

# Notes of DP - Bayesian Inference

## 1 Problem Statement

Our research is based on an observation that under different measurement, mechanisms usually have different accuracy behavior. When adding noise according to a certain measurement, mechanism' accuracy behavior will be better than when adding noise according to other measurements. Here we study the private Bayesian inference based on Hellinger distance measurement. Showing that our newly-designed mechanism adding noise according to Hellinger distance performs better than baseline mechanism adding noise according to  $l_1$  norm.

## 2 Technical Problem Statement

Our study based on the Bernoulli-Beta Bayesian inference. 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.

## 3 Baseline Approach - Laplace Mechanism

Based on the posterior results from Bayesian inference, we calculate the Laplace mechanism updates on posterior and get the private results under Laplace mechanism. 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, where  $Lap_i = floor(Y)$ ,  $Y \sim Lap(\frac{2}{\epsilon})$ . The private posterior then will be  $DL(\alpha_1 + n_1 + 1, \alpha_2 + n_2 + 1, \dots, \alpha_m + n_m + 1 - (Lap_1 + \dots + Lap_{m-1}))$ .

## 4 Our Approach - Exponential Mechanism with Smooth Sensitivity

### 4.1 Algorithm Setting up

We define a new mechanism  $\mathcal{M}_H(x)$  which is similar to the exponential mechanism where we use  $\mathcal{R}$  as the set  $\mathcal{R}_B$  of beta distributions with integer parameters summing up to  $n + 2$ , as scoring function we use the

Hellinger distance from  $\text{Bl}(x)$ , i.e.  $H(\text{Bl}(x), -)$ , and we calibrate the noise to the smooth sensitivity [1]. The only difference is in the sensitivity part, since now we use the smooth sensitivity.

**Definition 4.1.** The mechanism  $\mathcal{M}_H(x)$  outputs a candidate  $r \in \mathcal{R}_B$  with probability

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

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

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

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  $\Pr_{z \sim \mathcal{M}_H(x)}[z = r]$  of every  $r \in \mathcal{R}_B$  and the probability  $\Pr_{z \sim \mathcal{M}_H(x)}[H(\text{Bl}(x), z) = H(\text{Bl}(x), r)]$  for the utility score for  $r$ . In particular, for every  $r \in \mathcal{R}_B$  we have:

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

To see this, it is enough to notice that:  $\Pr_{z \sim \mathcal{M}_H(x)}[z = r]$  is proportional too  $H(\text{Bl}(x), r)$ , i.e.,  $u(x, z)$ . We can derive, if  $u(r, x) = u(r', x)$  then  $\Pr_{z \sim \mathcal{M}_H(x)}[z = r] = \Pr_{z \sim \mathcal{M}_H(x)}[z = r']$ . We assume the number of candidates  $z \in \mathcal{R}$  that satisfy  $u(z, x) = u(r, x)$  is  $|r|$ , we have  $\Pr_{z \sim \mathcal{M}_H(x)}[u(z, x) = u(r, x)] = |r| \Pr_{z \sim \mathcal{M}_H(x)}[z = r]$ . 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.,  $\Pr_{z \sim \mathcal{M}_H(x)}[u(z, x) = u(r, x)] = 2 \Pr_{z \sim \mathcal{M}_H(x)}[z = r]$ . This parameter can be eliminate in both sides in proof.

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

## 4.2 Sliding Property of Exponential Mechanism

**Lemma 4.1.** Consider the exponential mechanism  $\mathcal{M}_E^S(x, u, \mathcal{R})$  calibrated on the smooth sensitivity. Let  $\lambda = f(\epsilon, \delta)$ ,  $\epsilon \geq 0$  and  $|\delta| < 1$ . Then, the following sliding property holds:

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

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

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

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

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

□

### 4.3 Dilation Property of Exponential Mechanism

**Lemma 4.2.** *for any exponential mechanism  $\mathcal{M}_H(x)$ ,  $\lambda < |\beta|$ ,  $\epsilon, |\delta| < 1$  and  $\beta \leq \ln(1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})})$ , the dilation property holds:*

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

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

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

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  $\exp(\frac{\epsilon}{2} \cdot \frac{u(\mathcal{M}_H(x))(1-e^\lambda)}{S(x)})$ . Consider the event  $G = \{\mathcal{M}_H(x) : u(\mathcal{M}_H(x)) \leq \frac{S(x)}{(1-e^\lambda)}\}$ . 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  $\Pr[c \cap G] + \Pr[\bar{G}] \leq e^{\frac{\epsilon}{2}} \Pr[e^\lambda c \cap G] + \frac{\delta}{2} \leq e^{\frac{\epsilon}{2}} \Pr[e^\lambda c] + \frac{\delta}{2}$ .

