\frac{dP}{d\lambda} - \frac{dQ}{d\lambda} \right)^2 d\lambda}$$

When we consider the Hellinger distance between Dirichlet distributions we can simplify this definition to close form. If $\mathrm{P} = \mathsf{Dir}(\boldsymbol{\alpha}_1)$ and $\mathrm{Q} = \mathsf{Dir}(\boldsymbol{\alpha}_2)$ we have:

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

Differential Privacy. *Differential privacy* (Dwork et al., 2006) is a quantitative notion of data privacy guaranteeing a bound on the that changing the input of a mechanism slightly only reflects in a controlled and bounded way on the output distribution. Formally, the ϵ parameter controls how much two outputs starting from *close* inputs can differ. When two inputs are close we call them *adjacent*. By adjacency we usually mean a relation over input databases which includes all the pair of databases which differ by at most one row. The formal definition of ϵ-differential privacy follows.

Definition 1. *A randomized mechanism $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ and codomain $\mathcal{O}$, is ϵ-differential private, if for all adjacent* [1] $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{N}^{|\mathcal{X}|}$ *and all $\mathcal{S} \subseteq \mathcal{O}$:*

$$\Pr[\mathcal{M}(\boldsymbol{x}) \in \mathcal{S}] \le e^{\epsilon} \Pr[\mathcal{M}(\boldsymbol{x}') \in \mathcal{S}].$$

It is possible to achieve ϵ-differential privacy through use and composition of a few basic components. Some of these components are now reviewed.

Laplace mechanism. The Laplace mechanism releases a random number drawn from the Laplace distribution with mean $\mu = q(d)$ and scale ν. Where q is a numeric query on a database d. If $\nu \ge \frac{\Delta_f}{\epsilon}$, with $\Delta_f = \max\limits_{d,d':\mathbf{adj}(d,d')} |f(d_1) - f(d_2)|$ then the mechanism is ϵ-differentially private as proven in (**?**).

Exponential mechanism.

Composition.

Postprocessing.

A note on input data representation. We will think of our input databases $\mathbf{x}$ of length n as histograms with k-1 bins where k is the dimensionality of the problem. This is w.l.o.g. because given the first k-1 counts, the k-th count is known.

4. Problem Statement

We are interested in exploring the design space of mechanisms for privately releasing the posterior distribution for the Dirichlet-Multinomial model by output perturbation with accuracy measured in Hellinger distance. We assume that the prior is given and non-private and we aim at protecting the privacy of the observed data. It is worth noticing that the posterior distribution of the Dirichlet-Multinomial model is fully characterized by its parameters. That is, we are interested in releasing a private version of the Dirichlet distribution with parameter $\boldsymbol{\alpha}' = (\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is the parameter of the prior, and $\Delta\boldsymbol{\alpha}$ is computed from the data as discussed in the previous section. In the case of the Beta-Binomial model this specialize to releasing a private version of the Beta distribution with parameters $(\alpha', \beta') = (\alpha + \Delta\alpha, \beta + n - \Delta\alpha)$, where α and β are the parameters of the prior, and $\Delta\alpha$ is the number of 1 in the data.

Previous work by Zhang et al. (2016) and Xiao & Xiong (2012) have considered the problem of releasing the posterior of the Beta-Binomial model by output perturbation with accuracy measured in terms of ℓ_1 norm. In addition, Zhang et al. (2016) has also considered KL-divergence between posteriors. In this work we consider instead the Hellinger distance between posteriors. Similarly to KL-divergence, Hellinger distance is an f-divergence measuring levels of similarities between distributions. However, unlikely KL-divergence, Hellinger distance is a proper distance satisfying reflexivity, symmetry and triangle inequality. These properties gives Hellinger distance analytical advantages and makes the minimum hellinger distance an efficient and robust estimator ().

One of the challenges in working with the Hellinger distance for the Dirichlet-multinomial model is that a difference in the parameters of two Dirichlet distributions can results in different values in terms of Hellinger distance depending on the value of the parameter themselves. The plot in Figure 1 illustrates this by showing the Hellinger distance between beta distributions whose parameters differ by one, for each value of the parameter α. The reason why this change is relevant for differential privacy is that this quantity corresponds to the local sensitivity of the function $\mathcal{H}$.

5. Private Mechanisms

In this section we describe and analyze a variety of mechanisms which privately compute and release a posterior

[1] Given $\mathbf{x}, \mathbf{x}' \in \{0,1\}^n$ we say that $\mathbf{x}$ and $\mathbf{x}'$ are adjacent and we write, iff $\sum_i^n [x_i = x'_i] \le 1$.

