Homework Assignment 1

Zhaoyang Xu
Study group: Yi Liu, Jingzheng Li, Zhaoyang Xu
September 8, 2020

Contents

1 Numerical comparison of kl inequality with its relaxations and with Hoeffding's inequality (25 points)

1.1

Hoeffding's inequality :

$$\mathbb{P}(|p-\hat{p}| \leq \sqrt{\frac{\ln\frac{1}{\delta}}{2n}}) \geq 1-\delta$$

upper bound : $p \leq \hat{p} + \sqrt{\frac{\ln\frac{1}{\delta}}{2n}}$

lower bound : $p \geq \hat{p} - \sqrt{\frac{\ln\frac{1}{\delta}}{2n}}$

kl**-inequality :** On page 17 in Yevgeny's lecture notes,

$$\mathrm{kl}(\hat{p}\|p) \leq \frac{\ln \dfrac{n+1}{\delta}}{n}$$

From question, we can get

upper bound : $p \leq \mathrm{kl}^{-1^+}(\hat{p}_n, \dfrac{\ln \dfrac{n+1}{\delta}}{n})$

lower bound : $p \geq \mathrm{kl}^{-1^-}(\hat{p}_n, \dfrac{\ln \dfrac{n+1}{\delta}}{n})$

Pinsker's relaxation of kl **inequality :** On page 18,

$$|p - \hat{p}| \leq \sqrt{\frac{\mathrm{kl}(\hat{p}\|p)}{2}} \leq \sqrt{\frac{\ln \dfrac{n+1}{\delta}}{2n}}$$

upper bound : $p \leq \hat{p} + \sqrt{\dfrac{\ln \dfrac{n+1}{\delta}}{2n}}$

lower bound : $p \geq \hat{p} - \sqrt{\dfrac{\ln \dfrac{n+1}{\delta}}{2n}}$

Refined Pinsker's relaxation of kl **inequality :** On page 18,

$$p \leq \hat{p} + \sqrt{\frac{2\hat{p}\ln \frac{n+1}{\delta}}{n}} + \frac{2\ln \frac{n+1}{\delta}}{n}$$

1.2

Hoeffding's, Pinsker's and Refined Pinsker's inequality is clearly to implement the bound. But kl-inequality is not so obvious, we want to get the inversion, which may not calculate directly. Hence, we use binary search to iterate p and get the maximum and minimum.
Firstly, we define an initial $p_i = \frac{\hat{p}_i+1}{2}$ and an initial step $\frac{(1-\hat{p}_i)}{4}$. If $\mathrm{kl} \leq \frac{\ln \frac{n+1}{\delta}}{n}$, it means the current p_i is available, we can still try to maximize it. If $\mathrm{kl} \geq \frac{\ln \frac{n+1}{\delta}}{n}$, it means the current p_i is not available, we shoule reduce it. After that, we will renew the step by $step = step/2$. Until the step smaller than a precision(I used the $\epsilon = 0.00001$), we get the final p_i. Then $\hat{p}_{i+1}$ will be used to repeat these steps. Finally, we will get the upper bound.
For lower bound, we will change the steps about renew p. If the condition about kl is satisfy, we will decrease p, otherwise increase.

