

# Notes of DP - Bayesian Inference

Anonymous Author(s)

## 1 BAYESIAN INFERENCE BASED ON DIRICHLET-BERNOULLI DISTRIBUTION

In the Bayesian inference, first there is a prior distribution  $\pi(\xi)$  to present our belief about parameter  $\xi$ . Then, we get some observed data  $x$  sizing  $n$ , and produce a posterior distribution  $Pr(\xi|x)$ . The Bayesian inference is based on the Bayes' rule to calculate the posterior distribution:

$$Pr(\xi|x) = \frac{Pr(x|\xi)\pi(\xi)}{Pr(x)}$$

It is denoted as  $Bl(x, \pi(\xi))$  taking an observed data set  $x \in \mathcal{X}^n$  and a prior distribution  $\pi(\xi)$  as input, outputting a posterior distribution of the parameter  $\alpha$ . For conciseness, when prior is given, we use  $Bl(x)$ .  $n$  is the size of the observed data size.

In our inference algorithm, we take a Dirichlet distribution as prior belief for the parameters,  $DL(\alpha)$ , where  $\pi(\xi) = DL(\xi|\alpha) = \frac{\prod_{i=1}^m \xi_i^{\alpha_i}}{B(\alpha)}$ , and the Bernoulli distribution as the statistic model for  $Pr(x|\alpha)$ .  $m$  is the order of the Dirichlet distribution.

We give a inference process based on a concrete example, where we throw an irregular  $m$  sides dice. We want to infer the probability of getting each side  $\xi$ . We get a observed data set  $\{s_{k_1}, s_{k_2}, \dots, s_{k_n}\}$  by throwing the dice  $n$  times, where  $k_i \in \{1, 2, \dots, m\}$  denotes the side we get when we throw the dice the  $i^{th}$  time. The posterior distribution is still a Dirichlet distribution with parameters  $(\alpha_1 + n_1, \alpha_2 + n_2, \dots, \alpha_m + n_m)$ , where  $n_i$  is the appearance time of the side  $i$  in total.

In the case that  $m = 2$ , it is reduced to a Beta distribution  $beta(\alpha, \beta)$ . The  $m$  side dice change into a irregular coin with side  $A$  and side  $B$ . The posterior is computed as  $(\alpha + n_1, \beta + n_0)$ , where  $n_1$  is the appearance time of side  $A$  in the observed data set and  $n_0$  is the appearance time of the other side.

## 2 ALGORITHM SETTING UP

For now, we already have a prior distribution  $\pi(\xi)$ , an observed data set  $x$ .

### 2.1 Exponential Mechanism with Global Sensitivity

In exponential mechanism, candidate set  $\mathcal{R}$  can be obtained by enumerating  $y \in \mathcal{X}^n$ , i.e.

$$\mathcal{R} = \{Bl(y) \mid y \in \mathcal{X}^n\}.$$

Hellinger distance  $H$  is used here to score these candidates. The utility function:

$$u(x, r) = -H(Bl(x), r); r \in \mathcal{R}. \quad (1)$$

Exponential mechanism with global sensitivity selects and outputs a candidate  $r \in \mathcal{R}$  with probability proportional to  $\exp(\frac{\epsilon u(x, r)}{2\Delta_{gu}})$ :

$$P[r] = \frac{\exp(\frac{\epsilon u(x, r)}{2\Delta_{gu}})}{\sum_{r' \in \mathcal{R}} \exp(\frac{\epsilon u(x, r')}{2\Delta_{gu}})},$$

where global sensitivity is calculated by:

$$\Delta_{gu} = \max_{\{|x', y'| \leq 1; x', y' \in \mathcal{X}^n\}} \max_{\{r \in \mathcal{R}\}} |H(Bl(x'), r) - H(Bl(y'), r)|$$

The basic exponential mechanism is  $\epsilon$ -differential privacy[1].

