Stat 6390.001: High Dimensional Probability and Statistics, Fall 2026 **Topic 17**

Gaussian and Rademacher Complexity and Sub-Gaussian Processes

Lecturer: Yun Wei *Scribe: Hyeonduk Sim and Mengfan Fu*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1 Introduction

These notes cover packing numbers, Rademacher and Gaussian complexities, sub-Gaussian processes, and Dudley's entropy integral bound. The final application is to uniform deviations of empirical measures. The notes follow Sections 5.1–5.3 of [1].

2 Packing Numbers

Definition 2.1 (δ-packing number)**.** Let (T, ρ) be a metric space. A set $\{\theta^1, \dots, \theta^M\} \subseteq T$ is a δ-packing if

$$\rho(\theta^i, \theta^j) > \delta \qquad \text{for all } i \neq j.$$

The δ-packing number $M(\delta; T, \rho)$ is the cardinality of the largest δ-packing.

Let $N(\delta; T, \rho)$ denote the covering number, the smallest number of closed ρ-balls of radius δ with centers in T needed to cover T. Packing and covering numbers satisfy

$$M(2\delta; T, \rho) \leq N(\delta; T, \rho) \leq M(\delta; T, \rho). \tag{1}$$

3 Canonical Rademacher and Gaussian Processes

Let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d)$ have independent Rademacher coordinates, each taking -1 and 1 with equal probability. Let $w \sim N(0, I_d)$. For $T \subseteq \mathbb{R}^d$, define

$$R_\theta = \langle \varepsilon, \theta \rangle, \qquad G_\theta = \langle w, \theta \rangle, \qquad \theta \in T.$$

The collections $\{R_\theta : \theta \in T\}$ and $\{G_\theta : \theta \in T\}$ are the canonical Rademacher and Gaussian processes. Their expected suprema are

$$R(T) = \mathbb{E} \sup_{\theta \in T} R_\theta, \qquad G(T) = \mathbb{E} \sup_{\theta \in T} G_\theta. \tag{2}$$

Example 3.1 (Example: the ℓ_2 unit ball)**.** Let $B_2^d = \{\theta \in \mathbb{R}^d : \|\theta\|_2 \leq 1\}$. Using the duality of the Euclidean norm,

$$R(B_2^d) = \mathbb{E} \sup_{\theta \in B_2^d} \langle \varepsilon, \theta \rangle = \mathbb{E}\|\varepsilon\|_2 = \mathbb{E}\Big(\sum_{i=1}^d \varepsilon_i^2 \Big)^{1/2} = \sqrt{d}, \tag{3}$$

$$G(B_2^d) = \mathbb{E} \sup_{\theta \in B_2^d} \langle w, \theta \rangle = \mathbb{E}\|w\|_2 \leq \big(\mathbb{E}\|w\|_2^2\big)^{1/2} = \Big(\sum_{i=1}^d \mathbb{E} w_i^2 \Big)^{1/2} = \sqrt{d}. \tag{4}$$

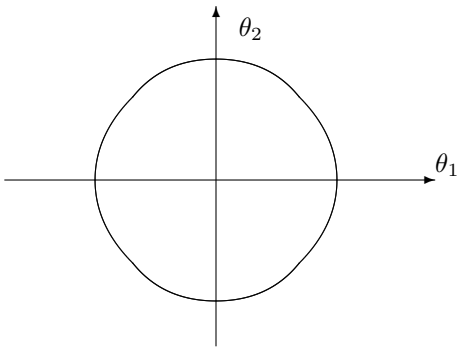
On the other hand,

$$\mathbb{E}\|w\|_2 \geq \sqrt{d}\,(1 - o(1)),$$

which is left as an exercise (hint: use the concentration of the sub-exponential variables $w_i^2 - 1$; see Chapter 2). Hence

$$\frac{G(B_2^d)}{\sqrt{d}} = 1 - o(1).$$

In two dimensions, B_2^2 is the Euclidean disk.



