Mathematical Foundations of Infinite-Dimensional Statistical Models

Chap. 2.3 The Metric Entropy Bound for Suprema of Sub-Gaussian Processes

Gyuseung Baek

July 27, 2018

Introduction

- Introduce sub-Gaussian preocess
- Express the bound for suprema of sub-Gaussian process by the metric entorpy of T.

Sub-Gaussian random variable

- A square integrable random variable ζ is said to be *sub-Gaussian* with parameter $\sigma > 0$ if for all $\lambda \in \mathbb{R}$,

$$Ee^{\lambda\zeta} \leq e^{\lambda^2\sigma^2/2}. \tag{1}$$

- By taylor's expansion, we can see
  - $E\zeta = 0$.
  - $E\zeta^2 \leq \sigma^2$.

Example - Rademacher random variable

- ϵ is a *Rademacher* random variable if $Pr\{\epsilon = 1\} = Pr\{\epsilon = -1\} = 1/2$.
- Suppose a_i is real and ϵ_i is i.i.d. Rademacher r.v.s for all i.
  The linear combinations of independent Rademacher random variables $\zeta = \sum_{i=1}^{n} a_i \epsilon_i$ is a sub-Gaussian r.v.s with parameter $\sum_i a_i^2/2$.

Sub-Gaussian tail

Let ζ be a sub-Gaussian r.v.. Then for all $t > 0$,

$$\begin{aligned} Pr\{\zeta \geq t\} &\leq e^{-t^2/2\sigma^2} \\ Pr\{|\zeta| \geq t\} &\leq 2e^{-t^2/2\sigma^2} \end{aligned} \tag{2}$$

From (2), ζ/c enjoys square exponential integrability for some $0 < c < \infty$: if $c^2 > 2\sigma^2$, then

$$Ee^{\zeta^2/c^2} = \int_0^\infty 2te^{t^2} Pr\{|\zeta| > ct\} dt \leq \frac{2}{c^2/2\sigma^2 - 1} < \infty \tag{3}$$

Orlicz norm

The collection of random variables ζ on (Ω, Σ, Pr) that satisfy (3) constitutes a vector space, denoted by $L^{\psi_2}(\Omega.\Sigma, Pr)$, and the functional

$$||\zeta||_{\psi_2} = \inf\{c > 0 : E\psi_2(|\zeta|/c) \leq 1\} \tag{4}$$

where $\psi_2(x) := e^{x^2} - 1$ is a pseudo-norm on it for which L^{ψ_2}, with identification of a.s. equal functions, is a Banach space.

With this notation, (3) shows that

$$Pr\{|\zeta| \geq t\} \leq 2e^{-t^2/2\sigma^2} \text{ for all } t > 0 \text{ implies } ||\zeta||_{\psi_2} \leq \sqrt{6}\sigma. \tag{5}$$

Conversely, if $\zeta \in L^{\psi_2}$ and $E\zeta = 0$, then ζ is sub-Gaussian with parameter

$\sigma \leq \sqrt{6}||\zeta||_{\pi_2}$.

Lemma 2.3.1

If ζ is centered, T.F.A.E.

- $\zeta \in L_2^{\psi}$.
- ζ satisties the sub-Gaussian tail inequality (2).

$$Pr\{|\zeta| \geq t\} \leq 2e^{-t^2/2\sigma^2}$$

- ζ is sub-Gaussian for some σ_2.

Lemma 2.3.2

(Extended version of lemma 2.3.1)

Assume that

$$Pr\{|\zeta| \geq t\} \leq 2Ce^{-t^2/2\sigma^2}, \quad t > 0$$

for some $C \leq 1$ and $\sigma > 0$, a condition implied by the Laplace transform condition

$$Ee^{\lambda\zeta} \leq Ce^{\lambda^2\sigma^2/2}, \quad \lambda \in \mathbb{R}$$