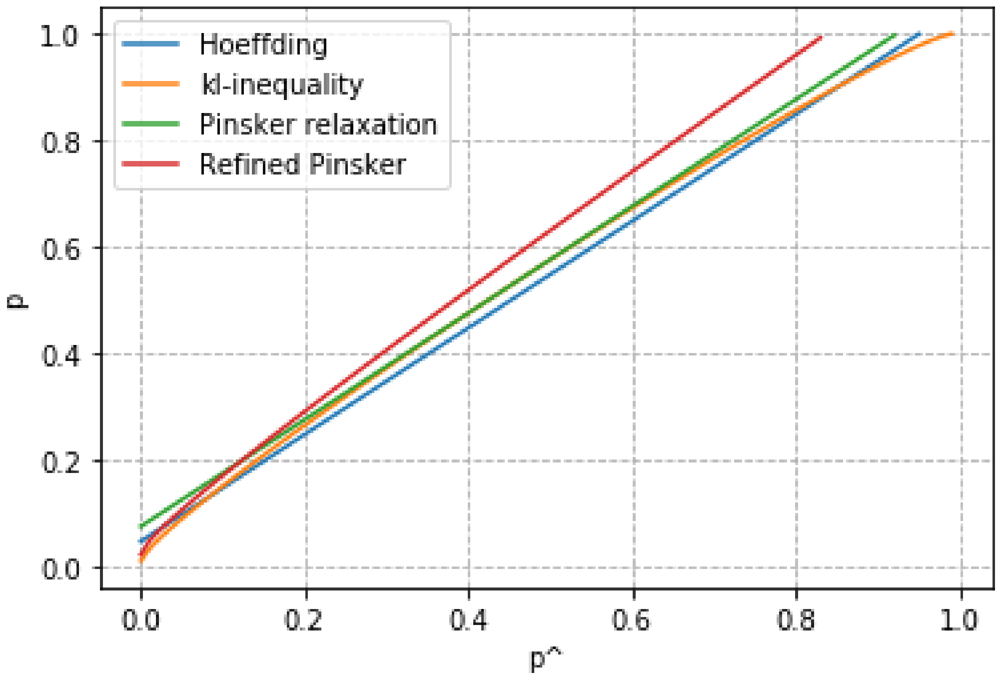


Figure 1: **Four upper bounds on p as a function of $\hat{p}_n$ (for $\hat{p}_n \in [0, 1]$, $n = 1000$, and $\delta = 0.01$)**

1.3

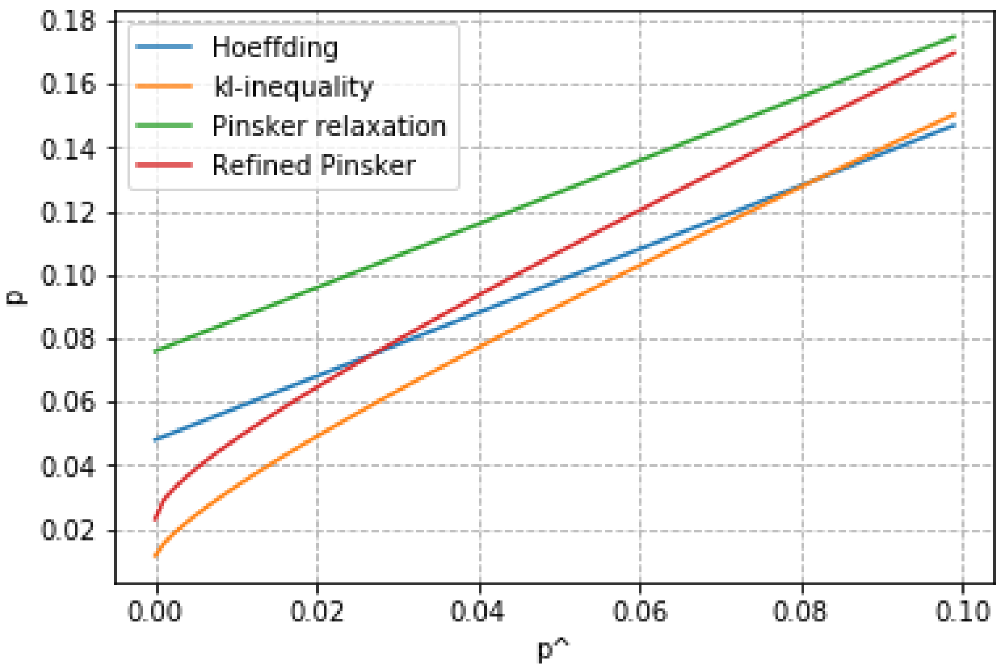
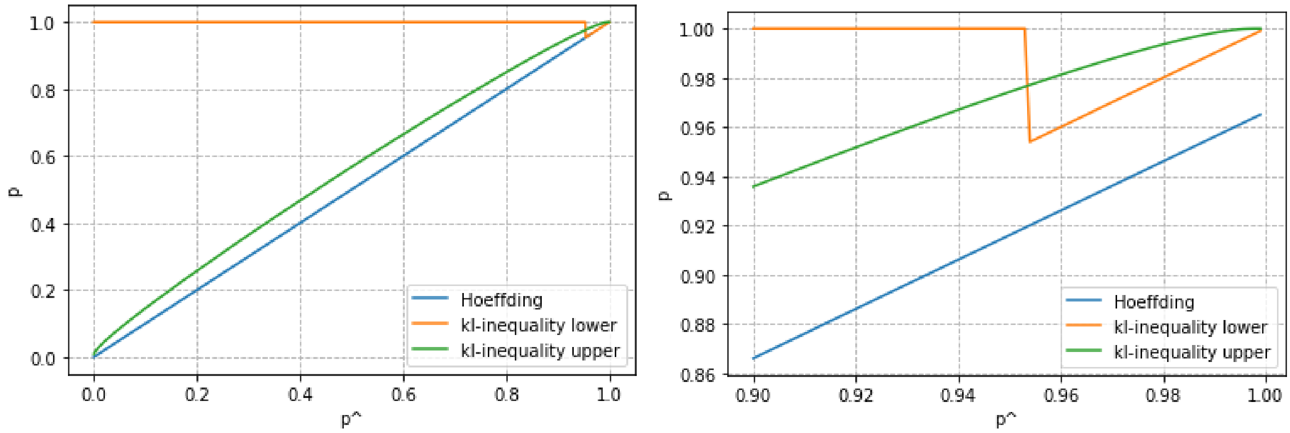