$$B_2^2 = \{(\theta_1, \theta_2) : \theta_1^2 + \theta_2^2 \le 1\}.$$

Example 3.2 (Example: the ℓ_1 unit ball)**.** Let $B_1^d = \{\theta \in \mathbb{R}^d : \|\theta\|_1 \le 1\}$.
Recall. For any norm $\|\cdot\|$ and its dual norm $\|\cdot\|_*$,

$$\sup_{\|\theta\| \le 1} \langle y, \theta \rangle = \|y\|_*.$$

In particular, the dual norm of ℓ_1 is ℓ_∞. Consequently,

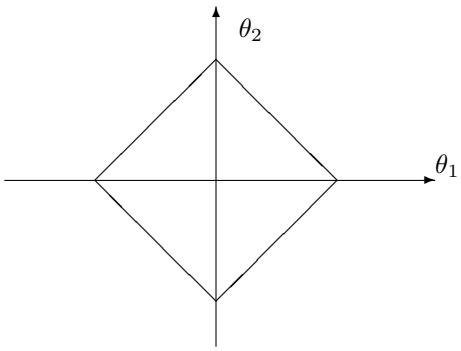
$$R(B_1^d) = \mathbb{E} \sup_{\theta \in B_1^d} \langle \varepsilon, \theta \rangle = \mathbb{E}\|\varepsilon\|_\infty = \mathbb{E} \max_{1 \le i \le d} |\varepsilon_i| = 1, \tag{5}$$

$$G(B_1^d) = \mathbb{E} \sup_{\theta \in B_1^d} \langle w, \theta \rangle = \mathbb{E}\|w\|_\infty = \mathbb{E} \max_{1 \le i \le d} |w_i|. \tag{6}$$

The Gaussian maximal inequality from Homework 1 gives

$$G(B_1^d) \le \sqrt{2\log(2d)} = \sqrt{2\log d}\,(1 + o(1)) \qquad (d \to \infty). \tag{7}$$

In two dimensions, B_1^2 is the diamond $\{(\theta_1, \theta_2) : |\theta_1| + |\theta_2| \le 1\}$.



$$B_1^2 = \{(\theta_1, \theta_2) : |\theta_1| + |\theta_2| \le 1\}.$$

Remark 3.3. Homework 5 will ask for the complexity of ℓ_0-balls, where $\|\theta\|_0$ counts nonzero coordinates.

4 Metric Entropy and Sub-Gaussian Processes

Definition 4.1 (Sub-Gaussian process)**.** A collection of mean-zero random variables $\{X_\theta : \theta \in T\}$ is a sub-Gaussian process with respect to ρ_X if

$$\mathbb{E}\exp\big(\lambda(X_\theta - X_{\widetilde{\theta}})\big) \le \exp\left(\frac{\lambda^2 \rho_X^2(\theta,\widetilde{\theta})}{2}\right) \tag{8}$$

for all $\theta, \widetilde{\theta} \in T$ and $\lambda \in \mathbb{R}$. Equivalently, each increment $X_\theta - X_{\widetilde{\theta}}$ is sub-Gaussian with parameter $\rho_X(\theta,\widetilde{\theta})$.

Example 4.2 (Canonical Gaussian process)**.** For $G_\theta = \langle w, \theta\rangle$,

$$G_\theta - G_{\widetilde{\theta}} = \langle w, \theta - \widetilde{\theta}\rangle \sim N\Big(0, \|\theta - \widetilde{\theta}\|_2^2\Big).$$

In particular,

$$\mathbb{E}\exp\big(\lambda(G_\theta - G_{\widetilde{\theta}})\big) = \exp\left(\frac{\lambda^2\|\theta - \widetilde{\theta}\|_2^2}{2}\right).$$

Thus the canonical Gaussian process is sub-Gaussian with respect to $\rho_X(\theta,\widetilde{\theta}) = \|\theta - \widetilde{\theta}\|_2$.

Example 4.3 (Canonical Rademacher process)**.** For $R_\theta = \langle \varepsilon, \theta\rangle$,

$$R_\theta - R_{\widetilde{\theta}} = \langle \varepsilon, \theta - \widetilde{\theta}\rangle = \sum_{i=1}^d \varepsilon_i(\theta_i - \widetilde{\theta}_i).$$

Recall:

- ε_i is 1-sub-Gaussian;
- hence $\varepsilon_i(\theta_i - \widetilde{\theta}_i)$ is $|\theta_i - \widetilde{\theta}_i|$-sub-Gaussian;
- since $\varepsilon_1, \dots, \varepsilon_d$ are independent, $\sum_{i=1}^d \varepsilon_i(\theta_i - \widetilde{\theta}_i)$ is sub-Gaussian with parameter $\big(\sum_{i=1}^d (\theta_i - \widetilde{\theta}_i)^2\big)^{1/2} = \|\theta - \widetilde{\theta}\|_2$.