□

### 4.4 Privacy Analysis

**Lemma 4.3.**  $\mathcal{M}_H$  is  $(\epsilon, \delta)$ -differential privacy.

*Proof.* of Lemma 4.3: For all neighboring  $x, y \in D^n$  and all sets  $\mathcal{S}$ , we need to show that:

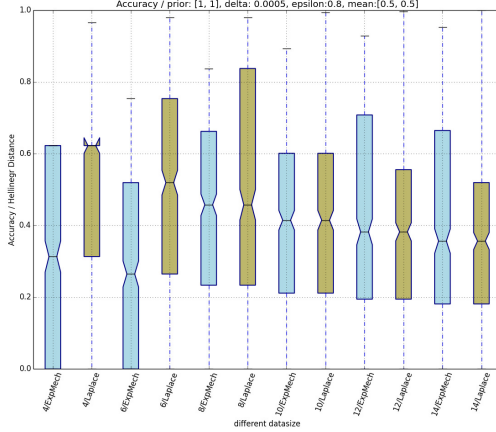
$$\Pr_{z \sim \mathcal{M}_H(x)} [z \in \mathcal{S}] \leq e^\epsilon \Pr_{z \sim \mathcal{M}_H(y)} [z \in \mathcal{S}] + \delta.$$

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

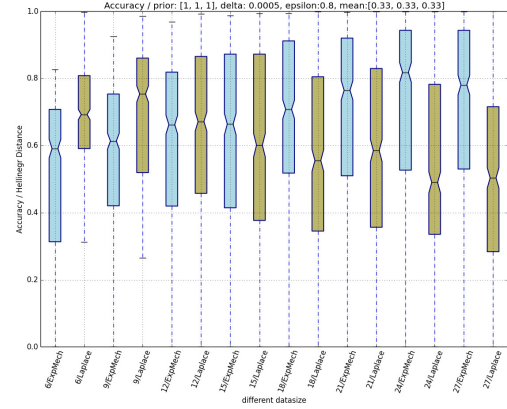
$$\begin{aligned}
2 \left( \Pr_{z \sim \mathcal{M}_H(x)} [z \in \mathcal{S}] \right) &= \Pr_{z \sim \mathcal{M}_H(x)} [u(x, z) \in \mathcal{U}] \\
&\leq e^{\epsilon/2} \cdot \Pr_{z \sim \mathcal{M}_H(x)} [u(x, z) \in \mathcal{U}_2] \\
&\leq e^\epsilon \cdot \Pr_{z \sim \mathcal{M}_H(x)} [u(x, z) \in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

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  $\frac{S(x)}{S(y)} \cdot \ln(\frac{NL(x)}{NL(y)}) \leq 1 - \frac{\epsilon}{2\ln(\frac{\delta}{2(n+1)})}$ .

□



(a) two dimensions with  $\text{beta}(1,1)$  prior distribution



(b) three dimensions with  $\text{DL}(1,1,1)$  prior distribution

Figure 1: Accuracy measurement based on Hellinger distance wrt. different datasizes. Settings: observed data are uniformly distributed,  $\epsilon = 0.8$  and  $\delta = 0.0005$

## 5 Preliminary Experimental Results

### 5.1 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.

#### 5.1.1 Accuracy Evaluation wrt. Datasize

In Fig. 1, both of the two plots show that when data size go larger, accuracy of our exponential mechanism are decreasing. In Fig. 1(a), when the data size is smaller than 12, we can beta Laplace mechanism but fail when data size larger than or equal to 12. Same as in Fig. 1(b), we can beat Laplace mechanism when data size is smaller than 15 and fail otherwise.