Figure 2: **Four upper bounds on p as a function of $\hat{p}_n$ (for $\hat{p}_n \in [0, 0.1]$, $n = 1000$, and $\delta = 0.01$)**

1.4



(a) $\hat{p} \in [0, 1]$ (b) $\hat{p} \in [0.9, 1]$

Figure 3: **Two lower bounds on p as a function of $\hat{p}_n$ (for $n = 1000$, and $\delta = 0.01$)**

By Fig.3(a), a part of lower bound is greater than upper bound. Because the lower bound is smaller than upper bound, we only pay attention on the interval $[0.9, 1]$, the part that the line of upper bound is higher than lower bound. When $\hat{p}$ close to 1, the lines seem to overlap. In fact, the kl-inequality still tighter. From Fig.3(b), we can see when close to 0.955, the difference is not significant, but the line of kl-inequality is still a little higher. It means the kl-inequality's lower bound is always tighter than Hoeffding's.

1.5

By Fig.1, overall, the Refined Pinsker inequality's upper bound is always the worst, if $\hat{p}$ is greater than nearly 0.3. When $\hat{p}$ in the area about $[0.2, 0.6]$, the kl-inequality's upper bound is similar with Pinsker relaxtion inequality's. Meanwhile, the Hoeffding's upper bound is the tightest in this interval. However, with $\hat{p}$ become close to 1, the kl-inequality's becomes tighter than others.
In Fig.2, it is clearly that the kl-inequality's upper bound is the tightest when $\hat{p}$ close to 0. It means that if $\hat{p}$ close to extrem value :0 and 1, the kl-inequality's will be the best. Also, the Refined Pinsker's shows tighter than Hoeffind's. When $\hat{p}$ is small, the Pinsker relaxation is the worst bound.

2 Refined Pinsker's Lower Bound (20 points)

By Lemma 2.18 on page 18 in Yevgeny's lecture notes, we can get Refined Pinsker's inequality,

$$\mathrm{kl}(p\|q) \geq \frac{(p-q)^2}{2\max\{p,q\}} + \frac{(p-q)^2}{2\max\{(1-p),(1-q)\}}$$

If $p \geq q$,

$$\begin{aligned}
\mathrm{kl}(p\|q) &\geq \frac{(p-q)^2}{2\max\{p,q\}} + \frac{(p-q)^2}{2\max\{(1-p),(1-q)\}} \\
&= \frac{(p-q)^2}{2p} + \frac{(p-q)^2}{2(1-p)} \\
&\geq \frac{(p-q)^2}{2p}
\end{aligned}$$

