Lecture 7

II. Method 2: Covering numbers and chaining

Proceeding as before, we get

$$\mathbb{E}\left[\Delta_n(\mathcal{A})\right] \le \frac{2}{n}\,\mathbb{E}\left[\sup_{A\in\mathcal{A}}\left|\sum_{i=1}^{n}\sigma_i \mathbb{1}_{\{x_i\in A\}}\right|\right]$$

Notation:

$$\mathcal{A}(x_1,\dots,x_n) = \left\{\left(\mathbb{1}_{\{x_1\in A\}},\dots,\mathbb{1}_{\{x_n\in A\}}\right), A\in\mathcal{A}\right\}.$$

Consider

$$\mathbb{E}\left[\sup_{A\in\mathcal{A}}\left|\sum_{i=1}^{n}\sigma_i \mathbb{1}_{\{x_i\in A\}}\right|\right]$$

$$= \mathbb{E}\left[\max_{b\in\mathcal{A}(x)}\left|\sum_{i=1}^{n}\sigma_i b_i\right|\right]$$

Let b^* attain the max. Thus, we are interested in $\mathbb{E}\left[\left|\sum_{i=1}^{n}\sigma_i b_i^*\right|\right]$.

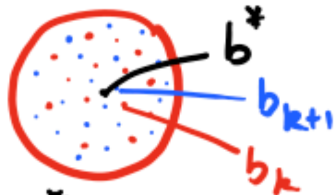
Definition (covering)

Given a metric space (M,ρ) and $A\subseteq M$, a set C constitutes a covering of radius r for A if $\forall\, a\in A\ \exists\, c\in C$ s.t. $\rho(a,c)\le r$.

Denote by $N(r,A)$ the min. cardinality of a covering of radius r for A.

We consider the n-dimensional binary hypercube with $\rho(x,y) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i \neq y_i\}}$.

Let $B_1, \dots, B_M$ be s.t. that B_k is a min. size covering of $A(x)$ of radius 2^{-k}.



Let $b^{(k)}$ denote the closest neighbor of b^* in B_k, with $b^{(0)} = 0$ and $b^{(M+1)} = b^*$

$$b^* = \sum_{k=1}^{M+1} b^{(k)} - b^{(k-1)}$$

Note that it suffices to consider $M = \log \frac{1}{\sqrt{n}}$

Also, $\rho\left(b^{(k)}, b^{(k+1)}\right)$

$$\leq \rho\left(b^*, b^{(k)}\right) + \rho\left(b^*, b^{(k+1)}\right)$$

$$\leq 2^{-k} + 2^{-k+1} = 3 \cdot 2^{-k}$$

Thus, $\mathbb{E}\left[\left|\sum_{i=1}^{n} \sigma_i b_i^*\right|\right]$

$$\leq \sum_{k=1}^{M+1} \mathbb{E}\left[\left|\sum_{i=1}^{n} \sigma_i \left(b_i^{(k)} - b_i^{(k-1)}\right)\right|\right]$$

$$\leq \sum_{k=1}^{M+1} \mathbb{E}\left[\sup_{\substack{b \in B_k,\, c \in B_{k-1}, \\ \rho(b,c) \leq 3 \cdot 2^{-k}}} \left|\sum_{i=1}^{n} \sigma_i (b_i - c_i)\right|\right] \quad (4)$$

$$\sum_{i=1}^{n} \sigma_i (b_i - c_i) \in \mathcal{G}\left(n \rho(b,c)^2\right)$$

Also, $|B_k| \geq |B_{k-1}|$.

Thus, the right-side of (4) is bounded above by ③

$$\sum_{k=1}^{M+1} 3 \cdot 2^{-k} \cdot \sqrt{2n \log 2 \, |B_k|^2}$$

$$\leq 12\sqrt{n} \sum_{k=1}^{M+1} 2^{-k+1} \sqrt{\log 2 N(2^{-k}, A(x))}$$

$$\leq 12\sqrt{n} \sum_{k=1}^{\infty} 2^{-k+1} \sqrt{\log 2 N(2^{-k}, A(x))}$$

$$\leq 12\sqrt{n} \int_0^1 \sqrt{\log 2 N(r, A(x))} \, dr .$$

Thus,
$$\boxed{\mathbb{E}\left[\Delta_n(A)\right] \leq \frac{24}{\sqrt{n}} \int_0^1 \sqrt{\log 2 N(r, A(x))} \, dr}$$

__Example__

For $A = \{(-\infty, \theta), \theta \in \mathbb{R}\}$, $A(x)$ consists of sequences of the form $(0, \dots, 0)$, $(1, 0, \dots, 0)$, $\dots, (1, 1, \dots, 1)$, where we assume $x_1 < x_2 < \dots < x_n$.

