STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Mar 22

Lecturer: Xianyang Zhang

1 Symmetrization

1.1 A theorem

For $p \geq 1$,

$$\mathbb{E}\left\|\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right\|^p \leq 2^p \mathbb{E}\left\|\sum_{i=1}^{n}\epsilon_i X_i\right\|^p.$$

Proof. Let X_i' be an independent copy of X_i. Then,

$$\begin{aligned}
\mathbb{E}\left\|\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right\|^p &= \mathbb{E}\left\|\sum_{i=1}^{n}(X_i - \mathbb{E}X_i')\right\|^p = \mathbb{E}\left\|\sum_{i=1}^{n}\mathbb{E}\left[(X_i - X_i')\middle|X_i\right]\right\|^p \\
&= \mathbb{E}\left\|\mathbb{E}\left[\sum_{i=1}^{n}(X_i - X_i')\middle|X_1, ..., X_n\right]\right\|^p \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\left\|\sum_{i=1}^{n}(X_i - X_i')\right\|^p \middle|X_1, \ldots, X_n\right]\right] \text{(Jensen for } f(x) = x^p\text{)} \\
&= \mathbb{E}\left[\left\|\sum_{i=1}^{n}(X_i - X_i')\right\|^p\right].
\end{aligned}$$

Since $X_i - X_i' =^d X_i' - X_i =^d \epsilon_i(X_i - X_i')$, we have

$$\begin{aligned}
L.H.S. &\leq \mathbb{E}\left[\left\|\sum_{i=1}^{n}(X_i - X_i')\right\|^p\right] \\
&= \mathbb{E}\left\|\sum_{i=1}^{n}\epsilon_i(X_i - X_i')\right\|^p = 2^p\mathbb{E}\left[\left\|\frac{1}{2}\sum_{i=1}^{n}\epsilon_i X_i - \frac{1}{2}\sum_{i=1}^{n}\epsilon_i X_i'\right\|^p\right] \\
&\leq 2^p\left(\frac{1}{2}\mathbb{E}\left\|\sum_{i=1}^{n}\epsilon_i X_i\right\|^p + \frac{1}{2}\mathbb{E}\left\|\sum_{i=1}^{n}\epsilon_i X_i'\right\|^p\right) = 2^p\mathbb{E}\left\|\sum_{i=1}^{n}\epsilon_i X_i\right\|^p.
\end{aligned}$$

2 Uniform law of large numbers

Using symmetrization argument, we can find a sufficient condition for ULLN.

2.1 Rademancher complexity

$$P\left(\sup_{f\in\mathcal{F}}|P_n f - Pf| > \epsilon\right) \leq \frac{2}{\epsilon n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\epsilon_i f(X_i)\right|\right].$$

Proof.

$$
\begin{aligned}
P\left(\sup_{f\in\mathcal{F}}|P_nf-Pf|>\epsilon\right) &\leq \epsilon^{-1}\mathbb{E}\left[\sup_{f\in\mathcal{F}}|P_nf-Pf|\right] \text{(Markov ineq)}\\
&= \epsilon^{-1}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^n(f(X_i)-\mathbb{E}f(X_i))\right|\right]\\
&= \frac{1}{\epsilon n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\mathbb{E}\left[\sum_{i=1}^n(f(X_i)-f(X_i'))\middle|X_1,...,X_n\right]\right|\right] \text{(Symmetrization)}\\
&\leq \frac{1}{\epsilon n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\mathbb{E}\left[\left|\sum_{i=1}^n(f(X_i)-f(X_i'))\right|\middle|X_1,...,X_n\right]\right]\\
&\leq \frac{1}{\epsilon n}\mathbb{E}\left[\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n(f(X_i)-f(X_i'))\right|\middle|X_1,...,X_n\right]\right]\\
&= \frac{1}{\epsilon n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n(f(X_i)-f(X_i'))\right|\right] = \frac{1}{\epsilon n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n\epsilon_i(f(X_i)-f(X_i'))\right|\right]\\
&= \frac{1}{\epsilon n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n\epsilon_if(X_i)\right|+\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n\epsilon_if(X_i')\right|\right]\\
&= \frac{2}{\epsilon n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n\epsilon_if(X_i)\right|\right].
\end{aligned}
$$

Therefore, to show ULLN, one way is to find a bound of $\mathbb{E}\left[\sup_{f\in\mathcal{F}}|\sum_{i=1}^n\epsilon_if(X_i)|\right]$, and this quantity is called the **Rademancher complexity** of $\mathcal{F}$, and denote by $R_n(\mathcal{F})$. Therefore if $R_n(\mathcal{F})=o(n)$, ULLN holds.