From question, we know $\mathrm{kl}(p\|q) \leq \epsilon$.

$$\begin{aligned}
\epsilon \geq \mathrm{kl}(p\|q) \geq \frac{(p-q)^2}{2p} &\Longrightarrow 2p\epsilon \geq (p-q)^2 \quad (p \geq q \text{ , so } p-q \geq 0) \\
&\Longrightarrow \sqrt{2p\epsilon} \geq p-q \\
&\Longrightarrow q \geq p - \sqrt{2p\epsilon}
\end{aligned}$$

If $q \geq p$,

$$\begin{aligned}
q \geq p &\Longrightarrow q - p \geq 0 \\
&\Longrightarrow q - p \geq -\sqrt{2p\epsilon} \\
&\Longrightarrow q \geq p - \sqrt{2p\epsilon}
\end{aligned}$$

So, the inequality $q \geq p - \sqrt{2p\epsilon}$ is proved.

3 Occam's razor with kl inequality (20 points)

From question, we know that we want to prove,

$$\mathbb{P}(\forall h \in \mathcal{H}, \mathrm{kl}(\hat{L}(h,S)\|L(h)) \leq \frac{\ln\frac{n+1}{\pi(h)\delta}}{n}) \geq 1-\delta$$

It can be processed like,

$$\begin{aligned}
&\Longrightarrow 1 - \mathbb{P}(\forall h \in \mathcal{H}, \mathrm{kl}(\hat{L}(h,S)\|L(h)) \leq \frac{\ln\frac{n+1}{\pi(h)\delta}}{n}) \leq \delta \\
&\Longrightarrow 1 - \mathbb{P}(\bigcup_i h_i, \mathrm{kl}(\hat{L}(h,S)\|L(h)) \leq \frac{\ln\frac{n+1}{\pi(h)\delta}}{n}) \leq \delta \\
&\Longrightarrow \mathbb{P}(\bigcap_i h_i, \mathrm{kl}(\hat{L}(h,S)\|L(h)) > \frac{\ln\frac{n+1}{\pi(h)\delta}}{n}) \leq \delta \\
&\Longrightarrow \mathbb{P}(\exists h \in \mathcal{H}, \mathrm{kl}(\hat{L}(h,S)\|L(h)) \geq \frac{\ln\frac{n+1}{\pi(h)\delta}}{n}) \leq \delta
\end{aligned}$$

Hence, our aim is to prove the last line,

$$\begin{aligned}
\mathbb{P}(\exists h \in \mathcal{H}, \mathrm{kl}(\hat{L}(h,S) \| L(h)) \geq \frac{\ln \frac{n+1}{\pi(h)\delta}}{n}) &\leq \sum_{h \in \mathcal{H}} \mathbb{P}(\mathrm{kl}(\hat{L}(h,S) \| L(h)) \geq \frac{\ln \frac{n+1}{\pi(h)\delta}}{n}) \\
&= \sum_{h \in \mathcal{H}} \mathbb{P}(n\, \mathrm{kl}(\hat{L}(h,S) \| L(h)) \geq \ln \frac{n+1}{\pi(h)\delta}) \\
&= \sum_{h \in \mathcal{H}} \mathbb{P}(e^{n\, \mathrm{kl}(\hat{L}(h,S) \| L(h))} \geq e^{\ln \frac{n+1}{\pi(h)\delta}}) \\
&= \sum_{h \in \mathcal{H}} \mathbb{P}(e^{n\, \mathrm{kl}(\hat{L}(h,S) \| L(h))} \geq \frac{n+1}{\pi(h)\delta}) \\
&\leq \sum_{h \in \mathcal{H}} \frac{\mathbb{E}(e^{n\, \mathrm{kl}(\hat{L}(h,S) \| L(h))})}{\frac{n+1}{\pi(h)\delta}} \\
&\leq \frac{n+1}{\frac{n+1}{\sum_{h \in \mathcal{H}} \pi(h)\delta}} = \sum_{h \in \mathcal{H}} \pi(h)\delta \\
&\leq \delta
\end{aligned}$$