For a covering of radius r, we can ignore nr^2 1's. Thus, we can consider sequences in $A(x)$ with the number of ones which are multiples of $\lfloor nr^2 \rfloor$; there are at most $\left(\frac{1}{r^2} + 1\right)$ such seq.

Thus, $N(r, A(x)) \le \frac{1}{r^2} + 1$. Therefore,

$$\mathbb{E}[\Delta_n(A)] \le \frac{24}{\sqrt{n}} \int_0^1 \sqrt{\log 2\left(\frac{1}{r^2}+1\right)}\, dr$$

$$= \frac{24}{\sqrt{n}} \int_0^1 \sqrt{\log \frac{4}{r^2}}\; dr$$

$$= \frac{24}{\sqrt{n}} \int_0^\infty \sqrt{2u}\; e^{-u} \,.\, 2 \,.\, du$$

$$= \frac{24 \,.\, 2 \,.\, \sqrt{2}}{\sqrt{n}} \,.\, \sqrt{\frac{\pi}{4}} = \frac{24\sqrt{2\pi}}{\sqrt{n}} \,.$$

(Massart 90: The constant can be improved to 1.)

—— × —— × —— × ——

This completes our digression and discussion on applications.

Lessons:
- BDP $\Rightarrow$ conc. makes handling deviations easy
- Handling expected value of max. also uses sub-Gaussian behaviour
- The # of elements we take max. over can be controlled by methods developed in empirical process theory.

Agenda for the remaining class

→ Effective "variance" of a function of independent r.v.s: The Efron-Stein Inequality

[A] Efron-Stein Inequality

→ Sum of uncorrelated zero mean r.v.s conc. around zero

→ $Z_i \stackrel{\text{def}}{=} \mathbb{E}[f(X) \mid X^i]$, $\Delta_i = Z_i - Z_{i-1}$

Δ_i's are uncorrelated (mult. family)

$\Rightarrow$ $f(X)$ conc. around $\mathbb{E}[f(X)]$ with "variance parameter" $\sum_{i=1}^{n} c_i^2$.

Now, we shall focus only on the variance of f and get a bound for it; this is a sort of precursor to Azuma-Hoeffding-McDiarmid.

$$Z - \mathbb{E}[Z] = \sum_{i=1}^{n} \Delta_i$$

Thus,

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\sum_{i=1}^{n} \Delta_i\right)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(\Delta_i) \qquad [\Delta_i\text{'s are uncorr.}]$$

$$(1) \qquad = \sum_{i=1}^{n} \mathbb{E}[\Delta_i^2] \qquad [\Delta_i\text{'s are zero-mean}]$$

Observation to simplify Δ_i

Let a subscript i denote conditioning on X^i and a subscript (i) denote conditioning on $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

Claim: $\mathbb{E}_i\left[\mathbb{E}_{(i)}[Z]\right] = \mathbb{E}_{i-1}[Z]$

With this simplification, we have

$$\Delta_i = \mathbb{E}_i[Z] - \mathbb{E}_{i-1}[Z]$$
$$= \mathbb{E}_i\left[Z - \mathbb{E}_{(i)}[Z]\right]$$
$$\Rightarrow \Delta_i^2 \le \mathbb{E}_i\left[\left(Z - \mathbb{E}_{(i)}[Z]\right)^2\right] \quad \text{a.s.}$$

(by conditional Jensen's)

Therefore, by (1), we get

$$\mathrm{Var}(Z) \le \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}_i\left[\left(Z - \mathbb{E}_{(i)}[Z]\right)^2\right]\right]$$
$$= \sum_{i=1}^{n} \mathbb{E}\left[\underbrace{\left(Z - \mathbb{E}_{(i)}(Z)\right)^2}_{\mathrm{Var}_{(i)}(Z)}\right]$$

Efron-Stein Inequality Form 1:

$$\boxed{\mathrm{Var}(Z) \le \sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}_{(i)}(Z)\right]}$$