Stat 6390.001: High Dimensional Probability and Statistics, Fall 2026 **Topic 18**

Dudley's Entropy Integral, VC Classes, and Chaining

Lecturer: Yun Wei *Scribe: Meng He, Shirou Zeng*

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 Sub-Gaussian Processes and Metric Entropy

Let $\{X_\theta\}_{\theta\in T}$ be a stochastic process indexed by a set T. The purpose of this lecture is to control the expected supremum of this process by measuring the size of T at a sequence of increasingly fine scales. This multiscale method is called *chaining*.

Definition 1.1 (Sub-Gaussian process)**.** A process $\{X_\theta\}_{\theta\in T}$ is sub-Gaussian with respect to a metric ρ_X if, for every $\theta, \widetilde{\theta} \in T$ and every $\lambda \in \mathbb{R}$,

$$\mathbb{E}\exp\left\{\lambda\left(X_\theta - X_{\widetilde{\theta}}\right)\right\} \le \exp\left\{\frac{\lambda^2\rho_X^2(\theta,\widetilde{\theta})}{2}\right\}. \tag{1}$$

Definition 1.2 (Covering number and Dudley entropy integral)**.** For $u > 0$, the covering number $N(u; T, \rho_X)$ is the smallest number of closed ρ_X-balls of radius u needed to cover T. If

$$D = \sup_{\theta,\widetilde{\theta}\in T} \rho_X(\theta,\widetilde{\theta}),$$

then the truncated Dudley entropy integral is

$$J(s; D) := \int_s^D \sqrt{\log N(u; T, \rho_X)}\, du, \qquad 0 \le s \le D. \tag{2}$$

The function $u \mapsto N(u; T, \rho_X)$ is nonincreasing. Consequently, the entropy $\log N(u; T, \rho_X)$ records how complicated the index set appears at resolution u.

Theorem 1.3 (Dudley's entropy integral bound)**.** *Let $\{X_\theta \mid \theta \in T\}$ be a sub-Gaussian process with respect to ρ_X. Then*

$$\mathbb{E}\left[\sup_{\theta,\widetilde{\theta}\in T}(X_\theta - X_{\widetilde{\theta}})\right] \le 16J(0,\infty).$$

2 Application to Rademacher Complexity

Example 2.1 (Dudley's bound for Rademacher complexity)**.** For a function $f : \mathcal{X} \to \mathbb{R}$, the empirical $L^2(P_n)$-norm is

$$\|f\|_{\mathbb{P}_n} := \left(\frac{1}{n}\sum_{i=1}^n f^2(x_i)\right)^{1/2}.$$

For a function class $\mathcal{F}$, define the conditional empirical Rademacher complexity by

$$\mathfrak{R}\left(\frac{\mathcal{F}(x_1^n)}{n}\right) := \mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i f(x_i)\right|\right].$$

The expectation is over the Rademacher variables, conditional on the observed sample. Introduce the Rademacher process

$$Z_f := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i f(x_i), \qquad f \in \mathcal{F}.$$

The process $\{Z_f\}_{f\in\mathcal{F}}$ is sub-Gaussian with respect to the empirical pseudometric

$$\rho(f,g) := \|f-g\|_{\mathbb{P}_n} = \left(\frac{1}{n}\sum_{i=1}^{n}\left(f(x_i)-g(x_i)\right)^2\right)^{1/2}, \qquad f,g \in \mathcal{F}.$$

Suppose that for any x_1^n, there exists a f_0 such that $f_0(x_1^n) = (0,0,\ldots,0)$. Using $|x| = \max\{x,-x\}$, we have

$$\begin{aligned}
\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(x_i)\right|\right] &= \frac{1}{\sqrt{n}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}|Z_f|\right] \\
&= \frac{1}{\sqrt{n}}\mathbb{E}_\epsilon\left[\max\left\{\sup_{f\in\mathcal{F}} Z_f, \sup_{f\in\mathcal{F}}(-Z_f)\right\}\right] \\
&\le \frac{1}{\sqrt{n}}\left(\mathbb{E}_\epsilon \sup_{f\in\mathcal{F}} Z_f + \mathbb{E}_\epsilon \sup_{f\in\mathcal{F}}(-Z_f)\right) \quad (\text{since } \sup_{f\in\mathcal{F}} Z_f \ge 0) \\
&= \frac{2}{\sqrt{n}}\mathbb{E}_\epsilon \sup_{f\in\mathcal{F}} Z_f \\
&\le \frac{32}{\sqrt{n}}\int_0^D \sqrt{\log N(u;\mathcal{F},\|\cdot\|_{\mathbb{P}_n})}\,du.
\end{aligned}$$

Here the equality in the fourth line follows from the symmetry of the Rademacher variables, and the last inequality follows from Dudley's entropy integral bound.