where in the forth line we apply Markov's inequality, and the next line follows Lemma 2.14 on page 17 in Yevgeny's lecture notes.
δ is the probability that we allow things to become bad. $\pi(h)$ is the distribution of δ. For each hypothesis $h \in \mathcal{H}$ the sample S is allowed to be "non representative" with probability at most $\pi(h)\delta$.
$\pi(h)$ should be independent of S, if not the fifth line will not hold. Because it used Markov's inequality $\mathbb{P}(X \geq \epsilon) \leq \frac{\mathbb{E}[X]}{\epsilon}$. The parameter ϵ has be indenpendent of X.

4 PAC-Bayes vs. Occam (10 points)

4.1

From question, we know that we want to prove,

$$\mathbb{P}(\forall h \in \mathcal{H}, \mathrm{kl}(\mathbb{E}_\rho[\hat{L}(h,S)] \| \mathbb{E}_\rho[L(h)]) \leq \frac{\mathbb{E}_\rho[\ln \frac{1}{\pi(h)}] + \ln \frac{n+1}{\delta}}{n}) \geq 1 - \delta$$

Similar with question 4, we will prove,

$$\mathbb{P}(\exists h \in \mathcal{H}, \mathrm{kl}(\mathbb{E}_\rho[\hat{L}(h,S)] \| \mathbb{E}_\rho[L(h)]) \geq \frac{\mathbb{E}_\rho[\ln \frac{1}{\pi(h)}] + \ln \frac{n+1}{\delta}}{n}) \leq \delta$$

$$
\begin{aligned}
&\mathbb{P}(\exists h \in \mathcal{H}, \mathrm{kl}(\mathbb{E}_\rho[\hat{L}(h,S)] \| \mathbb{E}_\rho[L(h)]) \geq \frac{\mathbb{E}_\rho[\ln \frac{1}{\pi(h)}] + \ln \frac{n+1}{\delta}}{n}) \\
\leq &\mathbb{P}(\exists h \in \mathcal{H}, \mathrm{kl}(\mathbb{E}_\rho[\hat{L}(h,S)] \| \mathbb{E}_\rho[L(h)]) \geq \frac{\mathbb{E}_\rho[\ln \frac{1}{\pi(h)}] + \ln \frac{n+1}{\delta} - \mathrm{H}(p)}{n}) \\
= &\mathbb{P}(\exists h \in \mathcal{H}, \mathrm{kl}(\mathbb{E}_\rho[\hat{L}(h,S)] \| \mathbb{E}_\rho[L(h)]) \geq \frac{\mathrm{KL}(\rho\|\pi) + \ln \frac{n+1}{\delta}}{n}) \\
\leq &\delta
\end{aligned}
$$

The second step follows the decomposion of KL-divergence, the third was applied Theorem 3.26 PAC-Bayes-kl inequality on page 39 in notes.
The decomposion of KL-divergence:

$$\mathrm{KL}(\rho\|\pi) = \mathbb{E}_\rho[\ln \frac{\rho}{\pi}] = \mathbb{E}_\rho[\ln \frac{1}{\pi}] - \mathrm{H}(\rho)$$

4.2

In this question, we will prove,

$$\frac{\mathbb{E}_\rho[\ln \frac{1}{\pi(h)}] + \ln \frac{n+1}{\delta}}{n} \geq \frac{\mathrm{KL}(\rho\|\pi) + \ln \frac{n+1}{\delta}}{n}$$

We can get,

$$
\begin{aligned}
\frac{\mathbb{E}_\rho[\ln \frac{1}{\pi(h)}] + \ln \frac{n+1}{\delta}}{n} &= \frac{\mathrm{KL}(\rho\|\pi) + \mathrm{H}(\rho) + \ln \frac{n+1}{\delta}}{n} && (\mathrm{H}(\rho) \geq 0) \\
&\geq \frac{\mathrm{KL}(\rho\|\pi) + \ln \frac{n+1}{\delta}}{n}
\end{aligned}
$$

Hence, we can get the answer.