#### 5.1.2 Accuracy Evaluation wrt. Dimensions

In Fig. 2, x-axis are observed data sets of different size and dimensions. The plot shows that dimensions have similar influence on our exponential mechanism and the Laplace mechanism. Accuracy of two mechanisms both decrease when dimensions go larger. We will be beat by Laplace mechanism when data size increase but will not be affected when dimensions increase. In other words, dimension has little influence on whether we will beat Laplace mechanism.

#### 5.1.3 Accuracy Evaluation wrt. Data variance

In Fig. 3, x-axis are observed data sets of different variances (or means). We study this variable under two-dimension  $\text{beta}$  distribution in order to be concise. It shows that our mechanism's accuracy is better when data variance go smaller, meanwhile Laplace mechanism go worse. We will beat Laplace mechanism when observed data are more uniformly.

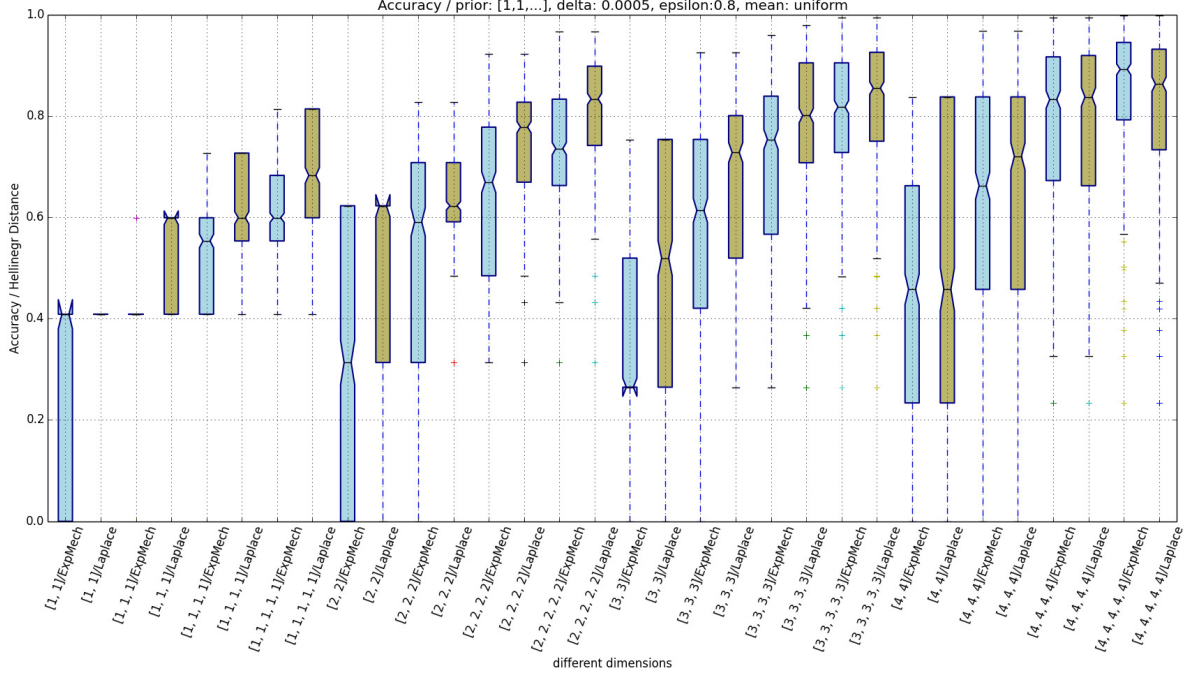


Figure 2: Accuracy measurement based on Hellinger distance wrt. different dimensions and data size. Settings: observed data are uniformly distributed,  $\epsilon = 0.8$  and  $\delta = 0.0005$ , prior distributions are all 1 in every dimension

#### 5.1.4 Accuracy Evaluation wrt. Prior Distribution

In Fig. 4, we study this variable under setting that observed data set is  $[5, 5, 5]$  because in Fig. 1 Laplace mechanism beat us when data size is 15 and uniformly distributed. The plot shows that in the beginning we cannot beat Laplace but when prior distribution grow larger, we perform better and better and beat Laplace mechanism finally.