Suppose that $\mathcal{F}$ is b-uniformly bounded. Then

$$\begin{aligned}
D &= \sup_{f,g\in\mathcal{F}} \|f-g\|_{\mathbb{P}_n} \\
&= \sup_{f,g\in\mathcal{F}}\left(\frac{1}{n}\sum_{i=1}^{n}(f(x_i)-g(x_i))^2\right)^{1/2} \\
&\le \sup_{f,g\in\mathcal{F}}\left(\frac{1}{n}\sum_{i=1}^{n}4b^2\right)^{1/2} = 2b.
\end{aligned}$$

Therefore,

$$\mathfrak{R}\left(\frac{\mathcal{F}(x_1^n)}{n}\right) \le \frac{32}{\sqrt{n}}\int_0^{2b}\sqrt{\log N(u;\mathcal{F},\|\cdot\|_{\mathbb{P}_n})}\,du.$$

Since $N(u;\mathcal{F},\|\cdot\|_{\mathbb{P}_n}) = 1$ for $u > 2b$, the integral may equivalently be extended to infinity.

2.1 VC Classes

Lemma 2.2 (Empirical covering bound for a VC class)**.** *Suppose that $\mathcal{F}$ is VC-class with VC-dimension ν. There is a universal constant $C > 0$ such that, for $0 < u \le b$,*

$$N(u;\mathcal{F},\|\cdot\|_{P_n}) \le C\nu(16e)^\nu\left(\frac{b}{u}\right)^{2\nu}. \tag{3}$$

Proof. Homework.

Example 2.3 (Rademacher complexity of a bounded VC class)**.** Let $\mathcal{F}$ be b-uniformly bounded with VC-dimension ν. By Example 2.1 and the preceding lemma,

$$\begin{aligned}
\mathfrak{R}\left(\frac{\mathcal{F}(x_1^n)}{n}\right) &\leq \frac{32}{\sqrt{n}}\int_0^{2b}\sqrt{\log N\left(u;\mathcal{F},\|\cdot\|_{\mathbb{P}_n}\right)}\,du \\
&\leq \frac{32}{\sqrt{n}}\int_0^{2b}\sqrt{\log\left[C\nu(16e)^\nu\left(\frac{b}{u}\right)^{2\nu}\right]}\,du \\
&= \frac{32}{\sqrt{n}}\int_0^{2b}\sqrt{\log(C\nu)+\nu\log(16e)+2\nu\log\left(\frac{b}{u}\right)}\,du \\
&\leq \frac{64\sqrt{\nu}}{\sqrt{n}}C_0\left(1+\int_0^{2b}\sqrt{\log\left(\frac{b}{u}\right)}\,du\right) \\
&\leq C_0'\sqrt{\frac{\nu}{n}},
\end{aligned}$$

where C_0' is independent of n and ν (and may depend on the uniform bound b). The integral is finite, so taking expectation over the sample gives

$$\mathfrak{R}_n(\mathcal{F}) := \mathbb{E}_{X_1^n}\mathfrak{R}\left(\frac{\mathcal{F}(X_1^n)}{n}\right) \leq C_0'\sqrt{\frac{\nu}{n}}.$$

Recall from Chapter 4 that polynomial discrimination gave the earlier bound

$$\mathfrak{R}_n(\mathcal{F}) \leq C\sqrt{\frac{\nu\log(n+1)}{n}}.$$

Thus, the entropy-integral argument removes the unnecessary factor $\sqrt{\log(n+1)}$.

One may use this result to Classical Glivenko-Cantelli Theorem; see Example 5.24 [1] for more details.

3 The Chaining Construction

We finish by proving the finite-set version that motivates Dudley's entropy integral. The central idea is to approximate every point by a sequence of representatives at dyadic resolutions and then telescope along this chain.

