Assumption 2 in the consistency

Chanwoo Lee, July 9, 2020

For notational clarification, here $\bar{f}_\pi = \text{sign}(f_\pi)$ and $f_\pi = p(\boldsymbol{X}) - \pi$ where $p(\boldsymbol{X}) = \mathbb{P}(y = 1|\boldsymbol{X})$. I feel previous Assumption 2 in the theorem is quite strict condition. A ground truth function $p(\boldsymbol{X})$ that satisfies this condition is a linear combination of step functions. I cannot come up with another general function which holds Assumption 2 true. Because of this reason, I change Assumption 2 to more general case in the theorem and suggest previous assumption as an example.

Theorem 0.1. *Assume that*

A.1 For some positive sequence such that $s_n \to 0$ as $n \to \infty$, there exists $f_\pi^* \in \mathcal{F}_r(M)$ such that $e_V(f_\pi^*, \bar{f}_\pi) \leq s_n$.

A.2 There exist constant $0 \leq \alpha < \infty$, $a_1 > 0$ such that, for any sufficiently small $\delta > 0$.

$$\sup_{\{f \in \mathcal{F}: e_{V^T}(f, \bar{f}_\pi) \leq \delta\}} \|sign(f) - sign(\bar{f}_\pi)\|_1 \leq a_1 \delta^\alpha.$$

A.3 Considered feature space is uniformly bounded such that there exists $0 < G < \infty$ satisfying $\sqrt{\mathbb{E}\|\boldsymbol{X}\|^2} \leq G$

Then, for the estimator $\hat{p}$ obtained from our algorithm, there exists a constant a_2 such that

$$\mathbb{P}\left\{\|\hat{p} - p\|_1 \geq \frac{1}{2m} + \frac{a_1}{2}(m+1)\delta_n^{2\alpha}\right\} \leq 15 \exp\{-a_2 n (\lambda J_\pi^*)^{2-\alpha\wedge 1}\},$$

Proof. We apply Theorem 3 in [1] to our case.

The second condition of the assumption is

$$\sup_{\{f \in \mathcal{F}: e_{V^T}(f, \bar{f}_\pi) \leq \delta\}} \text{var}\{V(f, \boldsymbol{X}, y) - V(\bar{f}_\pi, \boldsymbol{X}, y) \leq a_2 \delta^\beta\}.$$

Notice that

$$\begin{aligned}
\text{var}\{V^T(f, \boldsymbol{X}, y) - V(\bar{f}_\pi, \boldsymbol{X}, y)\} &\leq \mathbb{E}|V(^T f, \boldsymbol{X}, y) - V(\bar{f}_\pi, \boldsymbol{X}, y)|^2 \\
&\leq T\mathbb{E}|V^T(f, \boldsymbol{X}, y) - V(\bar{f}_\pi, \boldsymbol{X}, y)| \\
&= T(\lambda_1 + \lambda_2).
\end{aligned}$$

where

$$\begin{aligned}
\lambda_1 &= \mathbb{E}\left|S(y)(1 - \text{sign}(yf(\boldsymbol{X})) - V(\bar{f}_\pi, \boldsymbol{X}, y)\right| = \mathbb{E}|S(y)||\text{sign}(f) - \text{sign}(\bar{f}_\pi)| \\
&\leq \|\text{sign}(f) - \text{sign}(\bar{f}_\pi)|\|_1 \leq a_1\delta^\alpha \quad \text{ from } A.2.
\end{aligned}$$

and

$$\begin{aligned}
\lambda_2 &= \mathbb{E}\left[V^T(f, \boldsymbol{X}, y) - S(y)(1 - \text{sign}(yf(\boldsymbol{X}))\right] \\
&\leq e_{V^T}(f, \bar{f}_\pi) + \mathbb{E}\left\{V(\bar{f}_\pi, \boldsymbol{X}, y) - S(y)(1 - \text{sign}(yf(\boldsymbol{X}))\right\} \\
&\leq 2e_{V^T}(f, \bar{f}_\pi) \leq 2\delta
\end{aligned}$$

Therefore, β in [1] can be replaced by $1 \wedge \alpha$.