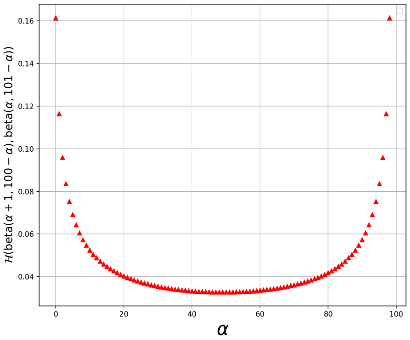


Figure 1. Hellinger distance between the Beta distributions corresponding to two adjacent datasets of size 100 when α (the number of 1s) changes. This quantity corresponds to the local sensitivity of the function $\mathcal{H}$.

distribution. Section 5.1 presents a first mechanism, LSDim, based on Laplace noise addition calibrated to sensitivity measured with respect to the ℓ_1 norm. In Section 5.2 we observe that the accuracy of LSDim can be optimized through a better analysis of the sensitivity of the inference process. Section 5.3 describes an instantiation of the exponential mechanism, EHD, where the scoring function is based on the Hellinger distance. We will see that this mechanism performs poorly due to the very high global sensitivity of the inference process with respect to this metric. We then proceed to introduce EHDL, a variation of EHD based on local sensitivity, a quantity which is at most the global sensitivity but which can be much lower. Unfortunately, EHDL by itself is non-private, but its accuracy guarantees serve as a useful benchmark. Building on EHDL, we introduce EHDS which instead relies on the notion of *smooth* sensitivity. This mechanism is ϵ-differentially private and has good accuracy in specific cases.

When describing these mechanisms, we denote by $\mathsf{DirP}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{x})$ the posterior distribution $\mathsf{Dir}(\boldsymbol{\alpha}')$ obtained from the prior $\mathsf{Dir}(\boldsymbol{\alpha})$, likelihood function $\mathsf{Mult}(|\mathbf{x}|, \boldsymbol{\theta})$, and observed data $\mathbf{x}$. To minimize notation we will identify the distribution $\mathsf{Dir}(\boldsymbol{\alpha})$ with its vector of parameters. Two different upper bounds on the value of

$$\max_{\substack{\mathbf{adj}(\mathbf{x},\mathbf{x}'),\\ \boldsymbol{\alpha}=\mathsf{DirP}(\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{x}),\\ \boldsymbol{\alpha}'=\mathsf{DirP}(\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{x}')}} ||\boldsymbol{\alpha} - \boldsymbol{\alpha}'||_1 \tag{1}$$

will be used to calibrate the noise in the mechanisms based on the Laplace primitive. Finally, we use the abbreviation $\mathsf{DirP}(\mathbf{x}, \boldsymbol{\theta})$ to denote $\mathsf{DirP}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{x})$ when the underlying vector $\boldsymbol{\alpha}$ is understood.

5.1. Calibrating noise w.r.t. ℓ_1 norm (LSDim)

We first present LSDim, in Algorithm 1, which is based on the Laplace primitive. The idea of LSDim is to add Laplace noise with scale proportional to $|\mathcal{X}|$ (the dimension of the parameter space) to the vector of numeric parameters of the posterior distribution.

Algorithm 1 LSDim

```
input $\mathbf{x} \in \mathcal{X}^n$, $\mathsf{Dir}(\boldsymbol{\alpha})$
    let $\boldsymbol{\alpha}' = \mathsf{DirP}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{x})$
    Initialize a vector $\tilde{\boldsymbol{\alpha}} = (0, \dots, 0) \in \mathbb{N}^{|\mathcal{X}|}$
    For $i = 1 \dots |\mathcal{X}| - 1$:
        let $\eta \sim \mathsf{Lap}(0, \frac{|\mathcal{X}|}{\epsilon})$
        $\tilde{\alpha_i} = \alpha_i + \lfloor (\alpha'_i - \alpha_i) + \eta \rceil_0^n$
    $\tilde{\alpha}_{|\mathcal{X}|} = \alpha_{|\mathcal{X}|} + \lfloor n - \sum_{i=1}^{|\mathcal{X}|-1} \lfloor (\alpha'_i - \alpha_i) + \eta_i \rceil_0^n \rceil_0^n$
  return $\tilde{\boldsymbol{\alpha}}$
```

Lemma 5.1. *Algorithm 1 is ϵ-differentially private.*

Proof. Privacy follows by noticing that $|\mathcal{X}| - 1$ is an upper bound on the ℓ_1 sensitivity denoted by (1). □

5.2. Improving the sensitivity (LSHist)

By viewing the parameters of a Dirichlet distribution as a histogram, we can give a tighter upper bound on the sensitivity (1). Indeed a constant suffices independent of the dimension $|\mathcal{X}| - 1$. Algorithm LSHist in Algorithm 2 takes advantage of this fact. Indeed, when $\mathcal{X} = 2$ (corresponding to the Beta-Binomial model), we have an upper bound of 1 (1), whereas when $\mathcal{X} > 2$, we have an upper bound of 2. In Algorithm 2 we present the algorithm LSHist.

