Mathematical Foundations of Infinite-Dimensional Statistical Models Chapter 3.2

Evarist Giné and Richard Nickl

Presenter : Seonghyeon Kim

2018.10.26

3.2 Rademacher Processes

Rademacher Processes

$\epsilon_1, \cdots, \epsilon_n$: independent Rademacher variables, independent of the variables X_i

Conditionally on the variables X_i, $\sum_{i=1}^n \epsilon_i f(X_i)$ is a Rademacher process.

$$t \to \sum_{i=1}^n t_i \epsilon_i, t \in T \subseteq \mathbb{R}^n$$

Since Rademacher processes are sub-Gaussian, the metric entropy moment bounds for sub-Gaussian processes given in Section 2.3 apply to these processes.

3.2.1 A Comparison Principle for Rademacher Processes

Contraction vanishing at 0

$\varphi : \mathbb{R} \to \mathbb{R}$ is called contraction vanishing at 0 if φ satisfies

$$|\varphi(s) - \varphi(t)| \leq |s - t|, \quad \text{for all } s, t \in \mathbb{R}$$

$$\varphi(0) = 0$$

Theorem 3.2.1

Let F be a nonnegative, convex and nondecreasing function defined on $[0, \infty)$. Let $\varphi_i : \mathbb{R} \to \mathbb{R}$ be contractions vanishing at 0, and let T be a bounded set of $\mathbb{R}_n$, $n < \infty$. Then

$$EF\left(\frac{1}{2}\left\|\sum_{i=1}^{n} \epsilon_i \varphi_i(t_i)\right\|_T\right) \leq EF\left(\left\|\sum_{i=1}^{n} \epsilon_i t_i\right\|_T\right)$$

where $t = (t_1, \cdots, t_n)$ and $\|\cdot\|_T$ denotes, as usual, supremum over all $t \in T$.

Notation for Corollary 3.2.2

$\mathcal{F}$: a countable class of measurable functions $S \to \mathbb{R}$ such that $F(x) < \infty$ for all $x \in S$, where $F(x) = \sup_{f \in \mathcal{F}} |f(x)|$.

$U = max_{i=1}^{n} |F(X_i)|$
$\sigma^2 = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} Ef^2(X_i)/n$. We assume $\sigma^2 < \infty$.

Corollary 3.2.2

$$E\left\|\sum_{i=1}^{n}\epsilon_i f^2(X_i)\right\|_{\mathcal{F}} \leq 4E\left[U\left\|\sum_{i=1}^{n}\epsilon_i f(X_i)\right\|_{\mathcal{F}}\right]$$

.

$$E\left\|\sum_{i=1}^{n} f^2(X_i)\right\|_{\mathcal{F}} \leq n\sigma^2 + 8E\left[U\left\|\sum_{i=1}^{n}\epsilon_i f(X_i)\right\|_{\mathcal{F}}\right]$$

Proof of Corollary 3.2.2

For $X_1, \cdots, X_n$ fixed, we take in Theorem 3.2.1
$t_i = Uf(X_i)$,
$T = \{(Uf(X_1), \cdots, Uf(X_n)) : f \in \mathcal{F}\}$ and
$\varphi_i(s) = s^2/2U^2 \wedge U^2/2$, so
$\varphi_i(t_i) = \varphi_i(Uf(X_i)) = f^2(X_i)/2 \wedge U^2/2 = f^2(X_i)/2.$

Proof of Corollary 3.2.2

The preceding theorem gives

$$\frac{1}{4}E_\epsilon \left\|\sum_{i=1}^n \epsilon_i f^2(X_i)\right\|_{\mathcal{F}} \leq UE_\epsilon \left\|\sum_{i=1}^n \epsilon_i f(X_i)\right\|_{\mathcal{F}}$$

. Integrating with respect to the variables X_i and then applying the basic randomisation inequality

$$\begin{aligned} E\left\|\sum_{i=1}^n f^2(X_i)\right\|_{\mathcal{F}} &\leq n\sigma^2 + E\left\|\sum_{i=1}^n (f^2(X_i) - Ef^2(X_i))\right\|_{\mathcal{F}} \\ &\leq n\sigma^2 + 2E\left\|\sum_{i=1}^n \epsilon_i f^2(X_i)\right\|_{\mathcal{F}} \\ &\leq n\sigma^2 + 8E\left[U\left\|\sum_{i=1}^n \epsilon_i f(X_i)\right\|_{\mathcal{F}}\right] \end{aligned}$$

3.2.2 Convex Distance Concentration and Rademacher Processes

Weighted Hamming Distance and Convex Distance

- Weighted Hamming distance : Given a vector $a \in \mathbb{R}^n$, $a_i \geq 0$

$$d_a(x, y) = \sum_{i=1}^{n} a_i I_{x_i \neq y_i}, \quad x, y \in S$$

$$d_a(x, A) = \inf\{d_a(x, y) : y \in A\}, \quad x \in S, A \subset S$$

- Convex distance :

$$d_c(x, y) = \sup_{|a| \leq 1} d_a(x, y), \quad x, y \in S$$

$$d_c(x, A) = \inf\{d_c(x, y) : y \in A\}, \quad x \in S, A \subset S$$

Lemma 3.2.3

$U_A(x) := \{u = (u_i)_{i=1}^n \in \{0,1\}^n : \exists y \in A \text{ with } y_i = x_i \text{ if } u_i = 0\}$
$V_A(x)$: convex hull of $U_A(x)$

$d_c(x, A) = \inf\{|v| : v \in V_A(x)\}$ and the infimum is attained at a point in $V_A(x)$.