Then ζ also satisfies

$$||\zeta||_{\psi_2} \leq \sqrt{2(2C+1)}\sigma$$

Morover, if in addition $E\zeta = 0$, then also

$$Ee^{\lambda\zeta} \leq e^{3\lambda^2(2(2C+1))\sigma^2}, \quad \lambda \in \mathbb{R}$$

that is, ζ is sub-Gaussian with constant $\tilde{\sigma}^2 = 12(2C+1)\sigma^2$

Lemma 2.3.3

(maximal inequality for variables in L_2^ψ not necessarily centred)

Let $\zeta \in L_2^\psi, i = 1, \cdots, N, 2 \leq N < \infty$. Then

$$\left|\left|\max_{i \leq N} |\zeta_i|\right|\right|_{\psi_2} \leq 4\sqrt{\log N} \max_{i \leq N} ||\zeta_i||_{\psi_2}, \tag{6}$$

and, in particular, there exist $K_p < \infty, 1 \leq p < \infty$, such that

$$\left|\left|\max_{i \leq N} |\zeta_i|\right|\right|_{L^p} \leq K_p\sqrt{\log N} \max_{i \leq N} ||\zeta_i||_{\psi_2}, \tag{7}$$

Bound for the Expectation of Maxima

- Let Φ be a nonnegative, strictly increasing, convex function on a finite or infinite interval I
- Let $\zeta_i, i \leq i \leq N$, be r.v.s taking values in I and s.t. $E\Phi(\zeta_i) < \infty$. Then

$$\begin{aligned} \Phi\left(E\max_{i\leq N}\zeta_i\right) \leq E\Phi\left(\max_{i\leq N}\zeta_i\right) &= E\max_{i\leq N}\Phi(\zeta_i) \\ \sum_{i=1}^{N} E\Phi(\zeta_i) &\leq N\max_{i\leq N}E\Phi(\zeta_i), \end{aligned} \tag{8}$$

and, inverting Φ,

$$E\max_{i\leq N}\zeta_i \leq \Phi^{-1}\left(N\max_{i\leq N}E\Phi(\zeta_i)\right) \tag{9}$$

Lemma 2.3.4 For sub-Gaussian ζ_is with parameter σ_i, respectively,

$$E\max_{i\leq N}\zeta_i \leq \sqrt{2\log N}\max_{i\leq N}\sigma_i, \quad E\max_{i\leq N}|\zeta_i| \leq \sqrt{2\log 2N}\max_{i\leq N}\sigma_i \tag{10}$$

Sub-Gaussian process

Definition 2.3.5 A centred stochastic process $X(t), t \in T$, is sub-Gaussian w.r.t. a distance(or pseudo-distance) d on T if its increments satistfy the sub-Gaussian inequality, that is, if

$$Ee^{\lambda(X(t)-X(s))} \leq e^{\lambda^2 d^2(s,t)/2} \quad \lambda \in \mathbb{R}, \quad s, t \in T \tag{11}$$

Therefore, $X(t), t \in T$ is sub-Gaussian if for all $s, t \in T$, $X(t) - X(s)$ is sub-Gaussian r.v. with parameter $d(s, t)$.

If the centred process X satisfies

$$Ee^{\lambda(X(t)-X(s))} \leq Ce^{\lambda^2 d^2(s,t)/2} \text{ or } Pr\{|X(t) - X(s)| \geq u\} \leq Ce^{-u^2/2d^2(s,t)}, \tag{12}$$

for all $\lambda \in mR, u > 0$ and $s, t \in T$ and some $C > 1$, then, by Lemma 2.3.2, X is sub-Gaussian for the distance $\tilde{d}(s, t) := \sqrt{12(2C + 1)}$ d.

Sub-Gaussian process

- Centred Gaussian processes $X(t)$ are sub-Gaussian with respect to the L^2-distance $d_X(s,t) = ||X(t) - X(s)||_{L^2}$.
- If X is a sub-Gaussian process with respect to d, From (2),

$$E(X(t) - X(s))^2 \leq d^2(s,t)$$

- Lemma 2.3.4 implies that if F is a finite subset of $T \times T$ of cardinality N, then

$$E \max_{(s,t)\in F} |X(t) - X(s)| \leq \sqrt{2 \log 2N} \max_{(s,t)\in F} d(s,t). \quad (13)$$

- We want to extend (13) for $\sup_{t\in T} |X(t)|$ or $\sup_{s,t\in T, d_X(s,t)\leq\delta} |X(t) - X(s)|$.