Algorithm 2 LSHist

```
input $\mathbf{x} \in \mathcal{X}^n$, $\mathsf{Dir}(\boldsymbol{\alpha})$
    let $\boldsymbol{\alpha}' = \mathsf{DirP}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{x})$
    let $k = \begin{cases} 1 & \text{if} \quad |\mathcal{X}| = 2 \\ 2 & \text{otherwise} \end{cases}$
    Initialize a vector $\tilde{\boldsymbol{\alpha}} = (0, \dots, 0) \in \mathbb{N}^{|\mathcal{X}|}$
    For $i = 1 \dots |\mathcal{X}| - 1$:
        let $\eta \sim \mathsf{Lap}(0, \frac{k}{\epsilon})$
        $\tilde{\alpha_i} = \alpha_i + \lfloor (\alpha'_i - \alpha_i) + \eta \rceil_0^n$
    $\tilde{\alpha}_{|\mathcal{X}|} = \alpha_{|\mathcal{X}|} + \lfloor n - \sum_{i=1}^{|\mathcal{X}|-1} \lfloor (\alpha'_i - \alpha_i) + \eta_i \rceil_0^n \rceil_0^n$
  return $\tilde{\boldsymbol{\alpha}}$
```

Lemma 5.2. *Algorithm 2 is ϵ-differentially private.*

Proof. This follows immediately from the upper bounds of 1 and 2 on the sensitivity (1), respectively, when $|\mathcal{X}| = 2$ and $|\mathcal{X}| > 2$. □

5.3. Variants of the Exponential Mechanism

In this section, we describe a mechanism which we denote by EHD. EHD is an instantiation of the exponential mechanism (McSherry & Talwar, 2007) where the scoring function is based on a metric over probability distributions. The idea is that EHD will sample and output a distribution with higher probability when it is closer to the real posterior; importantly, *closeness* is measured using a metric over probability distributions as opposed to a metric over the underlying space of parameters. In this work we will use the Hellinger distance, but this approach could be extended to other f-divergences. Given a prior distribution $\mathsf{Dir}(\boldsymbol{\alpha})$, we define the following set from which the mechanism samples:

$$\mathcal{R}_{\boldsymbol{\alpha}} = \{\mathsf{DirP}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{x}) \mid \mathbf{x} \in \mathcal{X}^n\}.$$

In general $|\mathcal{R}_{\boldsymbol{\alpha}}| = \binom{n+1}{|\mathcal{X}|-1}$ when $\boldsymbol{\alpha} \in \mathcal{X}^n$. In the specific case of $|\mathcal{X}| = 2$ with $\boldsymbol{\alpha} = (\alpha, \beta)$, i.e., the Beta-Binomial model, the above reduces to:

$$\mathcal{R}_{\boldsymbol{\alpha}} = \left\{ (\alpha', \beta') \middle| \begin{array}{l} \alpha' = \alpha + \Delta\alpha, \beta' = \beta + n - \Delta\alpha, \\ \Delta\alpha = \sum x_i \text{ for } \mathbf{x} \in \{0,1\}^n \end{array} \right\}.$$

A first approach: the EHD **mechanism.** EHD's pseudocode is shown in Algorithm 3. EHD is an instance of the Exponential mechanism where the score is taken to be the function: $r \mapsto -\mathcal{H}(\mathsf{DirP}(\mathbf{x}, \boldsymbol{\alpha}), r)$. Given the input $\mathbf{x} \in \mathcal{X}^n$ and a prior $\mathsf{Dir}(\boldsymbol{\alpha})$, the mechanism EHD outputs an element $r \in \mathcal{R}_{\boldsymbol{\alpha}}$ with probability:

$$\frac{\exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{DirP}(\mathbf{x},\boldsymbol{\alpha}),r)}{2 \cdot GS}\right)}{\Sigma_{r' \in \mathcal{R}_{\boldsymbol{\alpha}}} \exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{DirP}(\mathbf{x},\boldsymbol{\alpha}),r')}{2 \cdot GS}\right)},$$

where GS is the global sensitivity of the scoring function, defined as:

$$\max_{\mathbf{x},\mathbf{x}' : \mathbf{adj}(\mathbf{x},\mathbf{x}')} \max_{r \in \mathcal{R}_{\boldsymbol{\alpha}}} |\mathcal{H}(\mathsf{DirP}(\mathbf{x}, \boldsymbol{\alpha}), r) - \mathcal{H}(\mathsf{DirP}(\mathbf{x}', \boldsymbol{\alpha}), r)|$$

