STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Mar 8

Lecturer: Xianyang Zhang

1 Uniform laws of large numbers

1.1 Definition

Recall that for $X_1, \ldots, X_n \sim^{i.i.d} P$, we let $P_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$. Define $Pf = \int f dP$ and $P_n f = n^{-1}\sum_{i=1}^n f(X_i)$.

Let $\mathcal{F}$ be a collection of functions $f : \mathcal{X} \to \mathbb{R}$. Then $\mathcal{F}$ satisfies the uniform law of large numbers (ULLN) if

$$||P_n - P||_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |P_n f - Pf| \longrightarrow^p 0.$$

1.2 Glivenko-Cantelli

Consider $\mathcal{F} = \{f(x) = \mathbf{1}\{x \le t\} : t \in \mathbb{R}\}$. As shown in STAT 614,

$$\sup_{f\in\mathcal{F}} |P_n f - Pf| = \sup_{t\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i \le t\} - P(X \le t) \right| \longrightarrow^p 0,$$

where $X \sim P$.

1.3 Metric entropy

Let (Θ, ρ) be a metric space where $\rho : \Theta \times \Theta \longrightarrow [0, +\infty)$. For $\epsilon > 0$, we say $\{\theta^i\}_{i=1}^N$ is an ϵ-covering of Θ if for any $\theta \in \Theta$, there exists an i such that

$$\rho(\theta, \theta^i) \le \epsilon.$$

The ϵ-covering number of Θ is the smallest size of ϵ-covers, that is

$$N(\Theta, \rho, \epsilon) = \inf \left\{ N \in Z_+ : \text{ there exists an } \epsilon\text{-covering } \{\theta^i\}_{i=1}^N \text{ of } \Theta \right\}.$$

The metric entropy is defined as $H(\Theta, \rho, \epsilon) = \log N(\Theta, \rho, \epsilon)$.

1.4 Packing number

For $\delta > 0$, a set $\{\theta^i\}_{i=1}^N$ is a δ-packing of Θ if for any $i \ne j$, $\rho(\theta^i, \theta^j) > \delta$. The δ-packing number is

$$M(\Theta, \rho, \delta) = \sup\{M \in Z_+ : \text{ there exists an } \delta\text{-packing } \{\theta_i\}_{i=1}^M \text{ of } \Theta\}.$$

A useful observation:

$$M(\Theta, \rho, 2\epsilon) \le N(\Theta, \rho, \epsilon) \le M(\Theta, \rho, \epsilon).$$

1.5 Volumn comparison lemma

Let $\Theta = \{\theta \in \mathbb{R}^d : \|\theta\| \le r\}$. Consider the Euclidean distance, that is $\rho(x, y) = \|x - y\|$. We have

$$\left(\frac{r}{\epsilon}\right)^d \le N(\Theta, \rho, \epsilon) \le \left(1 + \frac{2r}{\epsilon}\right)^d$$

and thus

$$d \log \frac{r}{\epsilon} \le \log N(\Theta, \rho, \epsilon) \le d \log \left(1 + \frac{2r}{\epsilon}\right).$$

Proof. Let $B = \left\{\theta \in R^d, ||\theta|| \le 1\right\}$. Then $\Theta = rB$. We note that

$$\mathrm{Vol}(rB) = \frac{\pi^{d/2}}{\Gamma(d/2+1)} r^d.$$

Thus

$$\frac{\mathrm{Vol}(\Theta)}{\mathrm{Vol}(\epsilon B)} = \left(\frac{r}{\epsilon}\right)^d$$

which implies that $N(\Theta, \rho, \epsilon) \ge \left(\frac{r}{\epsilon}\right)^d$. Let $\{\theta^i\}_{i=1}^M$ be the maximum ϵ-packing of $\Theta = rB$. Then $\{\theta^i + \epsilon B/2\}_{i=1}^M$ are disjoint and

$$\cup_{i=1}^M (\theta^i + \epsilon B/2) \subset (r + \epsilon/2)B.$$

By comparing the volumes of both sides, we obtain

$$M \le \left(1 + \frac{2r}{\epsilon}\right)^d.$$

This result follows in view of $N(\Theta, \rho, \epsilon) \le M(\Theta, \rho, \epsilon)$.

