Tailoring Differentially Private Bayesian Inference to Distance Between Distributions

Jiawen Liu*, Mark Bun**, Gian Pietro Farina*, and Marco Gaboardi*

*Department of Computer Science and Engineering, University at Buffalo, SUNY. {jliu223,gianpiet,gaboardi}@buffalo.edu
**Department of Computer Science, Princeton University. mbun@cs.princeton.edu

October 23, 2018

1 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 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.

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 ($\sim$ *binomial*(θ)), where θ 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, $\mathsf{beta}(\alpha, \beta)$, 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 $\mathbf{x} = (x_1, \dots x_n), x_i \in \{0, 1\}$, 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:

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

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 $\boldsymbol{\theta}$, the vector of parameters of a categorically distributed random variable. The prior distribution over $\boldsymbol{\theta} \in [0, 1]^k$ is given by a Dirichelet distribution, $\mathsf{DL}(\boldsymbol{\alpha})$, 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 $\mid \mathcal{X} \mid = k$. 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 $(\Delta\alpha_1, \dots \Delta\alpha_k)$ 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.

Differential Privacy.

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

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

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

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

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

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

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

vwhere $\boldsymbol{x} = (x_i)_{i=1}^n$ and $\boldsymbol{x} = (x_i')_{i=1}^n$ is adjacent if there is only one j that $x_j \neq x_j'$ and $x_i = x_i'$ for $i = 1, 2, \cdots, n; i \neq j$.

2 Technical Problem Statement and Motivations

We are interested in designing a mechanism for privately releasing the full posterior distributions derived in section 1, 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 $\boldsymbol{\alpha}' = (\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha})$. [5] and [4] 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 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. Details are introduced in Sec. 3.1.

However, 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. 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 $\boldsymbol{\beta}_1 = \mathsf{beta}(\alpha_1, \beta_1)$, and $\boldsymbol{\beta}_2 = \mathsf{beta}(\alpha_2, \beta_2)$ the following equality holds

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

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 distingishing two distributions.

[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 $\sum_i^n [x_i = x_i'] \leq 1$.

- 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.

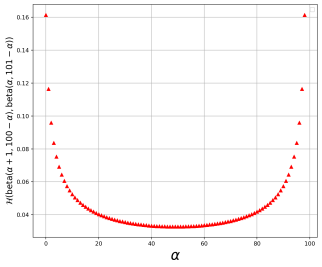

Figure 1: Sensitivity of $\mathcal{H}$

3 Mechanism Proposition

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

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

where $\Delta\alpha$ is as defined in Section1. 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 $\boldsymbol{\beta}_{\text{prior}}$.

3.1 Baseline Mechanisms

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

3.1.1 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 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[1].

3.1.2 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 $exp(\frac{\epsilon u(x,r)}{2LS(x)})$:

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

where $LS(x)$ is the local sensitivity 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[1].

3.1.3 Baseline Mechanism - Laplace Mechanism

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

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

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

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

where $Y_i \sim \mathcal{L}(0, \frac{\Delta \mathsf{BI}}{\epsilon})$ 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:

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

which is proportional to the dimensionality.

3.1.4 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 – $\mathsf{BI}(\mathbf{x}), \mathsf{BI}(\mathbf{x}')$ – which parameter differs at most in 2 dimensions even though extended to Dirichlet-multinomial mode, i.e., $\Delta \mathsf{BI} \leq 2$.

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

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

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

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

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

Both Laplace mechanism and improved one are $\epsilon-$differential privacy[1].

3.2 $\mathcal{M}_{\mathcal{H}}$: Smoothed Hellinger Distance Based Exponential Mechanism

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

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}}[z = r] = \frac{exp\big(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{BI}(\boldsymbol{x}), r)}{2 \cdot S(\boldsymbol{x})}\big)}{\sum_{r \in \mathcal{R}_{\text{post}}} exp\Big(\frac{-\epsilon \cdot \mathcal{H}(\mathsf{BI}(\boldsymbol{x}), r)}{2 \cdot S(\boldsymbol{x})}\Big)}.$$

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

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

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 [2], 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 [3]– 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 $\mathcal{M}_{\mathcal{H}}^D$ 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 $(\epsilon, \frac{e^{\frac{\epsilon}{2}}\delta}{2})-$differentially private. (or $\gamma = \ln(1 - \frac{\epsilon}{2\ln(\frac{\delta}{2|\mathcal{R}_{\text{post}}|})})$ when generalized to Dirichlet-multinomial System)

4 Privacy Analysis

4.1 Privacy Analysis for Baseline Mechanisms

In baseline mechanisms, *exponential mechanism*, *Laplace mechanism*, *improved Laplace mechanism* are $\epsilon-$differential privacy provided by [1]. The *exponential mechanism with local sensitivity* is non-differential privacy, also from [1].

