Mathematical Foundations of Infinite-Dimensional Statistical Models

Chapter 3.5 - 3.5.1

presented by Boyoung Kim

Seoul National University

December 21, 2018

Table of Contents

3.5 Metric Entropy Bounds for Suprema of Empirical Processes

- Good estimates for the mean of the supremum of an empirical process $E\|P_n - P\|_{\mathcal{F}}$.
- This section and the next are devoted to this important subject.

3.5.1 Random Entropy Bounds via Randomisation

Random (pseudo)distance

- For any $n \in \mathbb{N}$, let P_n denote the empirical measure corresponding to n i.i.d. S-valued random variables X_i of law P.
- Then, for any measurable real functions f, g on S, we let $e_{n,2}(f, g)$ denote their $L^2(P_n)$ (pseudo)distance, that is,

$$e_{n,2}^2(f, g) = \frac{1}{n} \sum_{i=1}^{n} (f - g)^2 (X_i)$$

Random or empirical metric entropies

- Given a class of measurable functions $\mathcal{F}$ on S, the empirical metric entropies of $\mathcal{F}$ are defined as $logN(\mathcal{F}, e_{n,2}, \tau)$ for any $\tau > 0$ (recall from Section 2.3).
- Often we will write $N(\mathcal{F}, L^2(P_n), \tau)$ for $N(\mathcal{F}, e_{n,2}, \tau)$.
- Recall also the packing numbers $D(T, d, \tau)$ and their relationship with covering numbers: for all $\tau > 0$

$$N(T, d, \tau) \leq D(T, d, \tau) \leq N(T, d, \tau/2) \qquad (3.164)$$

Theorem 3.5.1

In the preceding notation, assuming $\mathcal{F}$ countable and $0 \in \mathcal{F}$

$$E[\sqrt{n}\|P_n - P\|_{\mathcal{F}}] \leq 8\sqrt{2}E\left[\int_0^{\sqrt{\|P_n f^2\|_{\mathcal{F}}}} \sqrt{log 2D(\mathcal{F}, L^2(P_n), \tau)}d\tau\right] \quad (3.165)$$

and, for all $\delta > 0$,

$$E\left[\sqrt{n} \sup_{f,g\in\mathcal{F}:P_n|f-g|^2\leq\delta^2} |(P_n - P)(f-g)|\right]$$
$$\leq 2(16\sqrt{2}+2)E\left[\int_0^{\delta} \sqrt{log 2D(\mathcal{F}, L^2(P_n), \tau)}d\tau\right]. \quad (3.166)$$

Theorem 3.5.1

Key idea of the proof

- By Theorem 3.1.21, we can randomise the empirical process by Rademacher multipliers.
- The resulting process is sub-Gaussian conditionally on the variables X_i, and therefore, the metric entropy bounds in Section 2.3, in particular, Theorem 2.3.7, apply to it.

Measurable envelope (or envelope)

- If a measurable function F satisfies $|f| \leq F$, for all $f \in \mathcal{F}$, we say that F is a measurable envelope (or envelope) of the class of functions $\mathcal{F}$.

- As we see in the next section, there are many classes of functions $\mathcal{F}$, denoted by $Vapnik - Cervonenkis$ classes of functions, whose covering numbers admit the bound

$$N(\mathcal{F}, L^2(Q), \tau \|F\|_{L^2(Q)}) \leq \left(\frac{A}{\tau}\right)^{\nu}, \quad 0 < \tau \leq 1, \tag{3.168}$$

  for some A, ν positive and finite and for all probability measures Q on $(S, \mathcal{S})$, where F is a measurable envelop of $\mathcal{F}$.

- The next theorem will cover in particular the $Vapnik - Cervonenkis$ case.

Koltchinskii-Pollard entropy

- For ease of notation, we set, for all $0 < \delta < \infty$,

$$J(\mathcal{F}, F, \delta) := \int_0^\delta \sup_Q \sqrt{N(\mathcal{F}, L^2(Q), \tau\|F\|_{L^2(Q)})} d\tau, \qquad (3.169)$$

  where the supremum is taken over all discrete probabilities with a finite number of atoms and rational weights.

- The integrand of J is denoted as the Koltchinskii-Pollard entropy of $\mathcal{F}$.