Algorithm 3 EHD

```
input  observed data set $\mathbf{x} \in \mathcal{X}^n$, prior: $\mathsf{Dir}(\boldsymbol{\alpha})$, $\epsilon$
    let $\mathsf{Dir}(\boldsymbol{\alpha}') = \mathsf{DirP}(\mathbf{x}, \boldsymbol{\alpha})$.
    set $z = r$ with probability $\frac{\exp(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{DirP}(\mathbf{x},\boldsymbol{\alpha}),r)}{2 \cdot GS})}{\Sigma_{r' \in \mathcal{R}_{\boldsymbol{\alpha}}} \exp(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{DirP}(\mathbf{x},\boldsymbol{\alpha}),r')}{2 \cdot GS})}$
  return $z$
```

Lemma 5.3. *Algorithm 3 is ϵ-differentially private.*

A second approach based on local sensitivity: EHDL. Figure 2 shows that the accuracy of EHD is poor in comparison to that of LSHist. This is because GS is relatively large and hence the variance of the distribution EHD is large as well. To resolve this issue we will proceed in two steps. We will first define, in Algorithm 4, EHDL, which instead of using the global sensitivity, dampens the score function proportional to the local sensitivity. This approach by itself does not guarantee differential privacy, but it suggests new insights and provides a useful benchmark for comparing accuracy. We will then proceed to modify this algorithm by using a smooth upper bound on the local sensitivity.

Algorithm 4 EHDL

```
input  observed data set $\mathbf{x} \in \mathcal{X}^n$, prior: $\mathsf{Dir}(\boldsymbol{\alpha})$, $\epsilon$
    let $\mathsf{Dir}(\boldsymbol{\alpha}') = \mathsf{DirP}(\mathbf{x}, \boldsymbol{\alpha})$.
    set $z = r$ with probability $\frac{\exp(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{DirP}(\mathbf{x},\boldsymbol{\alpha}),r)}{2 \cdot LS(\mathbf{x})})}{\Sigma_{r' \in \mathcal{R}_{\boldsymbol{\alpha}}} \exp(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{DirP}(\mathbf{x},\boldsymbol{\alpha}),r')}{2 \cdot LS(\mathbf{x})})}$
  return $z$
```

Given $\mathbf{x}$, and a prior $\mathsf{Dir}(\boldsymbol{\alpha})$, EHDL samples an element $r \in \mathcal{R}_{\boldsymbol{\alpha}}$ with probability:

$$\frac{\exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{DirP}(\mathbf{x},\boldsymbol{\alpha}),r)}{2 \cdot LS(\mathbf{x})}\right)}{\Sigma_{r' \in \mathcal{R}_{\boldsymbol{\alpha}}} \exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{DirP}(\mathbf{x},\boldsymbol{\alpha}),r')}{2 \cdot LS(\mathbf{x})}\right)}$$

where $LS(\mathbf{x})$ is the local sensitivity of the scoring function, defined as:

$$\max_{\mathbf{x}' \in \mathcal{X}^n, \mathbf{adj}(\mathbf{x},\mathbf{x}'), r \in \mathcal{R}_{\boldsymbol{\alpha}}} |\mathcal{H}(\mathsf{DirP}(\mathbf{x}, \boldsymbol{\alpha}), r) - \mathcal{H}(\mathsf{DirP}(\mathbf{x}', \boldsymbol{\alpha}), r)| \tag{2}$$

Lemma 5.4 provides a useful characterization of the above function, the interested reader can check the proof in the Appendix.

Lemma 5.4. *For $\boldsymbol{x} \in \mathcal{X}^n$,*

$$LS(\boldsymbol{x}) = \max_{\boldsymbol{x}' \in \mathcal{X}^n : \boldsymbol{adj}(\boldsymbol{x},\boldsymbol{x}')} \mathcal{H}(\mathsf{DirP}(\boldsymbol{x}, \boldsymbol{\alpha}), \mathsf{DirP}(\boldsymbol{x}', \boldsymbol{\alpha}))$$

A third approach based on smooth sensitivity: EHDS. In this section, we explore a new mechanism, which compromises between EHD and EHDL and achieves good accuracy and ϵ-differential privacy simultaneously, by scaling the noise to a smooth upper bound on the local sensitivity.

Definition 2. *For $\boldsymbol{x} \in \mathcal{X}^n$, and $\gamma \in \mathbb{R}^{>0}$, the γ-smooth sensitivity of $\mathcal{H}(\mathsf{DirP}(\boldsymbol{x}), \cdot)$ is defined as:*

$$S(\boldsymbol{x}) = \max_{\boldsymbol{x}' \in \mathcal{X}^n} \left\{ \frac{1}{\frac{1}{LS(\boldsymbol{x}')} + \gamma \cdot d(\boldsymbol{x}, \boldsymbol{x}')} \right\}, \tag{3}$$

where d is the Hamming distance between two data sets.