1.6 Bracketing

Let $\mathcal{F}$ be a collection of funcitons, and μ be a measure on $\mathcal{X}$. A set $\{[l_i, u_i]\}_{i=1}^M$ is an ϵ-bracket of $\mathcal{F}$ in $L_p(\mu)$ if $\forall f \in \mathcal{F}$, there exists an $1 \le i \le M$ such that $l_i(x) \le f(x) \le u_i(x), \forall x \in \mathcal{X}$ and $\|l_i - u_i\|_{L_p(\mu)} \le \epsilon$. The bracketing number of Θ defined as

$$N_{[\ ]}(\mathcal{F}, L_p(\mu), \epsilon) = \inf \left\{M \in Z_+ : \text{ there exists an } \epsilon\text{-bracket } \{[l_i, u_i]\}_{i=1}^M \text{ of } \mathcal{F}\right\}.$$

1.7 Example: Lipschitz functions

Let $\Theta \subset \mathbb{R}^d$ be a compact subset. Let $\mathcal{F} = \{l_\theta(x) : \theta \in \Theta\}$ where l_θ is $L(x)$-Lipschitz in Θ, namely, for all $x \in \mathcal{X}$, $\theta_1, \theta_2 \in \Theta$,

$$|l_{\theta_1}(x) - l_{\theta_2}(x)| \leqslant L(x)\|\theta_1 - \theta_2\|.$$

Suppose that random variable X has a distribution P, which induces a measure μ. Assume that $\mathbb{E}[L(X)] < \infty$. Then

$$N_{[\ ]}(\mathcal{F}, L_1(\mu), \epsilon\mathbb{E}[L(X)]) \le N\left(\Theta, \|\cdot\|, \frac{\epsilon}{2}\right).$$

Proof. Let $\{\theta_i\}_{i=1}^N$ be an $\frac{\epsilon}{2}$-covering of Θ. For all $\theta \in \Theta$, there exists $1 \le i \le N$, such that $\|\theta - \theta_i\| \le \frac{\epsilon}{2}$. Then,

$$|l_\theta(x) - l_{\theta_i}(x)| \le L(x)\|\theta - \theta_i\| \le \frac{\epsilon}{2} L(x).$$

Define $u_i(x) = l_{\theta_i}(x) + \frac{\epsilon}{2}L(x)$ and $l_i(x) = l_{\theta_i}(x) - \frac{\epsilon}{2}L(x)$. Thus,

$$l_i(x) \le l_\theta(x) \le u_i(x),$$

for all $x \in \mathcal{X}$ and

$$\mathbb{E}[|u_i(X) - l_i(X)|] \le \epsilon\mathbb{E}[L(X)].$$

1.8 Example: Increasing functions

Let $\mathcal{F}$ be a class of increasing functions with form $f : \mathbb{R} \to [0, 1]$. Let $\mathcal{X} \subset \mathbb{R}$ be a finite set with cardinality n. Define $\|f\|_\infty = \sup_{x\in\mathcal{X}} |f(x)|$. Let $N_\infty(\mathcal{F}, \epsilon) = N(\mathcal{F}, L_\infty(\mathcal{X}), \epsilon)$. Thus, $N_\infty(\mathcal{F}, \epsilon)$ is the smallest value of N such that there exists $\{f_j\}_{j=1}^N$ with $\sup_{f\in\mathcal{F}} \min_{1\le j\le N} \|f - f_j\|_\infty < \epsilon$. Then

$$N_\infty(\mathcal{F}, \epsilon) \le \left(n + \frac{1}{\epsilon}\right)^{\frac{1}{\epsilon}}.$$