Theorem 3.2.4 (Talagrand's inequality for the convex distance)

For any $n \in \mathbb{N}$, if $X = (X_1, \cdots, X_n)$ is a vector of independent random variables taking values in the product space $S^{(n)} = \prod_{k=1}^{n} S_k$, and $A \subseteq S^{(n)}$, then

$$E\left(e^{d_c^2(X,A)/4}\right) \leq \frac{1}{Pr\{X \in A\}}$$

hence, for all $t \geq 0$,

$$Pr\{X \in A\} Pr\{d_c(X, A) \geq t\} \leq e^{` t^2/4}.$$

Corollary 3.2.5

Let $S = S_1 \times \cdots \times S_n$ be a product of measurable spaces, and let P be a product probability measure on it. Let $F : S \to \mathbb{R}$ be a measurable function satisfying the following Lipschitz property for the distance d_a: for every $x \in S$, there is $a = a(x) \in \mathbb{R}_n$ with $|a| = 1$ such that

$$F(x) \leq F(y) + d_a(x, y), \quad y \in S.$$

Let m_F be a median of F for P. Then, for all $t \geq 0$,

$$P\{|F - m_F| \geq t\} \leq 4e^{-t^2/4}.$$

Proof of Corollary 3.2.5

Taking $A = \{F \leq m\}$,

$$F(x) \leq m + d_a(x, A) \leq m + d_c(x, A)$$

$$P\{F \geq m + t\} \leq P\{d_c(x, A) \geq t\}$$

Corollary 3.2.6 (Concentration inequality for Rademacher (and other) processes)

Let $X_i, 1 \le i \le n$, be independent real random variables such that, for real numbers a_i, b_i,

$$a_i \le X_i \le b_i, 1 \le i \le n.$$

Let T be a countable subset of $\mathbb{R}^n$, and set

$$Z = \sup_{t \in T} \sum_{i=1}^{n} t_i X_i,$$

where $t = (t_1, \cdots, t_n)$. Let m_Z be a median of Z. Then, if $\tilde{\sigma} := \sup_{t \in T} (\sum_{i=1}^{n} t_i^2 (b_i - a_i)^2)^{1/2}$ is finite, we have that, for every $r \ge 0$,

$$Pr\{|Z - m_Z| \ge r\} \le 4e^{-r^2/4\tilde{\sigma}^2},$$

$$|EZ - m_Z| \le 4\sqrt{\pi}\tilde{\sigma} \quad and \quad Var(Z) \le 16\tilde{\sigma}^2.$$

Theorem 3.2.7

For $n < \infty$ and a countable set $T \subset \mathbb{R}_n$, set

$$Z = \sup_{t \in T} \sum_{i=1}^{n} t_i \epsilon_i, \sigma = \sup_{t \in T} \left(\sum_{i=1}^{n} t_i^2 \right)^{1/2},$$

and let m_Z be a median of Z. Then, if $\sigma < \infty$,

$$Pr\{|Z - m_Z| \geq r\} \leq 4e^{-r^2/8\sigma^2}$$

and, consequently,

$$E|Z - m_Z| \leq 4\sqrt{2\pi}\sigma \quad and \quad Var(Z) \leq 32\sigma^2.$$

Corollary 3.2.6 vs Theorem 3.2.7

- For Rademacher Process

$$\tilde{\sigma} = \sup_{t \in T} \left(\sum_{i=1}^{n} t_i^2 (b_i - a_i)^2 \right)^{1/2} = 2 \sup_{t \in T} \left(\sum_{i=1}^{n} t_i^2 \right)^{1/2} = 2\sigma$$

- Corollary 3.2.6

$$Pr\{|Z - m_Z| \geq r\} \leq 4e^{-r^2/16\sigma^2}$$

- Theorem 3.2.7

$$Pr\{|Z - m_Z| \geq r\} \leq 4e^{-r^2/8\sigma^2}$$

Proposition 3.2.8 (Khinchin-Kahane inequalities)

For Z as in Theorem 3.2.7, for all $p > q > 0$, there exists $C_q < \infty$ such that

$$(E|Z|^p)^{1/p} \le C_q \sqrt{p} (E|Z|^q)^{1/q}.$$

Moreover, there are $\tau > 0$ and $c > 0$ such that $Pr\{|Z| > c \|Z\|_2\} \ge \tau$.

3.2.3 A Lower Bound for the Expected Supremum of a Rademacher Process

Theorem 3.2.9(Variation on Sudakov's inequality

There exists a finite constant $K > 0$ such that for every $n \in \mathbb{N}$ and $\epsilon > 0$, if T is a bounded subset of $\mathbb{R}^n$ such that

$$E \sup_{t \in T} \left| \sum_{i=1}^{n} \epsilon_i t_i \right| \leq \frac{1}{K} \frac{\epsilon^2}{max_{1 \leq i \leq n} |t_i|}, \textit{for all } t \in T,$$

then

$$\epsilon \sqrt{\log N(T, d_2, \epsilon)} \leq KE \sup_{t \in T} \left| \sum_{i=1}^{n} \epsilon_i t_i \right|$$

.