Theorem 5.1. *For any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$, and $\gamma \in \mathbb{R}^+$ if $\boldsymbol{adj}(\boldsymbol{x}, \boldsymbol{x}')$ then $\frac{1}{S(\boldsymbol{x})} - \frac{1}{S(\boldsymbol{x}')} \leq \gamma$.*

again, the interested reader will find the proof in Appendix. We are now ready to describe the mechanism EHDS in Algorithm 5. EHDS outputs a candidate $r \in \mathcal{R}_{\boldsymbol{\alpha}}$ with probability

$$\frac{\exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{DirP}(\mathbf{x},\boldsymbol{\alpha}),r)}{4 \cdot S(\mathbf{x})}\right)}{\Sigma_{r' \in \mathcal{R}_{\boldsymbol{\alpha}}} \exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{DirP}(\mathbf{x},\boldsymbol{\alpha}),r')}{4 \cdot S(\mathbf{x})}\right)}$$

where $S(\mathbf{x})$ is the 1-smooth sensitivity of the function $\mathcal{H}(\mathsf{DirP}(\mathbf{x}), \cdot)$.

Theorem 5.2. *Algorithm 5 is ϵ-differentially private.*

```
Algorithm 5 EHDS
input observed data set $\mathbf{x} \in \mathcal{X}^n$, prior: $\mathsf{Dir}(\boldsymbol{\alpha})$, $\epsilon$
    let $\mathsf{Dir}(\boldsymbol{\alpha}') = \mathsf{DirP}(\mathbf{x}, \boldsymbol{\alpha})$.
    set $z = r$ with probability $\frac{\exp(\frac{\epsilon \cdot u(\mathbf{x}, r)}{4 \cdot S(\mathbf{x})})}{\Sigma_{r' \in \mathcal{R}_{\boldsymbol{\alpha}}} \exp(\frac{\epsilon \cdot u(\mathbf{x}, r')}{4 \cdot S(\mathbf{x})})}$
  return $z$
```

6. Accuracy Analysis

We will observe accuracy of mechanisms. In this section, we fix α and β, and for simplicity, we define the term of difference from $\mathsf{Beta}(\alpha, \beta)$, the true posterior distribution.

$$c(t) = \mathcal{H}(\mathsf{Beta}(\alpha, \beta), \mathsf{Beta}(\alpha + t, \beta - t).$$

6.1. Accuracy Bound for Baseline Mechanisms

6.1.1. Accuracy Bound for Laplace Mechanism

Lemma 6.1. *Given* $Y \sim \mathsf{Lap}(0, b)$*, the accuracy bound developed for* LSDim *is:*

$$\Pr_{z \sim \mathsf{LSDim}(\boldsymbol{x})} [\mathcal{H}(\mathsf{DirP}(\boldsymbol{x}), z) = c(t)] = \begin{cases} \frac{1}{2}(e^{-\frac{\epsilon(t)}{\Delta}} - e^{-\frac{\epsilon(t+1)}{\Delta}}) & t \geq 0 \\ \frac{1}{2}(e^{\frac{\epsilon(t+1)}{\Delta}} - e^{\frac{\epsilon(t)}{\Delta}}) & t < 0 \end{cases}$$

where $\mathsf{Beta}(\alpha, \beta)$ *is the true posterior distribution, i.e.,* $\mathsf{DirP}(\boldsymbol{x}) = \mathsf{Beta}(\alpha, \beta)$*, and* z *be the posterior produced by Laplace mechanism, i.e.,* $z = \mathsf{Beta}(\alpha + \lfloor Y \rfloor, \beta - \lfloor Y \rfloor)$.

(Proof is in Appendix)

Instantiating t with specific values $(0, 1, 2$ for example) and $\epsilon = 1$, we get the following accuracy equations on Laplace mechanism:

$$\Pr[\mathcal{H}(\mathsf{DirP}(\mathbf{x}), z) = c(t)] = \begin{cases} 0.19673467014 & t = 0 \\ 0.11932560927 & t = 1 \\ 0.07237464051 & t = 2 \end{cases}.$$

6.1.2. Accuracy Bound for Improved Laplace Mechanism

Accuracy bound for improved Laplace mechanism is obtained from the standard Laplace Mechanism by replacing the sensitivity of $\Delta\mathsf{DirP}()$ with 1 in Beta-binomial model and 2 in Dirichlet-multinomial model.

$$\Pr_{z \sim \mathsf{LSHist}(\mathbf{x})} [\mathcal{H}(\mathsf{DirP}(\mathbf{x}), z) = c(t)] = \begin{cases} \frac{1}{2}(e^{-\frac{\epsilon(t)}{2}} - e^{-\frac{\epsilon(t+1)}{2}}) & t \geq 0 \\ \frac{1}{2}(e^{\frac{\epsilon(t+1)}{2}} - e^{\frac{\epsilon(t)}{2}}) & t < 0 \end{cases}$$