Theorem 3.1 (Finite-set chaining bound)**.** *Let $\{X_\theta \mid \theta \in T\}$ be a sub-Gaussian process with respect to ρ_X , and define*

$$D := \sup_{\theta,\widetilde{\theta}\in T}\rho_X(\theta,\widetilde{\theta}) < \infty.$$

let $U \in T$ be any subset with $|U| < \infty$, define

$$L := \min\left\{i \geq 1 : N(D2^{-i};T,\rho_X) \geq |U|\right\}.$$

Then

$$\mathbb{E}\left[\sup_{\theta,\widetilde{\theta}\in U}\left|X_\theta - X_{\widetilde{\theta}}\right|\right] \leq 8\sqrt{2}\int_{D2^{-(L+1)}}^{D/2}\sqrt{\log N(u;T,\rho_X)}\,du. \tag{4}$$

Proof. Write $U = \{\theta^1, \dots, \theta^N\}$, where $N = |U|$, and set $\epsilon_m = D2^{-m}$. If $N = 1$, the result is immediate, so assume $N \geq 2$. In particular, $D > 0$.

Since (T, ρ_X) is a metric space and U is finite, the minimum distance between distinct points of U is positive. A ball of radius less than half this minimum contains at most one point of U. Consequently,

$$L = \min\bigl\{m \geq 1 : N(\epsilon_m; T, \rho_X) \geq N\bigr\}$$

is finite.

Construction of the covering sets and chains. For $m = 0, 1, \dots, L-1$, let U_m be a minimal ϵ_m-cover of T, with centers in T, and set $U_L = U$. Then

$$|U_m| = N(\epsilon_m; T, \rho_X) \quad (0 \leq m < L), \qquad |U_L| = |U| \leq N(\epsilon_L; T, \rho_X).$$

For $1 \leq m < L$, the definition of L gives $|U_m| < N$, so all these sets are finite. Since $\epsilon_0 = D$, we may take $U_0 = \{\theta_0\}$ for some $\theta_0 \in T$. Here covering balls are understood to be closed.

For each $m = 0, \dots, L-1$, fix a nearest-point map $\pi_m : T \to U_m$, with ties resolved deterministically, such that $\pi_m(\alpha) \in \arg\min_{\beta \in U_m} \rho_X(\alpha, \beta)$. Because U_m covers T, we have $\rho_X(\alpha, \pi_m(\alpha)) \leq \epsilon_m$ for every $\alpha \in T$.

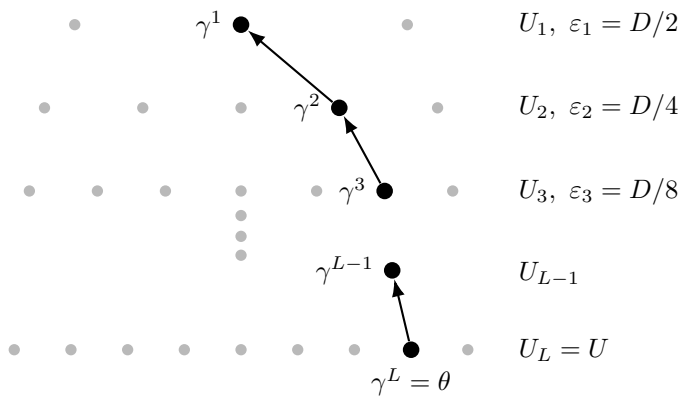


Figure 1: The multiscale covering sets and one backward chain from the target $\theta \in U$ to the coarsest displayed level. Each arrow is an increment controlled at the corresponding scale.

Fix $\theta \in U$, and define its chain backward by

$$\gamma^L = \theta, \qquad \gamma^{m-1} = \pi_{m-1}(\gamma^m), \qquad m = L, L-1, \dots, 1.$$

Thus $\gamma^m \in U_m$, $\gamma^0 = \theta_0$, and $\rho_X(\gamma^m, \gamma^{m-1}) \leq \epsilon_{m-1}$. The increments telescope:

$$X_\theta - X_{\theta_0} = \sum_{m=1}^{L} \bigl(X_{\gamma^m} - X_{\gamma^{m-1}}\bigr).$$

The deterministic chaining inequality. For $m = 1, \dots, L$ and $\gamma \in U_m$, define

$$Z_{m,\gamma} = X_\gamma - X_{\pi_{m-1}(\gamma)}, \qquad R_m = \max_{\gamma \in U_m} Z_{m,\gamma} - \min_{\gamma \in U_m} Z_{m,\gamma}.$$

The recursive definition of the chain gives $Z_{m,\gamma^m} = X_{\gamma^m} - X_{\gamma^{m-1}}$.