Metric entropy

- Need to measure the size of infinite metric space (T, d) to replace RHS of (13)
- The *covering number* $N(T, d, \epsilon)$ of (T, d)

$$N(T, d, \epsilon) := \min\left\{ n : \text{there exist } t_1, \cdots, t_n \in T \text{ such that } T \subseteq \bigcup_{i=1}^{n} B(t_i, \epsilon) \right\},$$

- The *packing number* $D(T, d, \epsilon)$ of (T, d)

$$D(T, d, \epsilon) := \min\left\{ n : \text{there exist } t_1, \cdots, t_n \in T \text{ such that } \min_{1 \le i,j \le n} d(t_i, t_j) > \epsilon \right\},$$

- For all $\epsilon > 0$,

$$N(T, d, \epsilon) \le D(T, d, \epsilon) \le N(T, d, \epsilon/2)$$

- The logarithm of the covering number of (T, d) is known as its **metric entropy**

Theorem 2.3.6

(Bound for Suprema of sub-Gaussian process with finite indicator set)

Let (T, d) be a pseudo-metric space, and let $X(t), t \in T$, be a stochastic process sub-Gaussian with respect to the pseudo-distance d. Then, for all finite subsets $S \subseteq T$ and points $t_0 \in T$, the following inequalities hold:

$$E \max_{t \in S} |X(t)| \leq E|X(t_0)| + 4\sqrt{2} \int_0^{D/2} \sqrt{\log 2N(T, d, \epsilon)} d\epsilon \tag{14}$$

where D is the diameter of (T, d), and

$$E \max_{s,t \in S, d(s,t) \leq \delta} |X(t) - X(s)| \leq (16\sqrt{2} + 2) \int_0^{\delta} \sqrt{\log 2N(T, d, \epsilon)} d\epsilon \tag{15}$$

for all $\delta > 0$, where the integrals are taken to be 0 if $D = 0$.

Theorem 2.3.7

(Generalization of theorem 2.3.6)

Let (T, d) be a (pseudo-)metric space, and let $X(t), t \in T$, be a sub-Gaussian process relative to d. Assume that

$$\int_0^\infty \sqrt{\log N(T, d, \epsilon)} d\epsilon < \infty \qquad (16)$$

Then (a) $X(t), t \in T$, is sample d-continuous (in particular, X admits a separable version)
(b) Any separable version of $X(t), t \in T$, that we keep denoting by $X(t)$ has almost all its sample paths bounded and uniformly d-continuous, and satisfies the inequalities

$$E \sup_{t \in T} |X(t)| \leq E|X(t_0)| + 4\sqrt{2} \int_0^\infty \sqrt{\log 2N(T, d, \epsilon)} d\epsilon, \qquad (17)$$

where $t_0 \in T, D$ is the diameter of (T, d) and

$$E \max_{s,t \in T, d(s,t) \leq \delta} |X(t) - X(s)| \leq (16\sqrt{2} + 2) \int_0^\delta \sqrt{\log 2N(T, d, \epsilon)} d\epsilon \qquad (18)$$

for all $\delta > 0$.

Theorem 2.3.8

(Dudley's Theorem)

If $X(t), t \in T$ is a sub-Gaussian process for a pseudo-metric d such that (T, d) has positive d-diameter and satisfies the metric entropy condition (16), then, for any separable version of X (still denoted by X), we haver, with the convention $0/0 = 0$,

$$E\left[\sup_{s,t \in T} \frac{|X(t) - X(s)|}{\int_0^{d(s,t)} \sqrt{\log N(T, d, \epsilon)} d\epsilon}\right] < \infty \tag{19}$$