Given t with the same specific values $(0, 1, 2)$ and $\epsilon = 1$ as above, we get the following accuracy equations on improved Laplace mechanism:

$$\Pr[\mathcal{H}(\mathsf{DirP}(\mathbf{x}), z) = c(t)] = \begin{cases} 0.31606027941 & t = 0 \\ 0.11627207896 & t = 1 \\ 0.04277410743 & t = 2 \end{cases}.$$

6.2. Accuracy Bound for EHDS

Lemma 6.2. *In Beta-binomial model, we have following accuracy bound for this mechanism:*

$$\Pr_{z \sim \mathsf{EHDS}(\boldsymbol{x})} [\mathcal{H}(\mathsf{DirP}(\boldsymbol{x}), z) = c(t)] = \frac{\exp\left(\frac{-\epsilon c(t)}{4 S(\boldsymbol{x})}\right)}{\sum\limits_{r' \in \mathcal{R}_{\text{post}}} \exp\left(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{BI}(\boldsymbol{x}), r')}{4 \cdot S(\boldsymbol{x})}\right)}$$

6.3. Accuracy Comparison between EHDS, LSDim and LSHist

For comparison with Laplace mechanism, we developed the Table 1. In this table, the first column is the value of distance c from the correct posterior distribution. The first row is three mechanisms: LSDim, LSHist and EHDS. For each mechanism, we calculate the probability of outputting corresponding candidates which is at distance c from the correct posterior distribution and put into the table. We calculate the first three nearest candidates and the others trival.

When ϵ and dimensions are fixed, the probability of getting the true posterior, or posterior with certain step from Laplace mechanism is fixed whatever the data size or prior changes.

By solving equations on accuracy such as:

$$\Pr_{z \sim \mathsf{EHDS}(\mathbf{x})} [\mathcal{H}(\mathsf{DirP}(\mathbf{x}), z) = c(1)] = \Pr_{z \sim \mathsf{LSDim}(\mathbf{x})} [\mathcal{H}(\mathsf{DirP}(\mathbf{x}), z) = c(1)] = 0.19673467$$

we can get the range where we can do better than LSDim or LSHist on outputting the answer with two steps from the correct one... etc.

7. Experimental Evaluations

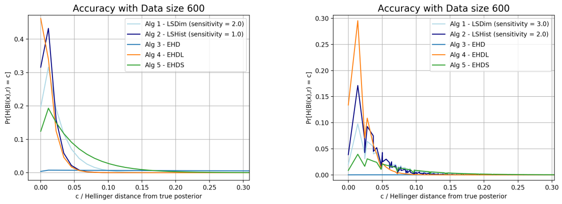
7.1. Accuracy Evaluation

7.1.1. Theoretical Results

In Fig. 2 and 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

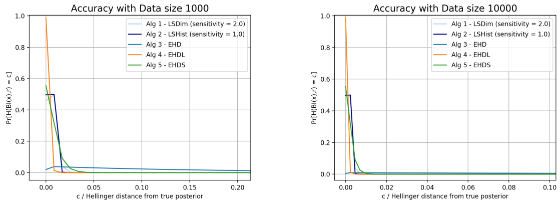
Table 1. Accuracy Comparison in Theoretical

c	0	$1-\frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)}\cdot\frac{\Gamma(\beta-\frac{1}{2})}{\Gamma(\beta)}$	$1-\sqrt{1-\frac{\frac{t}{2}}{a+\frac{t}{2}}}\cdot\sqrt{1-\frac{\frac{t}{2}}{\beta}}$	...
LSDim	0.19673467014	0.11932560927	0.07237464051	...
LSHist	0.31606027941	0.11627207896	0.04277410743	...
EHDS	$\frac{1}{\sum_{r'\in\mathcal{R}_{\text{post}}}\exp\left(\frac{-\epsilon\cdot\mathcal{H}(\mathsf{BI}(\mathbf{x}),r')}{4\cdot S(\mathbf{x})}\right)}$	$\frac{\exp\left(\frac{-\epsilon\sqrt{1-\frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)}\cdot\frac{\Gamma(\beta-\frac{1}{2})}{\Gamma(\beta)}}}{4S(\mathbf{x})}\right)}{\sum_{r'\in\mathcal{R}_{\text{post}}}\exp\left(\frac{-\epsilon\cdot\mathcal{H}(\mathsf{BI}(\mathbf{x}),r')}{4\cdot S(\mathbf{x})}\right)}$	$\frac{\exp\left(\frac{-\epsilon\sqrt{1-\sqrt{1-\frac{\frac{t}{2}}{a+\frac{t}{2}}}\cdot\sqrt{1-\frac{\frac{t}{2}}{\beta}}}}{4S(\mathbf{x})}\right)}{\sum_{r'\in\mathcal{R}_{\text{post}}}\exp\left(\frac{-\epsilon\cdot\mathcal{H}(\mathsf{BI}(\mathbf{x}),r')}{4\cdot S(\mathbf{x})}\right)}$	...