### 2.2 Exponential Mechanism with Local Sensitivity

Exponential mechanism with local sensitivity 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  $\exp(\frac{\epsilon u(x, r)}{2\Delta_{lu}})$ :

$$P[r] = \frac{\exp(\frac{\epsilon u(x, r)}{2\Delta_{lu}})}{\sum_{r' \in \mathcal{R}} \exp(\frac{\epsilon u(x, r')}{2\Delta_{lu}})},$$

where local sensitivity is calculated by:

$$\Delta_{lu}(x) = \max_{\{|x, y'| \leq 1; y' \in \mathcal{X}^n\}} \max_{\{r \in \mathcal{R}\}} |H(Bl(x), r) - H(Bl(y'), r)|$$

The exponential mechanism with local sensitivity is non differential privacy[1].

### 2.3 Exponential Mechanism with Smooth Sensitivity

**2.3.1 Algorithm Setting up.** We define a new mechanism  $\mathcal{M}_{\mathcal{H}}^B(x)$  which is similar to the exponential mechanism where we use  $\mathcal{R}$  as the set  $\mathcal{R}_{\text{post}}$  of beta distributions with integer parameters summing up to  $n + 2$ , as scoring function we use the Hellinger distance from  $Bl(x)$ , i.e.  $H(Bl(x), -)$ , and we calibrate the noise to the smooth sensitivity [2]. The only difference is in the sensitivity part, since now we use the smooth sensitivity.

**Definition 2.1.** The mechanism  $\mathcal{M}_{\mathcal{H}}^B(x)$  outputs a candidate  $r \in \mathcal{R}_{\text{post}}$  with probability

$$Pr_{z \sim \mathcal{M}_{\mathcal{H}}^B(x)} [z = r] = \frac{\exp(\frac{-\epsilon H(Bl(x), r)}{2s_{\beta}(x)})}{\sum_{r' \in \mathcal{R}} \exp(\frac{-\epsilon H(Bl(x), r')}{2s_{\beta}(x)})},$$

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

$$s_{\beta}(x) = \max(\Delta_l H(Bl(x), -), \max_{y \neq x; y \in D^n} (\Delta_l H(Bl(y), -) \cdot e^{-\beta d(x, y)})),$$

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

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

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

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

In our private Bayesian inference mechanism, we set the  $\beta$  as  $\ln(1 - \frac{\epsilon}{2 \ln(\frac{\delta}{2(n+1)})})$ .

### 2.3.2 Sliding Property of Exponential Mechanism.

LEMMA 2.1. Consider the exponential mechanism  $\mathcal{M}_{\mathcal{H}}^S(x, u, \mathcal{R})$  calibrated on the smooth sensitivity. Let  $\lambda = f(\epsilon, \delta)$ ,  $\epsilon \geq 0$  and  $|\delta| < 1$ . Then, the following sliding property holds:

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

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

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

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

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

□

### 2.3.3 Dilation Property of Exponential Mechanism.

LEMMA 2.2. for any exponential mechanism  $\mathcal{M}_{\mathcal{H}}^B(x)$ ,  $\lambda < |\beta|$ ,  $\epsilon$ ,  $|\delta| < 1$  and  $\beta \leq \ln(1 - \frac{\epsilon}{2 \ln(\frac{\delta}{2(n+1)})})$ , the dilation property holds:

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

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

PROOF. The sensitivity is always greater than 0, and our utility function  $-H(\text{Bl}(x), z)$  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  $h(c) = \Pr[u(\mathcal{M}_{\mathcal{H}}^B(x)) = c] = 2 \frac{\exp(\frac{\epsilon c}{2S(x)})}{NL(x)}$ .