For another point $\tilde{\theta} \in U$, construct $\tilde{\gamma}^L, \ldots, \tilde{\gamma}^0$ in the same way. Both chains end at θ_0, so subtracting their telescoping identities yields

$$X_\theta - X_{\tilde{\theta}} = \sum_{m=1}^{L} \big(Z_{m,\gamma^m} - Z_{m,\tilde{\gamma}^m}\big).$$

At each level m, both γ^m and $\tilde{\gamma}^m$ belong to U_m. Therefore, for every realization of the process, both increments lie between $\min_{\gamma \in U_m} Z_{m,\gamma}$ and $\max_{\gamma \in U_m} Z_{m,\gamma}$, and hence

$$\big|Z_{m,\gamma^m} - Z_{m,\tilde{\gamma}^m}\big| \le \max_{\gamma \in U_m} Z_{m,\gamma} - \min_{\gamma \in U_m} Z_{m,\gamma} = R_m.$$

By the triangle inequality,

$$\begin{aligned} |X_\theta - X_{\tilde{\theta}}| &\le \sum_{m=1}^{L} \big|Z_{m,\gamma^m} - Z_{m,\tilde{\gamma}^m}\big| \\ &\le \sum_{m=1}^{L} R_m. \end{aligned}$$

The last expression is independent of $\theta, \tilde{\theta}$. Taking the maximum over these two indices therefore gives the pointwise bound

$$\max_{\theta,\tilde{\theta} \in U} |X_\theta - X_{\tilde{\theta}}| \le \sum_{m=1}^{L} \left(\max_{\gamma \in U_m} Z_{m,\gamma} - \min_{\gamma \in U_m} Z_{m,\gamma} \right).$$

Notice that π_0 is constant, so R_1 is precisely the range of $\{X_\gamma : \gamma \in U_1\}$. The covering sets need not be nested, because the parent maps connect consecutive levels.

Application of the maximal inequality. By the sub-Gaussian increment assumption, for every $\gamma \in U_m$, since

$$\rho_X(\gamma, \pi_{m-1}(\gamma) \le \epsilon_{m-1},$$

$Z_{m,\gamma}$ is sub-Gaussian with parameter ϵ_{m-1}. Apply the known maximal inequality separately to $\{Z_{m,\gamma} : \gamma \in U_m\}$ and $\{-Z_{m,\gamma} : \gamma \in U_m\}$. This gives

$$\begin{aligned} \mathbb{E}R_m &= \mathbb{E} \max_{\gamma \in U_m} Z_{m,\gamma} + \mathbb{E} \max_{\gamma \in U_m} (-Z_{m,\gamma}) \\ &\le 2\epsilon_{m-1} \sqrt{2 \log |U_m|}. \end{aligned}$$

When $|U_m| = 1$, the same inequality holds because $R_m = 0$. Taking expectations in the deterministic chaining inequality yields

$$\mathbb{E} \max_{\theta,\tilde{\theta} \in U} |X_\theta - X_{\tilde{\theta}}| \le 2\sqrt{2} \sum_{m=1}^{L} \epsilon_{m-1} \sqrt{\log |U_m|}.$$

Comparison with the entropy integral. For every $u \in [\epsilon_{m+1}, \epsilon_m]$, monotonicity of covering numbers and the cardinality bounds above imply

$$N(u; T, \rho_X) \ge N(\epsilon_m; T, \rho_X) \ge |U_m|.$$

Since $\epsilon_m - \epsilon_{m+1} = \epsilon_{m-1}/4$, we obtain

$$\begin{aligned} \int_{\epsilon_{m+1}}^{\epsilon_m} \sqrt{\log N(u; T, \rho_X)}\, du &\ge (\epsilon_m - \epsilon_{m+1}) \sqrt{\log |U_m|} \\ &= \frac{\epsilon_{m-1}}{4} \sqrt{\log |U_m|}. \end{aligned}$$

Consequently,

$$\begin{aligned}\mathbb{E}\max_{\theta,\tilde{\theta}\in U}|X_\theta - X_{\tilde{\theta}}| &\leq 8\sqrt{2}\sum_{m=1}^{L}\int_{\epsilon_{m+1}}^{\epsilon_m}\sqrt{\log N(u;T,\rho_X)}\,du \\ &= 8\sqrt{2}\int_{\epsilon_{L+1}}^{\epsilon_1}\sqrt{\log N(u;T,\rho_X)}\,du\end{aligned}$$

□

References

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