Now we check Assumption 3 in [1]. From Lemma 2, we have

$$H_B(\epsilon, \mathcal{F}^V(k)) \leq \mathcal{O}\left(r(d_1+d_2)\log\left(\frac{Gk}{\epsilon}\right)\right).$$

Therefore, we have the following equation from Lemma 3.

$$\phi(\epsilon, k) \approx \int_{\mathcal{O}(L)}^{\mathcal{O}(\sqrt{L})} \sqrt{r(d_1+d_2)\log\left(\frac{kG}{\omega}\right)} d\omega / L \lessapprox \mathcal{O}\left(\sqrt{r(d_1+d_2)}\left(\log\left(\frac{kG}{\sqrt{L}}\right)/L\right)^{1/2}\right),$$

where $L = \min\{\epsilon^2 + \lambda(k/2-1)H_\pi^*, 1\}$. Solving Assumption 3 in [1] gives us $\epsilon_n^2 = \mathcal{O}\left(\frac{\log(n/r(d_1+d_2))+2\log(GM)}{n/r(d_1+d_2)}\right)$ by Lemma 4 when $\epsilon_n^2 \geq \lambda G J_\pi^*$. Plugging each variable into Theorem 3 proves the theorem. Notice that condition of λ is replaced because $\{\epsilon_n^2 \geq \lambda G J_\pi^*\} \subset \{\epsilon_n^2 \geq 2\lambda J_\pi^*\}$ when $rG \geq 2$. $\square$

provided that $\lambda^{-1} \geq \frac{GJ_\pi^*}{2\delta_n^2}$ *where* $J_\pi^* = \max(J(f_\pi^*), 1)$ *and* $\delta_n = \max\left(\mathcal{O}\left(\frac{\log(n/r(d_1+d_2))+2\log(GM)}{n/r(d_1+d_2)}\right), s_n\right)$.

Remark 1. We show that the Assumption 2 is satisfied when there exists $\eta > 0$ such that $|\mathbb{P}(y = 1|\boldsymbol{X}) - \pi| \geq \eta$ almost surely with respect to distribution $\boldsymbol{X}$. Smooth parameter is $a_1 = \frac{1}{\eta}$ and $\alpha = 1$ in this case.

Proof.

$$\begin{aligned}
e_{V^T}(f, \bar{f}_\pi) &= \mathbb{E}\left[S(y)L(yf(\boldsymbol{X})) \wedge T - S(y)L(y\bar{f}_\pi(\boldsymbol{X}))\right] \\
&\geq \mathbb{E}\left[S(y)(1-\text{sign}(yf(\boldsymbol{X})) - S(y)(1-\text{sign}(y\bar{f}_\pi(\boldsymbol{X}))\right] \\
&= \mathbb{E}\left[yS(y)\left(\text{sign}(\bar{f}_\pi) - \text{sign}(f)\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}(yS(y)|\boldsymbol{X})\left(\text{sign}(\bar{f}_\pi) - \text{sign}(f)\right)\right] \\
&= \mathbb{E}\left[|\mathbb{P}(y=1|\boldsymbol{X}) - \pi||\text{sign}(\bar{f}_\pi) - \text{sign}(f)|\right] \\
&\geq \eta\mathbb{E}|\text{sign}(\bar{f}_\pi) - \text{sign}(f)| = \eta\|\text{sign}(\bar{f}_\pi) - \text{sign}(f)\|_1.
\end{aligned}$$

$\square$

The main part of the proof is the following inequality

$$\mathbb{E}\left[|f_\pi||\text{sign}(f) - \text{sign}(\bar{f}_\pi)|\right] \geq \eta\mathbb{E}\left[|\text{sign}(f) - \text{sign}(\bar{f}_\pi)|\right]. \tag{1}$$

Therefore, we can replace the condition by

$$\text{For a given } \pi, \text{ there exists } \eta > 0 \text{ such that } \mathbb{E}\left[|f_\pi|\mathbb{1}_{\{\text{sign}(f)\neq\text{sign}(\bar{f}_\pi)\}}\right] \geq \eta\mathbb{E}\left[\mathbb{1}_{\{\text{sign}(f)\neq\text{ sign}(\bar{f}_\pi)\}}\right].$$

Example 1. When ground truth $p(\boldsymbol{X})$ is step function such that $p(\boldsymbol{X}) = \sum_{k=1}^K c_k \mathbb{1}_{\{\boldsymbol{X}\in A_k\}}$, then $\eta = \min_k\{|c_k - \pi|\}$.