We first consider  $\lambda < 0$ . In this case,  $1 < e^\lambda$ , so the ratio  $\frac{h(c)}{h(e^\lambda c)} = \frac{\exp(\frac{\epsilon c}{2S(x)})}{\exp(\frac{\epsilon(e^\lambda c)}{2S(x)})}$  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  $\exp(\frac{\epsilon}{2} \cdot \frac{u(\mathcal{M}_{\mathcal{H}}^B(x))(1 - e^\lambda)}{S(x)})$ . Consider the event  $G = \{\mathcal{M}_{\mathcal{H}}^B(x) : u(\mathcal{M}_{\mathcal{H}}^B(x)) \leq \frac{S(x)}{(1 - e^\lambda)}\}$ . 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  $\Pr[c \cap G] + \Pr[\bar{G}] \leq e^{\frac{\epsilon}{2}} \Pr[e^\lambda c \cap G] + \frac{\delta}{2} \leq e^{\frac{\epsilon}{2}} \Pr[e^\lambda c] + \frac{\delta}{2}$ .

### Detail proof:

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

- $\lambda < 0$

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

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

is always true when  $\lambda < 0$

- $\lambda > 0$

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

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

By applying the accuracy property of exponential mechanism, we bound the probability that the equation holds with probability:

$$\Pr[u(\mathcal{M}_{\mathcal{H}}^B(x)) \leq \frac{S(x)}{(1 - e^\lambda)}] \leq \frac{|\mathcal{R}| \exp(\frac{\epsilon S(x)}{(1 - e^\lambda)} / 2S(x))}{|\mathcal{R}_{OPT}| \exp(\epsilon OPT_{u(x)} / 2S(x))}$$

In our Bayesian Inference mechanism, the size of the candidate set  $\mathcal{R}$  is equal to the size of observed data set plus 1, i.e.,  $n + 1$ , and  $OPT_{u(x)} = 0$ , then we have:

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

When we set  $\lambda \leq \ln(1 - \frac{\epsilon}{2 \ln(\frac{\delta}{2(n+1)})})$ , it is easily to derive

$$\text{that } \Pr[u(\mathcal{M}_{\mathcal{H}}^B(x)) \leq \frac{S(x)}{(1 - e^\lambda)}] \leq \frac{\delta}{2}.$$

□

### 2.3.4 Privacy Analysis.

LEMMA 2.3.  $\mathcal{M}_{\mathcal{H}}^B$  is  $(\epsilon, \delta)$ -differential privacy.

PROOF. of Lemma 2.3: For all neighboring  $x, y \in D^n$  and all sets  $\mathcal{S}$ , we need to show that:

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

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

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

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

### 3 ACCURACY ANALYSIS

#### 3.1 Laplace Mechanism

Fixing a data set  $x$ , we will sample noise  $Lap_i = \text{floor}(Y)$  where  $Y \sim \text{Lap}(\frac{2}{\epsilon})$ . Since already had the accuracy bound based on the  $l_1$  norm in Laplace mechanism:

$$\Pr[|Y| \geq t] = e^{-\frac{t\epsilon}{2}}.$$

So we have the probability  $\Pr[|Lap_i|] = \Pr[|Lap_i| \leq |Y| < |Lap_i| + 1] = e^{-\frac{|Lap_i|\epsilon}{2}} - e^{-\frac{(|Lap_i|+1)\epsilon}{2}}$ .

When we are in the case of  $m$  dimension Dirichlet distribution, we will add  $(m-1)$  i.i.d. Laplace noises  $\{|Lap_1|, |Lap_2|, \dots, |Lap_{m-1}|\}$  to the output. Then the probability of each group of noises are calculated as  $\Pr[\{|Lap_1|, |Lap_2|, \dots, |Lap_{m-1}|\}] = \Pr[|Lap_1| \leq |Y| < |Lap_1| + 1] \times \Pr[|Lap_2| \leq |Y| < |Lap_2| + 1] \times \dots \times \Pr[|Lap_{m-1}| \leq |Y| < |Lap_{m-1}| + 1] = (e^{-\frac{|Lap_1|\epsilon}{2}} - e^{-\frac{(|Lap_1|+1)\epsilon}{2}}) \times (e^{-\frac{|Lap_2|\epsilon}{2}} - e^{-\frac{(|Lap_2|+1)\epsilon}{2}}) \times \dots \times (e^{-\frac{|Lap_{m-1}|\epsilon}{2}} - e^{-\frac{(|Lap_{m-1}|+1)\epsilon}{2}})$ .