Theorem 3.5.4

Theorem 3.5.4

Let $\mathcal{F}$ be a countable class of measurable functions with $0 \in \mathcal{F}$, and let F be a strictly positive envelope for $\mathcal{F}$. Assume that

$$J(\mathcal{F}, F, \delta) < \infty, \text{ for some (for all) } \delta > 0 \tag{3.170}$$

where J is defined in (3.169). Given $X_1, ..., X_n$ iid S-valued rvs. with common law P such that $PF^2 < \infty$, let P_n be the corresponding empirical measure and $\nu_n(f) = \sqrt{n}(P_n - P)(f), f \in F$. Set $U = max_{1 \le i \le n} F(X_i), \sigma^2 = \sup_{f \in \mathcal{F}} Pf^2$ and $\delta = \sigma / \|F\|_{L^2(P)}$. Then, for all $n \in \mathbb{N}$,

$$E\|\nu_n\|_{\mathcal{F}} \le \max \left[A_1 \|F\|_{L^2(P)} J(\mathcal{F}, F, \delta), \frac{A_2 \|U\|_{L^2(P)} J^2(\mathcal{F}, F, \delta)}{\sqrt{n}\delta^2} \right], \tag{3.171}$$

where we can take

$$A_1 = 8\sqrt{6} \text{ and } A_2 = 2^{15} 3^{5/2}. \tag{3.172}$$

- The next theorem covers when the Koltchinskii-Pollard entropy admits as upper bound a regularly varying function.
- The resulting bound for the expected value of the empirical process becomes particularly simple and applies in many situations including the $Vapnik - Cervonenkis$ case (3.168)

Theorem 3.5.6

Theorem 3.5.6

Let $\mathcal{F}$ be a countable class of functions with $0 \in \mathcal{F}$, let F be an envelope for $\mathcal{F}$ and let $H : [0,\infty) \to [0,\infty)$ be a function equal to log2 for $0 < x \leq 1$ and such that
(a) $H(x)$ is nondecreasing for $x > 0$, and so is $xH^{1/2}(1/x)$ for $0 < x \leq 1$, and
(b) there exists C_H finite such that $\int_0^c \sqrt{H(1/x)}dx \leq C_H c H^{1/2}(1/c)$ for all $0 < c \leq 1$. Assume that

$$\sup_Q log[2N(\mathcal{F}, L^2(Q), \tau\|F\|_{L^2(Q)})] \leq H(\frac{1}{\tau}), \text{ for all } \tau > 0, \tag{3.178}$$

where the supremum is taken over all discrete probability measures Q with a finite number of atoms and with rational weights. Then

$$E\|\nu_n\|_{\mathcal{F}} \leq \max\left[A_1 C_H \sigma\sqrt{H(\|F\|_{L^2(P)}/\sigma)}, A_2 C_H^2 \|U\|_{L^2(P)} H(\|F\|_{L^2(P)}/\sigma)/\sqrt{n}\right], \tag{3.179}$$

where A_1 and A_2 are the constants in (3.172).

Corollary 3.5.7

The uniformly bounded case in the preceding two theorems

Corollary 3.5.7

Assume that the hypotheses of Theorem 3.5.6 are satisfied and that, moreover, the functions in $\mathcal{F}$ are bounded in absolute value by a constant u. Then

$$E\|\nu_n\|_{\mathcal{F}} \leq 8\sqrt{2}C_H\sigma\sqrt{H(\|F\|_{L^2(P)}/\sigma)} + 2^7 C_H^2 uH(\|F\|_{L^2(P)}/\sigma)/\sqrt{n}. \quad (3.181)$$

Corollary 3.5.8

Corollary 3.5.8

Suppose that $sup_Q N(\mathcal{F}, L^2(Q), \epsilon\|F\|_{L^2(Q)}) \leq (A/\epsilon)^\nu$, *for* $0 < \epsilon < A$, *for some* $\nu \geq 1$ *and* $A \geq 2$, *the supremum extending over all Borel probability measures* Q, *and let* $u = \|F\|_\infty$. *Then*

$$E\|\nu_n\|_{\mathcal{F}} \leq 8\sqrt{2}C_A\sigma\sqrt{2\nu log\frac{A\|F\|_{L^2(Q)}}{\sigma}} + 2^8 C_A\frac{1}{\sqrt{n}}u\nu log\frac{A\|F\|_{L^2(Q)}}{\sigma}, \quad (3.184)$$

where $C_A = 2logA/(2logA - 1)$.

