Discussion for the smoothing parameter and the sample size

Chanwoo Lee, November 12, 2020

Discussion: sanity check of the theorem when $p(\boldsymbol{X}) = p$

Consider the case when $p(\boldsymbol{X}) = p$ for $p \in (0,1)$ for all $\boldsymbol{X} \in \mathbb{X}^{d_1 \times d_2}$. Then, we verified $\mathbb{E}|\hat{p} - p| \rightarrow \sqrt{\frac{2}{\pi}\frac{p(1-p)}{n}}$ because $\hat{p} - p \rightarrow N(0, \frac{p(1-p)}{n})$, where $\hat{p} = \sum_{i=1}^n y_i/n$. Notice that

$$\mathbb{P}(|p(\boldsymbol{X}) - \pi| \leq t) = \mathbb{P}(|p - \pi| \leq t) = 0, \tag{1}$$

for all $t \in [0, |p - \pi|)$ so we have $\alpha = 1$ and $\delta = |p - \pi|$ for the Assumption 1.

We define $\bar{S}(\pi) = \{\boldsymbol{X} : \hat{p} \geq \pi\}$ and $S_{\text{bayes}}(\pi) = \{\boldsymbol{X} : p \geq \pi\} = \begin{cases} \Omega & \text{if } p \geq \pi, \\ \phi & \text{if } p < \pi. \end{cases}$ Then following Theorem 3.1. without using our bound in Theorem 4.2. we have,

$$\begin{aligned} \mathbb{E}|p(\boldsymbol{X}) - \bar{p}(\boldsymbol{X})| &\leq \max_{\pi \in \Pi} \mathbb{P}\left[\bar{S}(\pi)\Delta S_{\text{bayes}}(\pi)\right] + \frac{1}{2H} \\ &= \max_{\pi \in \Pi} \frac{1}{2}\mathbb{E}|\text{sign}(\hat{p} - \pi) - \text{sign}(p - \pi)| + \frac{1}{2H} \\ &= \max_{\pi \in \Pi} \mathbb{E}\left[\mathbb{1}\{\hat{p} - p \geq \pi - p\}\mathbb{1}\{\pi \geq p\} + \mathbb{1}\{p - \hat{p} \geq p - \pi\}\mathbb{1}\{\pi \leq p\}\right] + \frac{1}{2H} \\ &= \max_{\pi \in \Pi} \mathbb{P}\left(\hat{p} - p \geq \pi - p\right)\mathbb{1}\{\pi \geq p\} + \mathbb{P}\left(p - \hat{p} \geq p - \pi\right)\mathbb{1}\{\pi < p\} + \frac{1}{2H} \\ &\leq \exp(-n \min_{\pi \in \Pi} |\pi - p|^2) + \frac{1}{2H} \end{aligned} \tag{2}$$

For the last inequality, I used CLT $\sqrt{n}(\hat{p} - p) \rightarrow N(0, p(1-p))$. Notice that $\min_{\pi \in \Pi} |\pi - p| = \mathcal{O}\left(\frac{1}{2H}\right)$, Therefore, we have

$$\mathbb{E}|p(\boldsymbol{X}) - \bar{p}(\boldsymbol{X})| \leq \mathcal{O}\left(\exp(-n/H^2) + \frac{1}{H}\right) \leq \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),$$

with the choice of $H = (\sqrt{n})^{1-\epsilon}$. From this calculation, one thing we should notice is that $\max_{\pi \in \Pi} \mathbb{P}\left[\bar{S}(\pi)\Delta S_{\text{bayes}}(\pi)\right]$ is related to the smoothing parameter H so that H can not be arbitrary large number. Our theorem calculated $\mathbb{P}\left[\bar{S}(\pi)\Delta S_{\text{bayes}}(\pi)\right]$ with fixed smoothing parameter H as

$$\mathbb{P}\left[\bar{S}(\pi)\Delta S_{\text{bayes}}(\pi)\right] \leq \mathcal{O}\left(\frac{r(s_1 + s_2)\log d}{n}\right).$$

Therefore, when This bound is true with fixed smoothing parameter H (constant $|\pi - p|$ case) considering the result in (2)

$$\mathbb{P}\left[\bar{S}(\pi)\Delta S_{\text{bayes}}(\pi)\right] \leq \mathcal{O}\left(\exp(-n|\pi - p|^2)\right). \tag{3}$$

Therefore, my current understanding is that when H is assumed to be fixed, every theorem works smoothly. However, when we set H to diverge, current term for $\mathbb{P}\left[\bar{S}(\pi)\Delta S_{\text{bayes}}(\pi)\right]$ should be changed to consider the term H. To be specific, Assumption 1 in (1) deviates when H is arbitrary large because we cannot find fixed constant δ in that case.