#### 3.2 Exponential Mechanism with Global Sensitivity

Also, according to [1], we already had the accuracy bound of exponential mechanism with global sensitivity as:

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}^B(x)} [H(\text{BI}(x), z) \geq c] \leq \frac{|R| \exp\left(\frac{-\epsilon c}{2\Delta_g}\right)}{|R_{OPT}| \exp\left(\frac{-\epsilon \text{OPT}_{H(\text{BI}(x), z)}(x)}{2\Delta_g}\right)},$$

where  $|R|$  is the size of the candidate set. Since  $\text{OPT}_{H(\text{BI}(x), z)}(x) = 0$  and  $|R_{OPT}| = 1$ , we simplified accuracy bound here  $\Pr_{z \sim \mathcal{M}_{\mathcal{H}}^B(x)} [H(\text{BI}(x), z) \geq c] \leq |R| \exp\left(\frac{-\epsilon c}{2\Delta_g}\right)$ .

#### 3.3 Exponential Mechanism with Smooth Sensitivity

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

First is the tight bound with very accurate calculation.

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

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

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

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

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

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  $N(n)$ .

Within the three accuracy bounds, we will mainly use the first one in following analysis.

#### 3.4 Accuracy Trade-off Between Laplace and Our Exponential Mechanism with Smooth Sensitivity

Based on accuracy study above, we are going to have a further study on the accuracy relationship between Laplace mechanism and our exponential mechanism in four aspects: data size, dimensions, prior distribution and data variance.

**3.4.1 Accuracy Analysis wrt. Data Size and Distribution Dimensions.** In this part, we will analyze the influence of data size on the two mechanisms' discrete probabilities of outputting each candidate separately. In following analysis, we will count the probabilities wrt. the steps from correct answer, in order to be more concise. For example, in the case where the correct posterior distribution is  $\text{beta}(5, 5)$ , the candidate  $\text{beta}(4, 6)$  and  $\text{beta}(6, 4)$  are of 1 steps from  $\text{beta}(5, 5)$ ; when correct posterior distribution is  $\text{DL}(5, 5, 5)$ , the candidate  $\text{DL}(4, 5, 3)$ ,  $\text{DL}(4, 3, 5)$ ,  $\text{DL}(5, 4, 3)$ ,  $\text{DL}(5, 3, 4)$ ,  $\text{DL}(3, 5, 4)$  and  $\text{DL}(5, 5, 5)$  are all of 1 step from  $\text{DL}(5, 5, 5)$ . Under the Hellinger distance measurement, if candidates are of the same steps from correct answer, they also have the same Hellinger distance from correct answer. i.e for every  $H(\text{BI}(x), z) = c$ , it will have a corresponding steps  $\text{step}(H(\text{BI}(x), z) = c) = k$ . This will make the results more clear and observable.

- In our exponential mechanism, the probability of outputting candidates wrt. steps (i.e., the hellinger distance) from the correct answer is calculated as:

$$\Pr_{z \sim \mathcal{M}_{\mathcal{H}}^B(x)} [H(\text{BI}(x), z) = c] = \sum_{\{z | H(\text{BI}(x), z) = c\}} \frac{\exp\left(\frac{-\epsilon c}{S(x)}\right)}{\sum_{r' \in R} \exp\left(\frac{-\epsilon H(\text{BI}(x), r')}{2S_\beta(x)}\right)},$$

Each candidate will occupy a portion of "1" as their outputting probability. Supposing candidates within three steps from the correct answer are good answers, the portion occupied by the good answers is decreasing when the

size of candidate set increasing. That's to say, the probabilities of outputting good answers are decreasing when the data size and dimension of prior distribution increasing. More specifically, the candidate set size is  $\sim n^{m-1}$ , which means the probabilities of outputting good answers are decreasing with speed  $\sim n^{m-1}$ .