- Perhaps the main observation regarding Theorem 3.5.6 is that if

$$n\sigma^2/\|U\|^2_{L^2(P)} \gtrsim H(\|U\|_{L^2(P)}/\sigma),$$

then the bound (3.179) becomes, disregarding constants,

$$E\Big\|\sum_{i=1}^{n}(f(X_i) - Pf)\Big\|_{\mathcal{F}} \lesssim \sqrt{n\sigma^2 H\left(\frac{2\|F\|_{L^2(P)}}{\sigma}\right)}.$$

Corollary 3.5.9

Corollary 3.5.9

Under the hypotheses of Theorem 3.5.6 and with the same notation, if, moreover, for some $\lambda \geq 1$,

$$\frac{n\sigma^2}{\|U\|^2_{L^2(P)}} \geq \left(\frac{A_2 C_H}{\lambda A_1}\right)^2 H(\|F\|_{L^2(P)}/\sigma), \tag{3.185}$$

then

$$E\left\|\sum_{i=1}^{n}(f(X_i) - Pf)\right\|_{\mathcal{F}} \leq \lambda A_1 C_H \sqrt{n\sigma^2 H\left(\frac{2\|F\|_{L^2(P)}}{\sigma}\right)} \leq \frac{\lambda^2 A_1^2}{A_2} \frac{n\sigma^2}{\|U\|_{L^2(P)}} \tag{3.186}$$

where A_1 *and* A_2 *are defined in (3.172). In the uniformly bounded case, if*

$$\frac{n\sigma^2}{u} \geq \frac{2^7 C_H^2}{(\lambda - 1)^2} H(\|F\|_{L^2(P)}/\sigma),$$

then

$$E\left\|\sum_{i=1}^{n}(f(X_i) - Pf)\right\|_{\mathcal{F}} \leq 8\sqrt{2}\lambda C_H \sqrt{n\sigma^2 H\left(\frac{2\|F\|_{L^2(P)}}{\sigma}\right)} \leq \lambda^2 \frac{n\sigma^2}{u}.$$

Definition 3.5.10

We need a definition just to describe how the function H must also be, up to constants, a lower bound for the metric entropy of $\mathcal{F}$.

Definition 3.5.10

A class of functions $\mathcal{F}$ that satisfies the hypotheses of Theorem 3.5.6 and such that $|f| \leq 1$ for all $f \in \mathcal{F}$ is full for H and P if, moreover, there exists $c > 0$ such that

$$logN(\mathcal{F}, L^2(P), \sigma/2) \geq cH\left(\frac{\|F\|_{L^2(P)}}{\sigma}\right), \qquad (3.187)$$

for a measurable envelope F of $\mathcal{F}$.

Theorem 3.5.11

Theorem 3.5.11

Let $\mathcal{F}$, H and F be as in Theorem 3.5.6 but further assume that the functions in $\mathcal{F}$ take values in $[-1,1]$, let $P_n, n \in \mathbb{N}$, be the empirical measure corresponding to samples from a probability measure P on $(S, \mathcal{S})$ and suppose as well that $Pf = 0$ for all $f \in \mathcal{F}$. Set $D_H = \int_0^1 \sqrt{H(1/\tau)}d\tau$. Assume that

$$n\sigma^2 \geq (2^{15} \vee 2^{22} K^2 C_H^2)\, H(6\|F\|_2/\sigma) \text{ and } n\sigma^2 \geq 32\sqrt{2}D_H/(3e^{1/2}), \tag{3.188}$$

where $K \geq 1$ is as in Theorem 3.2.9. Then

$$E\Big\|\sum_{i=1}^{n} f(X_i)\Big\|_{\mathcal{F}} \geq \frac{\sqrt{n}\sigma}{32K}\sqrt{logN(\mathcal{F}, L^2(P), \sigma/2)}. \tag{3.189}$$

If, moreover, the class $\mathcal{F}$ is full for H, P and F with constant c, then

$$\frac{c}{32K}\sqrt{n\sigma^2 H\left(\frac{2\|F\|_{L^2(P)}}{\sigma}\right)} \leq E\Big\|\sum_{i=1}^{n} f(X_i)\Big\|_{\mathcal{F}} \leq 8\sqrt{22}\sqrt{n\sigma^2 H\left(\frac{2\|F\|_{L^2(P)}}{\sigma}\right)} \tag{3.190}$$

(fullness is only required for the left-hand side inequality).