What about the reference paper? Under the same setting they have $\mathbb{P}\left[\bar{S}(\pi)\Delta S_{\text{bayes}}(\pi)\right] \leq \mathcal{O}\left(\frac{1}{n}\right)$. Their $l1$ norm bound is

$$\mathbb{E}|p(\boldsymbol{X}) - \bar{p}(\boldsymbol{X})| \leq H\mathbb{P}\left[\bar{S}(\pi)\Delta S_{\text{bayes}}(\pi)\right] + \frac{1}{H} \leq H\mathcal{O}\left(\frac{1}{n}\right) + \frac{1}{H}. \tag{4}$$

by setting $H = \sqrt{n}$, they have $\mathcal{O}(1/\sqrt{n})$ which looks good but if we plug real bound (3) into (4),

$$\mathbb{E}|p(\boldsymbol{X}) - \bar{p}(\boldsymbol{X})| \leq H\mathcal{O}\left(\exp(-n|\pi - p|^2)\right) + \frac{1}{H}.$$

It gives us very rough bound.

1 Solution?

I calculated main terms on the assumptions in this case.

$$\begin{aligned}
\mathbb{P}(S(\pi)\Delta S_{\text{bayes}}(\pi)) &= \begin{cases} \mathbb{P}(\hat{p} < \pi) & \text{if } p \geq \pi, \\ \mathbb{P}(\hat{p} \geq \pi) & \text{if } p < \pi. \end{cases} \\
R_\pi(S) - R_\pi(S_{\text{bayes}}) &= \begin{cases} (p-\pi)\mathbb{P}(\hat{p} < \pi) & \text{if } p \geq \pi, \\ (\pi - p)\mathbb{P}(\hat{p} \geq \pi) & \text{if } p < \pi. \end{cases} \\
\operatorname{Var}\left[\ell(yf(\boldsymbol{X})) - \ell(yf_{\text{bayes}}(\boldsymbol{X}))\right] &= (p-\pi)^2\mathbb{P}(\hat{p} < \pi)\mathbb{P}(\hat{p} \geq \pi).
\end{aligned}$$

It is clear that Assumption 1 might not hold when $p - \pi \to 0$

$$\mathbb{P}(S(\pi)\Delta S_{\text{bayes}}(\pi)) \leq C[R_\pi(S) - R_\pi(S_{\text{bayes}})]^\alpha \iff \mathbb{P}(S(\pi)\Delta S_{\text{bayes}}(\pi)) \leq (?)|p-\pi|[R_\pi(S) - R_\pi(S_{\text{bayes}})].$$

Let me take a close look at the term $R_\pi(S) - R_\pi(S_{\text{bayes}})$ for general case. If we define the sets

$$\begin{aligned}
I &= \{\boldsymbol{X} : f(\boldsymbol{X}) \geq 0, f_{\text{bayes}}(\boldsymbol{X}) < 0\}, \\
II &= \{\boldsymbol{X} : f(\boldsymbol{X}) < 0, f_{\text{bayes}}(\boldsymbol{X}) \geq 0\}.
\end{aligned}$$

We have

$$\begin{aligned}
R_\pi(S) - R_\pi(S_{\text{bayes}}) &= \mathbb{E}\left[(1-\pi)p(\boldsymbol{X})\left(\mathbb{1}\{1 \neq \operatorname{sign}(f)\} - \mathbb{1}\{1 \neq \operatorname{sign}(f_{\text{bayes}})\}\right)\right] \\
&\quad + \mathbb{E}\left[\pi(1-p(\boldsymbol{X}))\left(\mathbb{1}\{-1 \neq \operatorname{sign}(f)\} - \mathbb{1}\{-1 \neq \operatorname{sign}(f_{\text{bayes}})\}\right)\right] \\
&= \mathbb{E}_I\left(\pi - p(\boldsymbol{X})\right) + \mathbb{E}_{II}\left(p(\boldsymbol{X}) - \pi\right).
\end{aligned}$$

Therefore, the worst case of $R_\pi(S) - R_\pi(S_{\text{bayes}})$ has $\mathcal{O}(1/H)$ order. My suggestion is to change the Assumption 1 to

$$R_\pi(S) - R_\pi(S_{\text{bayes}}) \leq C\left(H[R_\pi(S) - R_\pi(S_{\text{bayes}})]\right)^\alpha. \tag{5}$$

This assumption is more relaxed assumption compared to Assumption 1 because

$$R_\pi(S) - R_\pi(S_{\text{bayes}}) \leq H[R_\pi(S) - R_\pi(S_{\text{bayes}})].$$

This new assumption still guarantee the Assumption 2-(ii) because (5) implies Assumption 1 and Assumption 1 implies Assumption 2-(ii). The main theorem needs to be modified a little bit and briefly speaking

$$\mathbb{P}(S(\pi)\Delta S_{\text{bayes}}(\pi)) \leq \mathcal{O}\left(\left(\frac{r(s_1+s_2)\log d}{n}\right)^{\frac{\alpha}{2-\alpha}} H^\alpha\right).$$

In the above example setting where $p(\boldsymbol{X}) = p$, we can easily verify that the new assumption holds with

$\alpha = 1$ and

$$\mathbb{E}|p(\boldsymbol{X}) - \bar{p}(\boldsymbol{X})| = \mathcal{O}(1/\sqrt{n}),$$

with the choice of $H = \sqrt{n}$.