Example 2. Assume that ground truth $p(x) = x$ and x is a random variable from Unif(0, 1). If considered function class is a set of functions with only one sign change, we show Assumption 2 holds with $\alpha = 1/2, a_1 = 2$.

Proof. We check two terms of Equation (1) in this case. If $f^{-1}(0) = \pi$, then the conclusion trivially holds. So we consider $f^{-1}(0) \neq \pi$ case. Let $f^{-1}(0) = \pi'$. Notice that right side of Equation (1) is

$$\mathbb{E}|\text{sign}(f) - \text{sign}(\bar{f}_\pi)| = 2|\pi - \pi'|.$$

The left side of the equation is

$$\mathbb{E}\left[|f_\pi||\text{sign}(f) - \text{sign}(\bar{f}_\pi)|\right] = \mathbb{E}\left[2|x-\pi|\mathbb{1}_{\{\pi\wedge\pi' < x < \pi\vee\pi'\}}\right] = \int_{\pi\wedge\pi'}^{\pi\vee\pi'} 2|x-\pi|dx = |\pi - \pi'|^2.$$

Therefore, $e_{VT}(f, \bar{f}_\pi) \geq \mathbb{E}\left[|f_\pi||\text{sign}(f) - \text{sign}(\bar{f}_\pi)|\right] = \frac{1}{4}\left(\mathbb{E}\left[|\text{sign}(f) - \text{sign}(\bar{f}_\pi)|\right]\right)^2$, which implies $\alpha = 1/2, a_1 = 2$

□

Remark 2. In Example 1, the order of ground truth function is 0 and we obtain the smooth parameter $\alpha = 1$. In Example 2, the order of ground truth function is 1 and we have the smooth parameter $\alpha = \frac{1}{2}$. We can conjecture that the smooth parameter $\alpha = \frac{1}{\text{order}(f_\pi)+1}$ because if we consider each term of the condition (1), the left side is calculated as

$$L \stackrel{\text{def}}{=} \mathbb{E}\left[|f_\pi|\mathbb{1}_{\{\text{sign}(f)\neq\text{sign}(\bar{f}_\pi)\}}\right] = \int_{\{\text{sign}(f)\neq\text{sign}(\bar{f}_\pi)\}} |f_\pi| dF(x)$$

where $F(x)$ is distribution of x. The right side is

$$R \stackrel{\text{def}}{=} \mathbb{E}\left[\mathbb{1}_{\{\text{sign}(f)\neq\text{sign}(\bar{f}_\pi)\}}\right] = \int_{\{\text{sign}(f)\neq\text{sign}(\bar{f}_\pi)\}} 1 dF(x)$$

If we consider the simple case where $\{\text{sign}(f) \neq \text{sign}(\bar{f}_\pi)\}$ is an interval, we can easily see that $L = \mathcal{O}\left((R)^{\text{order}(\bar{f}_\pi)+1}\right)$ which explains the conjecture. Therefore, Assumption 2 consider features of ground truth probability.

Remark 3. *A.1 measures how well our considered function class $\mathcal{F}_r$ approximate the ground truth function. A.2 considers the complexity of ground truth function as in Remark 2. A.3 has to do with calculating the covering number which measures the complexity of considered function class $\mathcal{F}_r$.*

Remark 4. We can think of our estimation method consisting of two parts.

S.1 Approximation of the target probability function

$$p(\boldsymbol{X}) \approx \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}_{\{\boldsymbol{X}:p(\boldsymbol{X})\leq\frac{i}{m}\}}.$$

S.2 For each i, Estimation of sublevel set

$$\mathbb{1}_{\{\boldsymbol{X}:p(\boldsymbol{X})\leq\frac{i}{m}\}} \approx \mathbb{1}_{\{\boldsymbol{X}:\text{sign}\left[\hat{f}_{\pi_i}(\boldsymbol{X}\right]=-1\}}.$$

Those estimation procedures are reflected in Theorem 0.1. As parameter m increases, the approximation in *S.1 becomes accurate but gets difficult in S.2 because of increased function complexity. The probability consistency term $\mathbb{P}\left\{\|\hat{p} - p\|_1 \geq \frac{1}{2m} + \frac{a_1}{2}(m+1)\delta_n^{2\alpha}\right\}$ shows this trade off.*

References

[1] Junhui Wang, Xiaotong Shen, and Yufeng Liu. Probability estimation for large-margin classifiers. *Biometrika*, 95(1):149–167, March 2008.