- However in Laplace mechanism, the probability of producing noise has little relevance with the size of the candidate set. In  $m$  dimensional Dirichlet distribution, when the correct posterior distribution is  $DL(\alpha_1, \alpha_2, \dots, \alpha_m)$ , we are adding Laplace noise in this way:  $DL(\alpha_1 + Lap_1, \alpha_2 + Lap_2, \dots, \alpha_m + Lap_m)$  where  $Lap_i \sim \text{Floor}(Lap(\frac{2}{\epsilon}))$  are i.i.d. So, the probabilities of outputting a single candidate in Laplace mechanism can be obtained by:

$$\begin{aligned} Pr[(Lap_1, Lap_2, \dots, Lap_{m-1})] &= Pr[Lap_1 \leq Lap(\frac{2}{\epsilon}) < Lap_1 + 1] \\ &\times Pr[Lap_2 \leq Lap(\frac{2}{\epsilon}) < Lap_2 + 1] \times \dots \\ &\times Pr[Lap_{m-1} \leq Lap(\frac{2}{\epsilon}) < Lap_{m-1} + 1], \end{aligned}$$

where  $Pr[Lap_i \leq Lap(\frac{2}{\epsilon}) < Lap_i + 1] = \frac{1}{2}Pr[|Lap_i| \leq Lap(\frac{2}{\epsilon}) < |Lap_i| + 1]$  when  $Lap_i > 0$  and  $\frac{1}{2}Pr[|Lap_i| - 1 \leq Lap(\frac{2}{\epsilon}) < |Lap_i|]$  else.

From analysis above, we can see the probabilities of outputting the good answers will not change a lot as the size of the candidate set increasing. Specifically, we can see that the probabilities of outputting good answers are decreasing with speed upper bounded by , which means the probability of correct answer will decrease very little no matter how large the candidate set is.

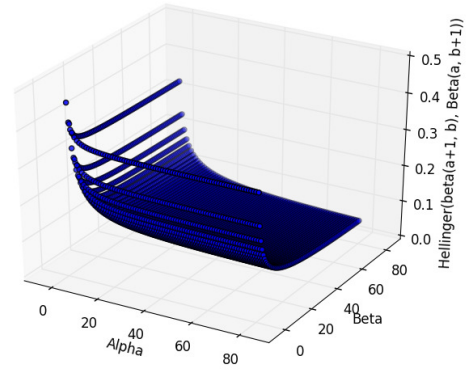
- Then, we do some concrete cases analysis.
  - when the prior is  $\text{beta}(1, 1)$ , the observed data set  $x = (1, 1, 0, 0, 1, 1, 0, 0)$ , it is easy to compute the posterior distribution:  $\text{beta}(5, 5)$ , the probability from the two mechanisms are shown in Tab. 1

Here, there are only 5 kinds of steps from correct answer. Our mechanism in the second column is clearly better than Laplace mechanism in the third column. When the candidates are close to correct answer (for example, 0, 1, 2 steps from correct answer), our mechanism can output them with higher probabilities than Laplace mechanism. On the other hand, when candidates are far away from correct answer (for example, 3, 4 steps), our mechanism can output them with lower probabilities.

- keep the prior unchanged,  $\text{beta}(1, 1)$ , the observed data set  $x = (20, 20)$  (suppose a black box will produce  $A, B$  with a certain distribution, after observing this black box continuously for 40 times, we get 20 times  $A$  and 20 times  $B$ , we can get the posterior distribution  $\text{beta}(21, 21)$  shown in Tab. 2

This case shows that the Laplace mechanism do much better than our exponential mechanism significantly. These good answers with few steps from correct answer can be outputted with higher probabilities in Laplace mechanism than in our mechanism. Moreover, in our mechanism the bad answers and good answers have very similar outputting probabilities.

**3.4.2 Accuracy Analysis wrt. prior distribution.** In this part, we firstly study the local sensitivity of Hellinger distance of different beta distributions as in Fig. ??, in order to have a better understanding of the relationships between accuracy and next two aspects.