#### 5.1.5 Accuracy Evaluation wrt. Prior Distribution and Data Variance

Here, we change the prior distribution and data variance in the same time. As shown in Fig. 5, our exponential mechanism do better in uniform data set than in edging data set while Laplace mechanism on the contrary. Moreover, our mechanism is improving continuously and significantly as prior distribution increasing while Laplace mechanism isn't.

## 6 Conclusions

We can obtain some preliminary conclusions:

1. We can beat Laplace mechanism when data size is small.
2. We will beat Laplace mechanism when observed data are more uniformly.
3. When prior distribution grow larger, we perform better and better and beat Laplace mechanism finally.

In consequence, we have a better accuracy in small data size, in larger prior and in uniformly data.

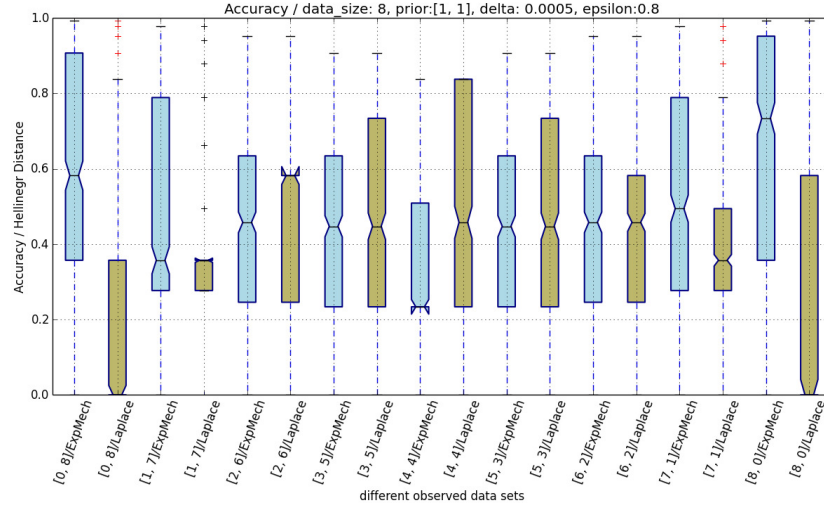


Figure 3: Accuracy measurement based on Hellinger distance wrt. different data variance. Settings:  $\epsilon = 0.8$  and  $\delta = 0.0005$ , prior distributions are all 1 in every dimension

## References

- [1] 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.