(a) 2 dimensions, data size 600 (b) 3 dimensions, data size 600

Figure 2. The theory probabilities of outputting candidates in certain distance from true posterior, with balanced data set and parameters $\epsilon = 1.0$



(a) data size 1000 (b) data size 10000

Figure 3. The theory probabilities of outputting candidates in certain distance from true posterior, with balanced data set and parameters $\epsilon = 5.0$, in 2 dimensions

0s, while for the Dirichelet-Multinomial (Figure 3(b)) the data will be split in the $k = 3$ bins with perecentages of: 33%, 33% and 34% in 3 dimensionality. Same concept in 4 dimensionality.

We consider 5 mechanisms in our comparison experiments, including the Laplace mechanism (LSDim), improved Laplace mechanism (LSHist), standard exponential mechanism (EHD), non private exponential mechanism (EHDL) and the newly designed mechanisms (EHDS with 1−smooth sensitivity achieving ϵ−dp).

Candidates of smaller distance from true posterior are considered to be good results, which can result in good accuracy. From Fig. 2, it can be derived that all these mechanisms (except standard exponential mechanism EHD in green line) can output good results with larger probability. The EHDL can output good results with higher probability than others, but EHDL is non private. For these are ϵ−differentially private, the improved laplace mechanism LSHist with ℓ_1 norm metric can produce good results with higher probability, i.e., perform better in terms of accuracy than others. As the dimension increases from 2 to 3, mechanisms' behavior remain the same.

Increasing the privacy bound ϵ, we get theoretical results as in Fig. 3 in 2 dimensions. The EHDL is still the one with the best accuracy, but it is non-private. However, the LSDim and LSHist having the similar performance, and EHDS can produce the correct posterior with higher probability than the others. And the standard exponential mechanism EHD is still the worst one.

7.1.2. Experimental Results

In this section, we evaluate the accuracy of the mechanisms defined in Section (5) w.r.t. data sizes and dimensions by experimenting on real data sets. Every plot is over 1000 runs. In all following experiments we set $\epsilon = 1.0$.

In the following whiskery-plots, the y-axis shows the accuracy (or equivalently, the error measured by the Hellinger distance) of the mechanisms, and the x-axis shows different data sets we are using. 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.

Experiments on Real Data Sets in Beta-binomial Model We use four real data sets from UCI Machine Learning repository[2]. Each data set contains 1 binary variable, which fits our Beta-binomial model. By setting the prior distribution as $\mathsf{Beta}(1, 1)$, we apply our 5 private inference algorithms on them to obtain the posterior distribution of these data. The first data set is the SHARING BIKE data, where 0 means user use sharing bike at workday and 1 stands for weekend. The second and third data set are the THERAPY data, where 1 means patient with disease is cured and 1

[2] https://archive.ics.uci.edu/ml/datasets.html

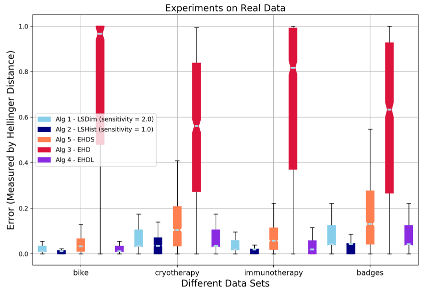


Figure 4. Accuracy on Real Data Sets, with $\epsilon = 1.0$

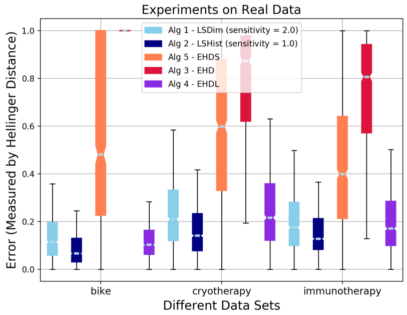


Figure 5. Accuracy on Real Data Sets, with $\epsilon = 1.0$

means not. The forth data set is the BADGE data, 1 means people with badge and 0 means without badge. The size of the four data set are $731, 90, 294, 90$. The accuracy are shown in Figure 4.