**Figure 1: Hellinger distance study**

As in Fig. ??, the local sensitivity of Hellinger distance will decrease when beta distribution's two parameters get closer (i.e. more uniform) and larger. In the same time, our smooth sensitivity will also decrease based on the Def. 2.1. From accuracy bound in Sec. 3.3, the accuracy will be improved when sensitivity go larger. In consequence, when we increase the prior distribution, the smooth sensitivity in our exponential mechanism will decrease, which means our accuracy will be improved. However, the sensitivity of  $l_1$  norm in Laplace mechanism is fixed regardless the prior distribution. We will study this trade-off in Sec. 4.2.2.

**3.4.3 Accuracy Analysis wrt. data variance.** Similar as above, when the data variance is small (i.e. the data are uniform), the parameters of posterior distribution will get more uniform (i.e. closer). Based on Fig. ??, the Hellinger distance's sensitivity will get smaller when parameter get closer. As a result, the accuracy will be improved in the same time. We will study this trade-off in Sec. ??

## 4 EXPERIMENTAL EVALUATIONS

### 4.1 Computation Efficiency

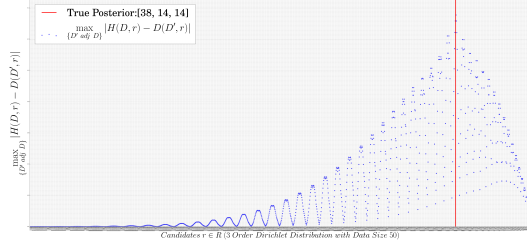
The formula for computing the local sensitivity presented in Sec. 2.2:  $\max_{\{x, y' | |x - y'| \leq 1; y' \in \mathcal{X}^n\}} \max_{\{r \in R\}} \{H(BI(x), r) - H(BI(y'), r)\}$  can be reduced

Hellinger Distance / steps	Our Exponential Mechanism	Laplace Mechanism
$Pr[H(BI(x), r) = 0.83737258593]/4$	0.0431193490585	0.066561234758
$Pr[H(BI(x), r) = 0.662174391701]/3$	0.0785621424847	0.0992976939175
$Pr[H(BI(x), r) = 0.457635865026]/2$	0.158265808563	0.148134752205
$Pr[H(BI(x), r) = 0.233629480709]/1$	0.340809715054	0.220991081918
$Pr[H(BI(x), r) = 0.0]/0$	0.37924298484	0.329679953964

**Table 1: Probability with Prior beta(1, 1) and data size 8**

Hellinger Distance / steps	Our Exponential Mechanism	Laplace Mechanism
$Pr[H(BI(x), r) = 0.99999984481]/\dots$	0.000149705644585	3.05988187701e-09
$\dots$		
$Pr[H(BI(x), r) = 0.187421762881]/2$	0.0548161224677	0.0285774941516
$Pr[H(BI(x), r) = 0.110122822057]/1$	0.192227323562	0.170131188571
$Pr[H(BI(x), r) = 0.0]/0$	0.0713016293602	0.108688872046