Hence $\{R_\theta : \theta \in T\}$ is sub-Gaussian with respect to $\rho_X(\theta,\widetilde{\theta}) = \|\theta - \widetilde{\theta}\|_2$.

Our target is to bound $\mathbb{E}\sup_{\theta\in T} X_\theta$.

Lemma 4.4. *Let $\{X_\theta : \theta \in T\}$ be a zero-mean random process. Then*

$$\mathbb{E}\sup_{\theta\in T} X_\theta \le \mathbb{E}\sup_{\theta,\widetilde{\theta}\in T}(X_\theta - X_{\widetilde{\theta}}). \tag{9}$$

Proof. Fix $\theta_0 \in T$. By the mean-zero assumption,

$$\mathbb{E}\sup_{\theta\in T} X_\theta = \mathbb{E}\sup_{\theta\in T}(X_\theta - X_{\theta_0}) \le \mathbb{E}\sup_{\theta,\widetilde{\theta}\in T}(X_\theta - X_{\widetilde{\theta}}).$$

□

4.1 Dudley's entropy integral

Let $D = \sup_{\theta,\widetilde{\theta}\in T} \rho_X(\theta,\widetilde{\theta})$ denote the diameter of T. The Dudley entropy integral is

$$J(\delta; D) = \int_{\delta}^{D} \sqrt{\log N(u; T, \rho_X)}\, du. \tag{10}$$

The term $\log N(u; T, \rho_X)$ is the metric entropy.

Theorem 4.5 (Dudley's entropy integral bound)**.** *Let $\{X_\theta : \theta \in T\}$ be a zero-mean sub-Gaussian process with respect to ρ_X. Then*

$$\mathbb{E}\left[\sup_{\theta,\widetilde{\theta}\in T} (X_\theta - X_{\widetilde{\theta}})\right] \le 32 J(0; D). \tag{11}$$

Proof: next class.

5 Application: the Classical Glivenko–Cantelli Theorem

Let $X_1, \dots, X_n$ be independent with common distribution P, and let $\mathcal{F}$ be a class of functions. Define

$$P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}, \qquad \|P_n - P\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}_P[f(X)] \right|.$$

Recall the symmetrization inequality:

$$\mathbb{E}\|P_n - P\|_{\mathcal{F}} \le 2R_n(\mathcal{F}) = 2\,\mathbb{E}_X R\big(\mathcal{F}(X_1^n)/n\big), \tag{12}$$

where $\mathcal{F}(x_1^n) = \{(f(x_1), \dots, f(x_n)) : f \in \mathcal{F}\}$ and

$$R\big(\mathcal{F}(x_1^n)/n\big) = \mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(x_i) \right|.$$

Fix $x_1^n = (x_1, \dots, x_n)$ and consider

$$Z_f := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \varepsilon_i f(x_i) = \Big\langle \varepsilon,\ \underbrace{\tfrac{1}{\sqrt{n}}\big(f(x_1), \dots, f(x_n)\big)}_{=:\,\theta(f)} \Big\rangle. \tag{13}$$

By Example 4.3, $\{Z_f : f \in \mathcal{F}\}$ is a sub-Gaussian process with respect to

$$\rho_X(f,g) = \|\theta(f) - \theta(g)\|_2 = \Big\| \tfrac{1}{\sqrt{n}}\big(f(x_1), \dots, f(x_n)\big) - \tfrac{1}{\sqrt{n}}\big(g(x_1), \dots, g(x_n)\big) \Big\|_2,$$

where $\theta(g) = \frac{1}{\sqrt{n}}(g(x_1), \dots, g(x_n))$ for $g \in \mathcal{F}$. In terms of f and g,

$$\rho_X(f,g) = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} (f(x_i) - g(x_i))^2 \right)^{1/2} = \|f - g\|_{P_n}. \tag{14}$$

By Lemma 4.4 and Theorem 4.5,

$$\begin{aligned} \mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}} Z_f &\le \mathbb{E}_\varepsilon \sup_{f,g\in\mathcal{F}} (Z_f - Z_g) \\ &\le 32 \int_0^\infty \sqrt{\log N(u; \mathcal{F}, \|\cdot\|_{P_n})}\, du. \end{aligned} \tag{15}$$

References

[1] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019. Sec 5.2 and Sec 5.3