Experiments on Real Data Sets in Dirichlet-multinomial Model We use three data sets from the same repository repository. Each data set contains one category variable of 3 values ($|\mathcal{X}| = 3$), which fits our 3 dimension Dirichlet-multinomial model. By setting the prior distribution as $\mathsf{Dir}(1, 1, 1)$, we apply our 5 private inference algorithms on them to obtain the posterior distribution of these data. The first data set is the SHARING BIKE data, where $\mathcal{X} = \{0, 1, 2\}$ stand for 3 weather conditions. The second and third data set are the THERAPY data, where $0, 1, 2 \in \mathcal{X}$ represent 3 types of the diseases. The size of the four data set are $731, 90, 90$. The accuracy are shown in Figure 5.

The red box representing the EHD plays worst in terms of accuracy. While the orange one representing the EHDS has a better performance even though they achieve the same privacy bound, because we are calibrating noise to a smooth bound on Hellinger distance (the accuracy metric we are using). The purple one representing the EHDL although performs better than EHD and EHDS, but it is non-private.

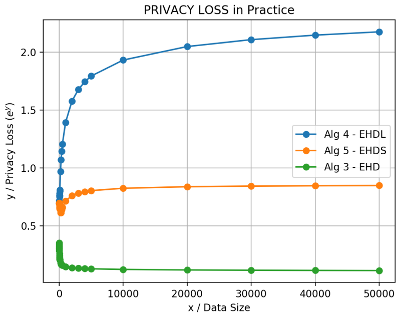


Figure 6. The privacy loss in practice under different data size when privacy bound $\epsilon = 1.0$ in 2 dimensions, prior:$\mathsf{Beta}(1, 1)$ and balanced data

7.2. Privacy Evaluation

In order to see our privacy behavior, we study the privacy loss under concrete cases. The ϵ - differential privacy we proved in Sec. 5 is just an upper bound, the ϵ in practice might be smaller than upper bound. We calculate the actual privacy loss of EHD, EHDL and EHDS w.r.t. the data size, and obtain plots in Fig. 6.

We can observe from Fig. 6 that the actual privacy loss of EHD and EHDS are smaller than the upper bound. That is to say, we achieved a higher privacy level than expected.

8. Conclusion

From what we have seen in the previous sections, we can obtain following conclusions. We explored the design space of the mechanisms for differentially privacy Bayesian inference, by considering different metrics and different algorithms.

- The accuracy can change a lot when considering different metrics and using different sensitivity values.
- Different algorithms have different performance in terms of accuracy and privacy. From the experimental results, mechanisms calibrating to ℓ_1 norm metric have fixed probability of outputting certain candidate when data size changes, while mechanisms calibrating to Hellinger distance have varying probability.
- Mechanisms calibrating to global sensitivity has large improve space in terms of accuracy. The accuracy of EHD is improved by applying a smooth bound to the sensitivity of scoring function in EHDS.

References

Barthe, G., Farina, G. P., Gaboardi, M., Arias, E. J. G., Gordon, A., Hsu, J., and Strub, P. Differentially private bayesian programming. In *CCS*, pp. 68–79, 2016.

Dimitrakakis, C., Nelson, B., Mitrokotsa, A., and Rubinstein, B. I. Robust and private bayesian inference. In *ALT*, pp. 291–305, 2014.

Dimitrakakis, C., Nelson, B., Zhang, Z., Mitrokotsa, A., and Rubinsein, B. I. Differential privacy in a bayesian setting through posterior sampling. *Technical Report 1306.1066, arXiv*, 2015.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *TCC*, pp. 265–284. Springer-Verlag, 2006. ISBN 3-540-32731-2, 978-3-540-32731-8.

Foulds, J. R., Geumlek, J., Welling, M., and Chaudhuri, K. On the theory and practice of privacy-preserving bayesian data analysis. In *UAI*, 2016.

Geumlek, J., Song, S., and Chaudhuri, K. Renyi differential privacy mechanisms for posterior sampling. In *NIPS*, pp. 5295–5304, 2017.

McSherry, F. and Talwar, K. Mechanism design via differential privacy. In *FOCS*, 2007.

Wang, Y.-X., Fienberg, S., and Smola, A. Privacy for free: Posterior sampling and stochastic gradient monte carlo. In *ICML*, pp. 2493–2502, 2015.

Williams, O. and McSherry, F. Probabilistic inference and differential privacy. In *NIPS*, pp. 2451–2459, 2010.

Xiao, Y. and Xiong, L. Bayesian inference under differential privacy. *arXiv preprint arXiv:1203.0617*, 2012.

Zhang, J., Cormode, G., Procopiuc, C. M., Srivastava, D., and Xiao, X. Privbayes: Private data release via bayesian networks. pp. 1423–1434, 2014.

Zhang, Z., Rubinstein, B. I., Dimitrakakis, C., et al. On the differential privacy of bayesian inference. In *AAAI*, pp. 2365–2371, 2016.

Zheng, S. The differential privacy of bayesian inference. In *Bachelor's thesis, Harvard College*, 2015.