**Table 2: Probability with Prior beta(1, 1) and data size 40**

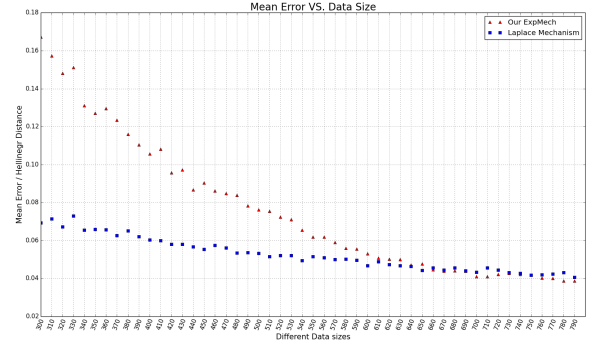


**Figure 2: Experimental Results for Finding the Local Sensitivity Efficiently**

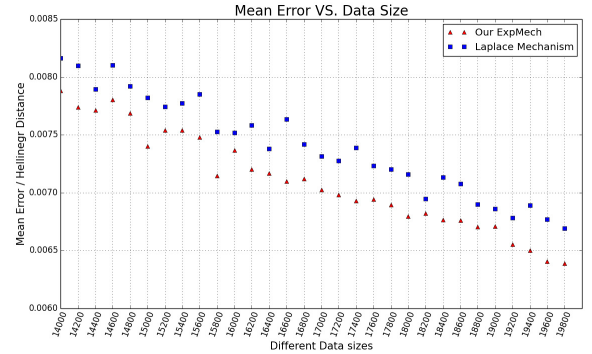
to  $\max_{\{|x, y'| \leq 1; y' \in \mathcal{X}^n\}} H(BI(x), BI(y'))$  by applying the distance triangle property. i.e., the maximum value over  $r \in R$  always happen when  $r = BI(x)$  itself, where  $\Delta_I u(x) = \max_{\{|x, y'| \leq 1; y' \in \mathcal{X}^n\}} \{H(BI(x), BI(x)) - H(BI(y'), BI(x))\} = \max_{\{|x, y'| \leq 1; y' \in \mathcal{X}^n\}} \{H(BI(y'), BI(x))\}$ . 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  $r = BI(x)$ .

## 4.2 Accuracy Trade-off Evaluation wrt. Different Variables

In this section, we evaluate the accuracy wrt. four variables, including data size, dimension, data variance and prior distribution, and some combinations of these variables. We experiment 1000 times under each value of variables and produce 4-quantile plots for each variable. In following 4-quantile plots, the y-axis is accuracy measured by Hellinger distance, x-axis is different value of variables. The blue boxes in plots represent our exponential mechanism and the next yellow box represents the Laplace mechanism under the same setting.



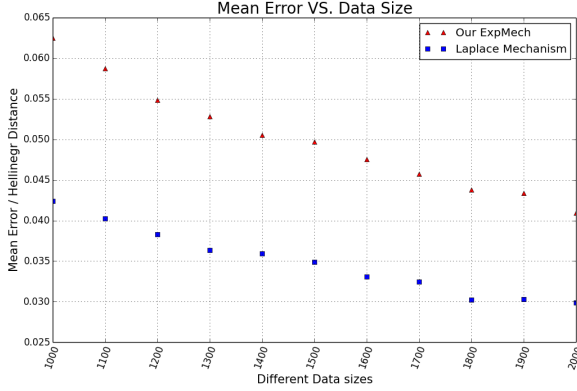
(a) Data set size from 300 to 800



(b) Data set size from 14000 to 20000

**Figure 3: Increasing data size with fixed prior beta(1, 1). Unbalanced datasets of mean (0.1, 0.9) and parameters  $\epsilon = 0.8$  and  $\delta = 10^{-8}$**

**4.2.1 Accuracy Evaluation wrt. Data size.** In Figures 3 and 4 we consider *unbalanced* datasets of observations. This means that in the



(a) Data set size from 14000 to 20000

**Figure 4: Increasing data size with fixed DL(1, 1, 1) prior distribution, Unbalanced datasets of mean (0.2, 0.3, 0.5) and parameters  $\epsilon = 0.8$  and  $\delta = 10^{-8}$**