1.8.1 Proof

Let $\mathcal{X}$ consist of points such that $x_1 \le x_2 \le \ldots \le x_n$. For $f \in \mathcal{F}$, define $M_i = \left\lfloor \frac{f(x_i)}{\epsilon} \right\rfloor$ for $1 \le i \le n$. Further let $\tilde{f}(x_i) = \epsilon M_i$. Notice that

$$\epsilon\left(\frac{f(x_i)}{\epsilon} - 1\right) \le \tilde{f}(x_i) \le \epsilon\left(\frac{f(x_i)}{\epsilon} + 1\right).$$

So $\|\tilde{f} - f\|_\infty \le \epsilon$. Since f is increasing, $0 \le M_1 \le \ldots \le M_n \le \lfloor\frac{1}{\epsilon}\rfloor$. We claim that

$$N_\infty(\mathcal{F}, \epsilon) \le \binom{\lfloor\frac{1}{\epsilon}\rfloor + n}{n}.$$

Recall that

$$\binom{n}{k} \le (\frac{en}{k})^k \le n^k.$$

Therefore,

$$\binom{\lfloor\frac{1}{\epsilon}\rfloor + n}{n} = \binom{\lfloor\frac{1}{\epsilon}\rfloor + n}{\lfloor\frac{1}{\epsilon}\rfloor} \le \left(n + \lfloor\frac{1}{\epsilon}\rfloor\right)^{\lfloor\frac{1}{\epsilon}\rfloor} \le \left(n + \frac{1}{\epsilon}\right)^{\frac{1}{\epsilon}}.$$

1.9 Theorem

Suppose $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$ and $N_{[\,]}(\mathcal{F}, L_1(\mu), \epsilon) < \infty$ for all $\epsilon > 0$. Then

$$\sup_{f\in\mathcal{F}} |P_n f - Pf| \xrightarrow{P} 0. \quad \text{ULLN}$$

1.9.1 Proof

For any $\epsilon > 0$, let $\{[l_i, u_i]\}_{i=1}^N$ be an ϵ-bracketing for $\mathcal{F}$. For any $f \in \mathcal{F}$, there exists i such that $l_i \le f \le u_i$. Then

$$P_n f - Pf \le P_n u_i - Pl_i \le (P_n - P)u_i + P(u_i - l_i) \le (P_n - P)u_i + \epsilon.$$

Similarly, $Pf - P_n f \le (P - P_n)l_i + \epsilon$. Then, for any $f \in \mathcal{F}$,

$$|P_n f - Pf| \le \max_{1\le i\le N} (P - P_n)l_i \vee (P_n - P)u_i + \epsilon.$$

Thus, for all $\epsilon > 0$,

$$\sup_{f\in\mathcal{F}} |P_n f - Pf| \le o_p(1) + \epsilon.$$

As ϵ is arbitrary, the result follows.

1.10 Example: Logistic regression

Suppose that we are given a pair $Z = (X, Y) \in \mathbb{R} \times \{\pm 1\}$, which is the setting of logistic regression. Set

$$Y = \begin{cases} 1, & p = \frac{1}{1+e^{-X^\top \beta}} \\ -1, & 1-p = \frac{1}{1+e^{X^\top \beta}}. \end{cases}$$

Define

$$l_\beta(Z) = l_\beta(X, Y) = \log(1 + e^{-\mathrm{sign}(Y) X^\top \beta}), \quad \beta \in \mathcal{B}.$$

We claim that $l_\beta(Z)$ is $\|X\|$-Lipschitz. In fact, consider the function $f(t) = \log(1 + e^{-t})$. Then $f'(t) = \frac{-e^{-t}}{1+e^{-t}}$, which means $|f'(t)| \le 1$. So

$$|l_{\beta_1}(Z) - l_{\beta_2}(Z)| \le |X^\top \beta_1 - X^\top \beta_2| \le \|X\| \|\beta_1 - \beta_2\|,$$

where $\beta_1, \beta_2 \in \mathcal{B}$. If $\mathbb{E}[\|X\|] < \infty$ and $\mathcal{B}$ is compact, we can apply the first theorem and then conclude that the collection of functions

$$\mathcal{F} = \{l_\beta(z) : \beta \in \mathcal{B}\}$$

satisfies the ULLN.