There is another way to define the Rademancher complexity in $\mathbb{R}^n$: Let $\mathcal{A}\in\mathbb{R}^n$, then,

$$R_n(\mathcal{A}):=\mathbb{E}\left[\sup_{a\in\mathcal{A}}|\langle a,\epsilon\rangle|\right]$$

where $\langle,\rangle$ is a dot product and $\epsilon=(\epsilon_1,\dots,\epsilon_n)$. Define the convex hull of $\mathcal{A}$ by,

$$\text{convex}(\mathcal{A})=\left\{\sum_{i=1}^n w_ia_i\middle|a=(a_1,a_2,\dots,a_n)\in\mathcal{A},w_i\geq 0,\sum_{i=1}^n w_i=1\right\}.$$

Then $R_n(\text{convex}(\mathcal{A}))=R_n(\mathcal{A})$. (Exercise)

2.2 A version of uniform law of large numbers

Assume there exists an $F\in L_1(P)$ such that for any $f\in\mathcal{F}$, $|f|\leq F$. The function F is called an **Envelope function**. For $M>0$ and $f\in\mathcal{F}$, define

$$f_M(x)=\begin{cases} f(x) & |f(x)|\leq M\\ 0 & |f(x)|>M\end{cases},$$

and define $\mathcal{F}_M=\{f_M|f\in\mathcal{F}\}$. The following result links the connection between the covering number and the uniform law of large numbers.

Let $\mathcal{F}$ be a class of functions with the envelop function $F\in L_1(P)$. Assume for all $M>0$ and $\epsilon>0$,

$$\log N\left(\mathcal{F}_M,L_1(P_n),\epsilon\right)=o_p(n),$$

where $N(\mathcal{F}_M, L_1(P_n), \epsilon)$ is the covering number. Then,

$$\|P_n - P\|_{\mathcal{F}} \to_p 0.$$

Proof. Using the Markov inequality, it is enough to bound $\mathbb{E}[\|P_n - P\|_{\mathcal{F}}]$. By the previous result,

$$\begin{aligned}
\mathbb{E}[\|P_n - P\|_{\mathcal{F}}] &\leq \frac{2}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \epsilon_i f(X_i)\right|\right] \\
&\leq \frac{2}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \epsilon_i (f(X_i) - f_M(X_i))\right|\right] + \frac{2}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \epsilon_i f_M(X_i)\right|\right] \\
&\leq \underbrace{\frac{2}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\sum_{i=1}^n |f(X_i)| I\{|f(X_i)| > M\}\right]}_{=:A} + \underbrace{\frac{2}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \epsilon_i f_M(X_i)\right|\right]}_{=:B}.
\end{aligned}$$

By the definition of the envelope function, we have

$$\sup_{f\in\mathcal{F}} |f(X_i)| I\{|f(X_i)| > M\} \leq |F(X_i)| I\{|F(X_i)| > M\}.$$

Therefore, using $F \in L_1$,

$$A \leq 2\mathbb{E}[|F(X_i)| I\{|F(X_i)| > M\}] \to 0,$$

as $M \to \infty$. To bound B, let $\mathcal{G}$ be the minimal ϵ-cover of $\mathcal{F}_M$ ($\mathcal{G}$ is finite by assumption). Then, by the definition of ϵ-cover, for all $f \in \mathcal{F}_M$, we can find $g \in \mathcal{G}$ which satisfies $n^{-1}\sum_{i=1}^n |f(X_i) - g(X_i)| \leq \epsilon$. Therefore,

$$\left\|n^{-1}\sum_{i=1}^n \epsilon_i f(X_i)\right\| \leq \max_{g\in\mathcal{G}}\left\|n^{-1}\sum_{i=1}^n \epsilon_i g(X_i)\right\| + \epsilon.$$

Taking $\sup_{f\in\mathcal{F}_M}$ gives an upper bound of B:

$$B = 2\mathbb{E}\left[\sup_{f\in\mathcal{F}_M}\left\|n^{-1}\sum_{i=1}^n \epsilon_i f(X_i)\right\|\right] \leq 2\mathbb{E}\left[\max_{g\in\mathcal{G}}\left\|n^{-1}\sum_{i=1}^n \epsilon_i g(X_i)\right\|\right] + 2\epsilon.$$

Next, conditioning on X_i, $(\sum_{i=1}^n \epsilon_i g(X_i))$ is $\left(\sum_{i=1}^n g(X_i)^2\right)$-sub-Guassian. Therefore, use the fact that $n^{-1}\sum_{i=1}^n g(X_i)^2 \leq M^2$, it is easy to show $\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i g(X_i)\right)$ is M^2-sub-Gaussian. Therefore use the lemma from previous lecture,

$$\begin{aligned}
\mathbb{E}\left[\max_{g\in\mathcal{G}}\left\|n^{-1}\sum_{i=1}^n \epsilon_i g(X_i)\right\|\right] &= \mathbb{E}\left[\mathbb{E}\left[n^{-1/2}\max_{g\in\mathcal{G}}\left\|n^{-1/2}\sum_{i=1}^n \epsilon_i g(X_i)\right\| \middle| X_1, \ldots, X_n\right]\right] \\
&\leq \mathbb{E}\left[M\left(\sqrt{2n^{-1}\log N(\mathcal{F}_M, L_1(P_n), \epsilon)} \wedge 1\right)\right] = \mathbb{E}[o_p(1) \wedge M] = o(1),
\end{aligned}$$

where we have used Lebesgue dominated convergence theorem to claim that $\mathbb{E}[o_p(1) \wedge M] = o(1)$. By letting $M \to +\infty$, $n \to +\infty$ and $\epsilon \downarrow 0$, $A + B$ will go to zero. Thus $\mathbb{E}[\|P_n - P\|_{\mathcal{F}}] \to 0$ as $n \to \infty$. $\square$