4.2 Privacy Analysis for $\mathcal{M}_{\mathcal{H}}$

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

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

Lemma 4.1. *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 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. More specifically, for 3 different beta distributions $\mathsf{beta}(\alpha_0, \beta_0)$, $\mathsf{beta}(\alpha_1, \beta_1)$ and $\mathsf{beta}(\alpha_2, \beta_2)$, $\mathcal{H}(\mathsf{beta}(\alpha_0, \beta_0), \mathsf{beta}(\alpha_1, \beta_1)) = \mathcal{H}(\mathsf{beta}(\alpha_0, \beta_0), \mathsf{beta}(\alpha_2, \beta_2))$ iff $\alpha_2 = \beta_1$, $\alpha_1 = \beta_2$ and $\alpha_0 + \alpha_1 = \beta_0 + \beta_2$ and $\alpha_0 + \alpha_2 = \beta_0 + \beta_1$.

From this, we can derive that there are either only one or two candidates in $\mathcal{R}_{\text{post}}$ can have the same score, i.e., for every $r \in \mathcal{R}_{\text{post}}$ we have: $\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z = r]$ either

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

or

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

We assume the number of candidates $z \in \mathcal{R}$ that satisfy $\mathcal{H}(\mathsf{BI}(x), z) = \mathcal{H}(\mathsf{BI}(x), r)$, i.e. $u(z,x) = u(r,x)$ is $|z|$, then $\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[u(z,x) = u(r,x)] = |z| \Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z = r]$. It can be infer that $|z| = 2$ or $|z| = 1$, i.e., $\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[u(z,x) = u(r,x)] = 2 \Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z = r]$ or $\Pr_{z\sim\mathcal{M}_{\mathcal{H}}(x)}[z = r]$.

The same way in Dirichlet-multinomial system. By combination and permutation, $|z|$ can take value of 1 or 3 or 6 in 3 dimensionality, and trivially in higher dimensions.

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 $\mathcal{M}_{\mathcal{H}}(x)$: $\sum_{r' \in \mathcal{R}} exp(\frac{\epsilon u(r',x)}{2S(x)})$ 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:

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

- 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 4.2. *for any exponential mechanism $\mathcal{M}_{\mathcal{H}}(x)$, $\lambda < |\beta|$, ϵ, $|\delta| < 1$ and $\beta \leq \ln(1 - \frac{\epsilon}{2\ln(\frac{\delta}{2|\mathcal{R}_{\text{post}}|})})$, 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.

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

Proof. We partition $\mathcal{R}_{\text{post}}$ in the same way as before, 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 $-\mathcal{H}(\mathsf{BI}(x), z)$ 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 $h(c) = Pr[u(\mathcal{M}_{\mathcal{H}}(x)) = c] = \frac{exp(\frac{\epsilon c}{2S(x)})}{NL(x)}$.

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 $G = \{\mathcal{M}_{\mathcal{H}}(x) : u(\mathcal{M}_{\mathcal{H}}(x)) \leq \frac{S(x)}{(1-e^{\lambda})}\}$. 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 $Pr[c \cap G] + Pr[\overline{G}] \leq e^{\frac{\epsilon}{2}} Pr[e^{\lambda}c \cap G] + \frac{\delta}{2} \leq e^{\frac{\epsilon}{2}} Pr[e^{\lambda}c] + \frac{\delta}{2}$.

Detail proof: To show:

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

for $z \in R_1$.

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

After simplification, we need to show: $u(\mathcal{M}_{\mathcal{H}}(x)) \leq \frac{S(x)}{(1-e^{\lambda})}$. Because 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 $\lambda > 0$:

- $\lambda < 0$

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

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

  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:

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

Based on the accuracy property of exponential mechanism:

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

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

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

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

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

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

In Dirichlet-multinomial system, λ 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.

□

$(\epsilon, \frac{e^{\frac{\epsilon}{2}}\delta}{2})$−Differential Privacy of $\mathcal{M}_{\mathcal{H}}$

Lemma 4.3. *$\mathcal{M}_{\mathcal{H}}$ is $(\epsilon, \frac{e^{\frac{\epsilon}{2}}\delta}{2})$-differential privacy.*

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

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

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

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

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

- for $r \in R_2$:

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

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

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

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

Based on two cases above, $\mathcal{M}_{\mathcal{H}}$ is $(\epsilon, \delta)-$differential privacy, where δ takes the maximum value form the two cases.

Proof in Dirichlet-multinomial system can be derived in the same way by proving in parts and taking the maximum δ value of each part. □

References

[1] Cynthia Dwork, Aaron Roth, et al. *The algorithmic foundations of differential privacy.* Now Publishers, Inc., 2014.

[2] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *FOCS, 2007.*

[3] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *STOC, 2007*, pages 75–84.

[4] Yonghui Xiao and Li Xiong. Bayesian inference under differential privacy.

[5] Zuhe Zhang, Benjamin IP Rubinstein, Christos Dimitrakakis, et al. On the differential privacy of bayesian inference. In *AAAI, 2016*, pages 2365–2371.