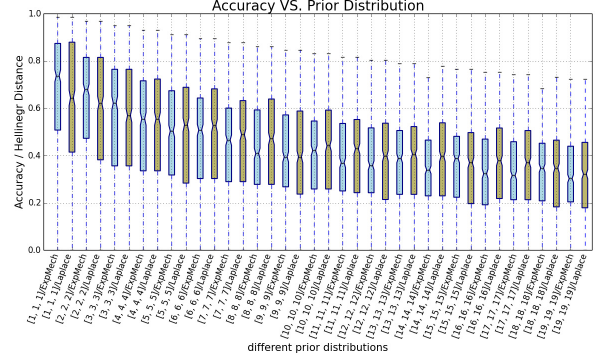
Beta-Binomial setting (Figure 3(a), and 3(b)) the datasets will consist of 10% 1s and the rest 0s, while for the Dirichelet-Multinomial (Figure ?? and 4(a)) the data will be split in the  $k = 3$  bins with percentages of: 20%, 30% and 50%. The results show that when the data size increases, the average errors of  $\mathcal{M}_{\mathcal{H}}^B$ ,  $\mathcal{M}_{\mathcal{H}}^D$ , and Laplace decrease. For small datasets, i.e with size less 650 in the case of Beta-Binomial systems, the Laplace mechanisms outperforms the exponential mechanisms, but for bigger data sets, that is, bigger than 650, or as in Figure 3(b) where we considered data sets of the order of 15 thousands elements, the exponential mechanisms outperforms the Laplace mechanism. Similar experimental tendencies were obtained for the Dirichlet-Multinomial system (Figure ?? and 4(a)), but we were not able to perform experiments with higher dimensions (4 or more) or, big datasets, e.g.  $10^4$  elements, due to too high run time of the algorithm.

**4.2.2 Accuracy Evaluation wrt. Prior Distribution.** In Figure 5, we fix the dataset to be  $\langle 5, 5, 5 \rangle$ . We also considered balanced priors with increasing values in their dimensions. The plot shows that in the beginning the Laplace mechanism performs better but it is outperformed after a while.

### 4.3 Experiment Evaluations on Privacy

In order to see our privacy behavior, we study the accurate epsilon under concrete cases in this section. The  $(\epsilon, \delta)$ -differential privacy we proved in Sec. 2.3 is just an upper bound, we concrete  $\epsilon$  should be smaller than upper bound in our exponential mechanism. We calculate the concrete privacy value in following ways wrt. the data size, and obtain plots in Fig. 6.

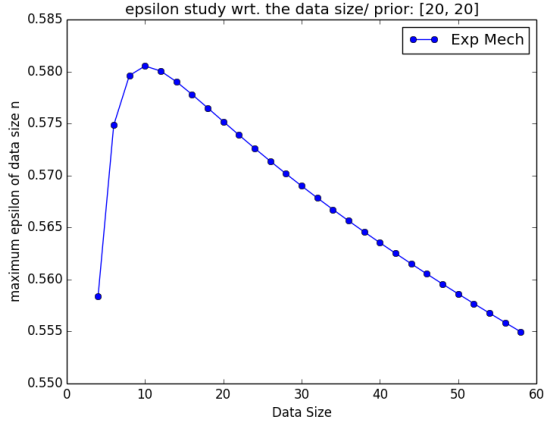
$\epsilon = 0.8$  is a privacy upper bound, we can observe that the concrete  $\epsilon$  values are smaller than the upper bound. That is to say, we achieved a higher privacy level than expected. In next step, we are going to improve the accuracy using this property.



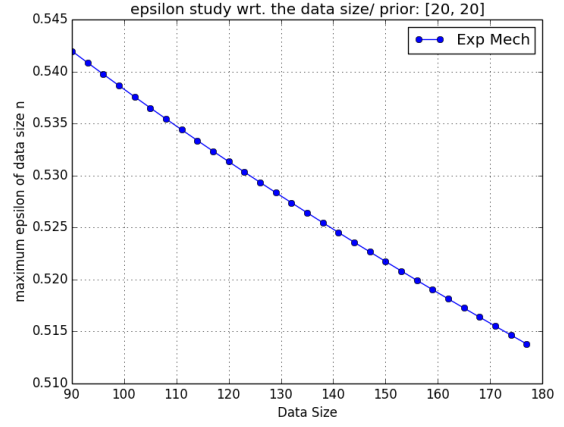
**Figure 5: Observed data set is:  $\langle 5, 5, 5 \rangle$ , varying balanced priors**

## REFERENCES

- [1] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.
- [2] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84. ACM, 2007.



(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:  $[1, 1]$ ,  $\epsilon = 0.8$ ,  $\delta = 0.0005$  and observed data are uniformly distributed**