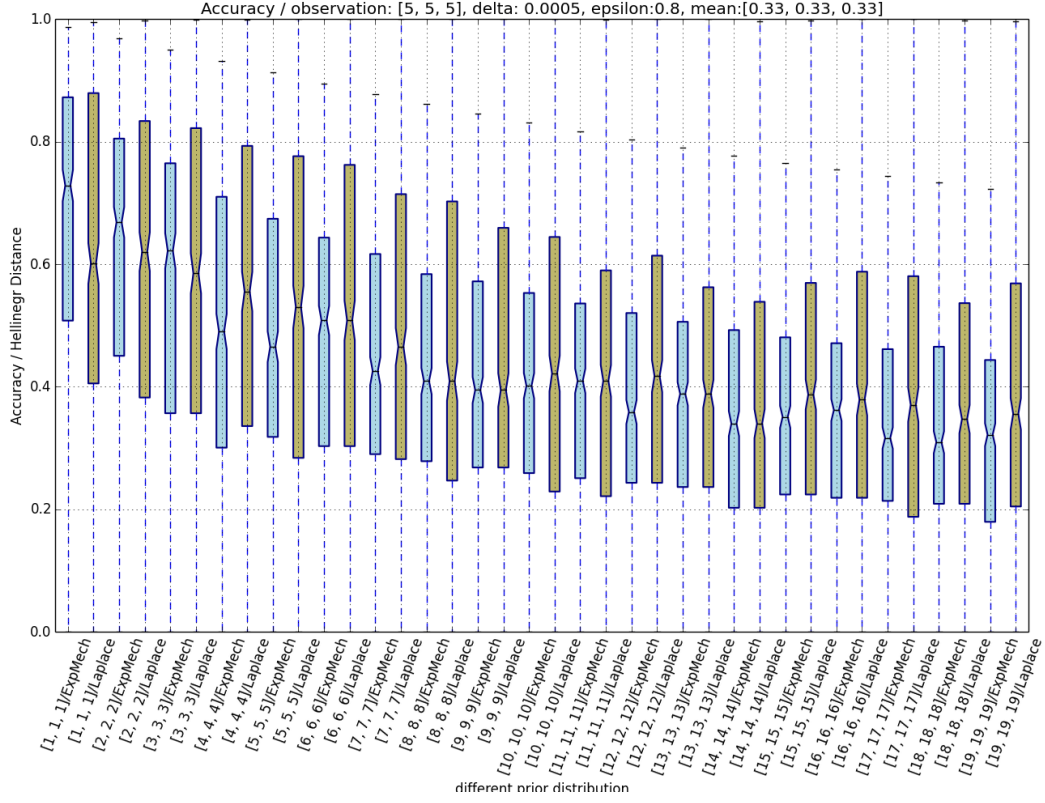


Figure 4: Accuracy measurement based on Hellinger distance wrt. different prior distribution. Settings:  $\epsilon = 0.8$  and  $\delta = 0.0005$ , observed data set is:  $[5, 5, 5]$

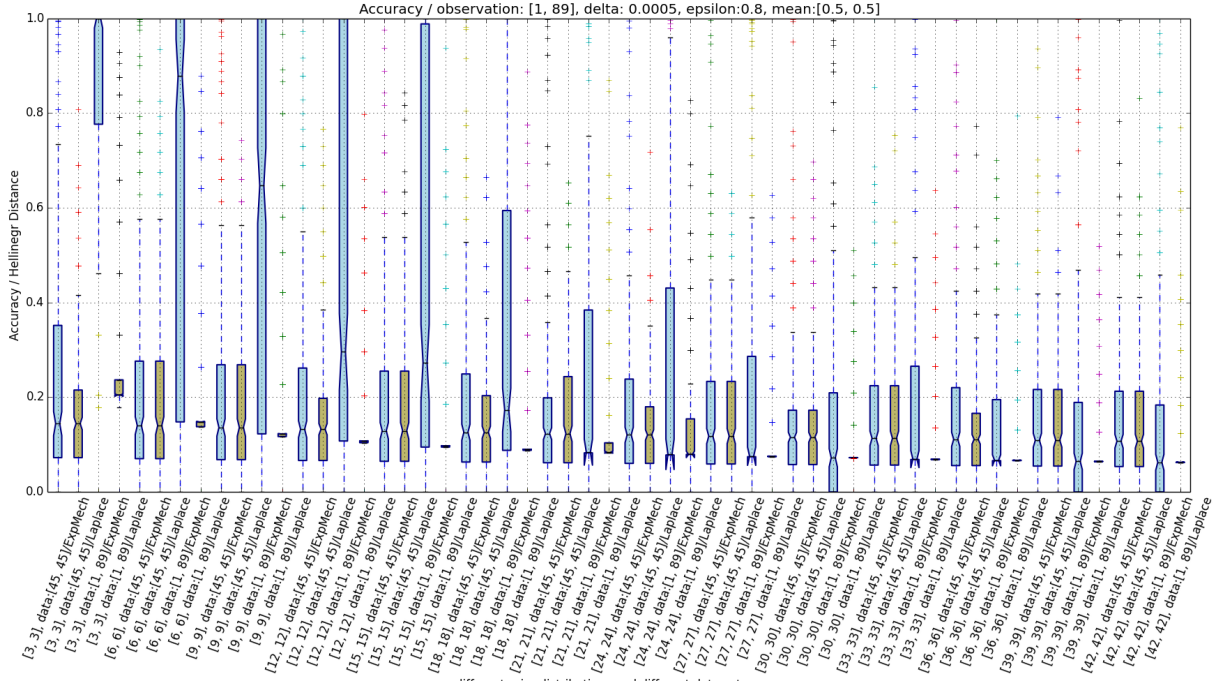


Figure 5: Experimental Results for Finding the Local Sensitivity Efficiently