Stat210B: Theoretical Statistics **Lecture Date: February 6th, 2007**

Uniformly Strong Law of Large Numbers

Lecturer: Michael I. Jordan *Scribe: Jian Ding*

In this lecture, we try to generalize the Glivenko-Cantelli theorem.

Let $\xi_1, \xi_2, \cdots, \xi_n \sim P$ and are i.i.d. sequences. We define $Pf := \mathbb{E}(f(X))$, in which $X \sim P$. We also define $P_n f$ with respect to the empirical measure but puts mass $\frac{1}{n}$ at $\{\xi_1, \cdots, \xi_n\}$. Notice that by definition $P_n f = \frac{1}{n}\sum_{i=1}^n f(\xi_i)$.

We point out that $P_n f - Pf$ is an object of interest; and $\sup_{f\in\mathcal{F}} |P_n f - Pf|$ is of even more interest. For example, let $\mathcal{F} = \{\mathbb{1}_{(-\infty,t]} : t \in \mathbb{R}\}$, then $P_n f - Pf$ becomes $F_n(t) - F(t)$ and $\sup_{f\in\mathcal{F}} |\cdot|$ becomes $\sup_t |F_n(t) - F(t)|$. In general, we are interested in statistics defined on a family of stochastic processes with index set $\mathcal{F}$.

Uniform Law of Large Numbers

Define $\|P_n - P\| := \sup_{f\in\mathcal{F}} |P_n f - Pf|$. Recalling the discussion in last lecture, we get

$$\begin{aligned}\mathbb{P}\{\|P_n - P\| > \epsilon\} &\leq 2\mathbb{P}\{\|P_n - P_n'\| > \frac{\epsilon}{2}\}\\ &\leq 4\mathbb{P}\{\|P_n^0\| > \frac{\epsilon}{4}\}\end{aligned}$$

where P_n^0 is a signed measure putting mass $\frac{1}{n}\sigma_i$ at $\{\xi_1, \cdots, \xi_n\}$. Again, σ_i independently pick value uniformly on $\{1, -1\}$.

Specialize $\mathcal{F}$ to indicators

Let $I_j = (-\infty, t_j]$ where $\{t_j\}$ lie between the points ξ_i, i.e., $t_0 < \xi_1 < t_1 < \xi_2 < t_2 < \cdots$. Consider

$$\begin{aligned}&\mathbb{P}\{\|P_n^0\| > \frac{\epsilon}{4}|\xi\}\\ &= \mathbb{P}\{\bigcup_{j=0}^n \{|P_n^0 I_j| > \frac{\epsilon}{4}\}|\xi\}\\ &\leq \sum_{j=0}^n \mathbb{P}\{|P_n^0 I_j| > \frac{\epsilon}{4}|\xi\}\\ &\leq (n+1)\max_j \mathbb{P}\{|P_n^0 I_j| > \frac{\epsilon}{4}|\xi\}.\end{aligned}$$

Recall Hoeffding's inequality. Let Y_i be independent, $\mathbb{E}(Y_i) = 0$, $a_i \leq Y_i \leq b_i$. Then, $\mathbb{P}\{|Y_1 + Y_2 + \cdots + Y_n| > \eta\} \leq \exp\{-\frac{2\eta^2}{\sum_i (b_i - a_i)}\}$. We apply this to $\sigma_i\{\xi_i \leq t\}$, and conclude

$$\begin{aligned}\mathbb{P}\{|P_n^0\{(-\infty, t]\}| > \frac{\epsilon}{4}|\xi\} &\leq 2\exp(-\frac{2(n\epsilon/4)^2}{4n})\\ &\leq 2\exp(-\frac{n\epsilon^2}{32}),\end{aligned}$$

notice that this is independent of ξ, so $\mathbb{P}\{\|P_n - P\| > \epsilon\} \leq 8(n+1)\exp(-\frac{n\epsilon^2}{32})$, i.e., we get Uniform Law of Large Numbers in probability and also almost surely (by Borel-Cantelli).

The conclusion, namely, Glivenko-Cantelli theorem is not new. However, this method can be generalized to richer class of functions immediately.

VC Classes

Consider a collection $\mathcal{C}$ of subsets of some set $\mathcal{X}$, and consider points $\xi_1, \cdots, \xi_n$ from $\mathcal{X}$. Define $\Delta_n^{\mathcal{C}} := \#\{C \bigcap \{\xi_1, \cdots, \xi_n\} : C \in \mathcal{C}\}$; $m(n) := \max_{\xi_1, \cdots, \xi_n} \Delta_n^{\mathcal{C}}(\xi_1, \cdots, \xi_n)$; $V^{\mathcal{C}} := \min\{n : m(n) < 2^n\}$.

Examples

1, $\mathcal{X} = \mathbb{R}, \mathcal{C} = \{(-\infty, t]\}$. Then, $V^{\mathcal{C}} = 2$.

2, $\mathcal{X} = \mathbb{R}, \mathcal{C} = \{(s, t] : s < t\}$. Then, $V^{\mathcal{C}} = 3$.

3, $\mathcal{X} = \mathbb{R}^d, \mathcal{C} = \{(-\infty, t] : t \in \mathbb{R}^d\}$. Then, $V^{\mathcal{C}} = d + 1$.

4, Rectangles in $\mathbb{R}^d$. $V^{\mathcal{C}} = 2d + 1$.

Sauer's Lemma

Lemma 1.

$$m(n) \leq \sum_{j=0}^{V^{\mathcal{C}}} \binom{n}{j} \leq (\frac{ne}{V^{\mathcal{C}} - 1})^{V^{\mathcal{C}} - 1}.$$

Proof. We prove the second part.

$$\begin{aligned}
\sum_{j=0}^{S} \binom{n}{j} &= 2^n \sum_{j=0}^{S} \binom{n}{j} (\frac{1}{2})^n \\
&= 2^n \mathbb{P}(Y \leq S), \quad Y \sim Bin(n, \frac{1}{2}) \\
&\leq 2^n \mathbb{E}(\theta)^{Y-S}, \quad 0 \leq \theta \leq 1 \\
&= 2^n \theta^{-S} (\frac{1}{2} + \frac{\theta}{2})^n, \quad \text{take } \theta = \frac{S}{n} \\
&= (\frac{n}{S})^S (1 + \frac{S}{n})^n \\
&\leq (\frac{n}{S})^S e^S.
\end{aligned}$$

□

This suggests

$$\begin{aligned}
\mathbb{P}\{\|P_n^0\| > \frac{\epsilon}{4} | \xi\} &= \mathbb{P}\{\bigcup_{i=0}^{m(n)} |P_n^0 \widetilde{f}_i| > \frac{\epsilon}{4} | \xi\} \\
& \quad (\widetilde{f}_i \text{ are indicators of subsets that achieve } m(n)). \\
&\leq \sum_{i=1}^{m(n)} \mathbb{P}\{|P_n^0 \widetilde{f}_i| > \frac{\epsilon}{4} | \xi\} \\
&\leq m_n \max(\cdot).
\end{aligned}$$

Then if $\mathcal{F}$ is a VC class, (i.e., $V^{\mathcal{C}} < \infty$), then

$$\mathbb{P}\{\|P_n - P\| > \epsilon\} \leq \text{ (Poly in } n)(\exp(-Cn)).$$