Lecture 17: Metric Entropy, Stochastic Processes

Isabella Zhu

10 April 2025

§0.1 Reading

Chapter 5, 5.1-5.3 (Wai).

§1 Setup

A **stochastic process** can be thought of as the set $\{X_t, t \in \Pi\}$ and we assume the process is always centered, i.e. $\mathbb{E}[X_t] = 0$.

Introduce a suitable metric ϕ_x on the plane Π. A δ**-covering** of Π in ϕ_x is a set $\{t^1, \dots t^N\} \subseteq \Pi$ such that $\forall t \in \Pi$, there exists j such that $\phi_x(t, t^j) \leq \delta$. $N(\delta, \Pi, \phi_x)$ is the cardinality of the smallest δ-covering.

Example 1.1

Let $\Pi = \{\theta \in \mathbb{R} | ||\theta|| \leq 1\}$. Then $N(\delta) \approx d \log(1/\delta)$.

We will measure distances in a way that connects to tail behavior of $\{X_t | t \in \Pi\}$.

§1.1 Sub-Gaussian Process

A process $\{X_t\}$ is a ϕ_x**-sub-gaussian process** if $\forall s, t \in \Pi$ and $\lambda \in \mathbb{R}$, we have

$$\mathbb{E}[e^\lambda(X_s - X_t)] \leq e^{\lambda^2 \phi_X(s,t)^2/2}$$

Example 1.2

Let $t \in \Pi \subseteq \mathbb{R}^n$. Let $X_t = \sum_{i=1}^n \epsilon_i t_i$, ϵ_i i.i.d Rademacher. In previous lectures, we have been bounding X_t. This is subgaussian with $\phi_x(s,t)^2 = ||s - t||_2^2$.

Example 1.3

Let $X_t = \sum_{i=1}^n w_i t_i$ where $w \sim N(0, Q)$. Then, we have

$$\text{cov}(X_t - X_s) = (t - s)^T Q (t - s) = \phi_x(s,t)^2$$

§1.2 Gaussian Random Matrices

Let $Z \in \mathbb{R}^{n\times d}$, where $Z_{ij} \sim N(0,1)$. Then, we have

$$\mathbb{E}[||Z||_{op}] = \mathbb{E}[\sup_{||v||_2=1} ||Zv||_2] = \mathbb{E}_2[\sup_{||u||_2=1,||v||_2=1} X_{u,v}]$$

so $t = (u, v)$ and $\Pi = S^{d-1}(1) \times S^{n-1}(1)$ for Euclid sphere in $\mathbb{R}^d$.

Example 1.4

Consider least squares formulation, where $\hat{g} \in \operatorname{argmin}_{g\in G} \frac{1}{n}\sum_{i=1}^n (y_i - g(x_i))^2$ and $g^+ = \operatorname{argmin} \mathbb{E}[(Y - g(X))^2] = \mathbb{E}[Y|X = x]$. We can show

$$||\hat{g} - g^+||_n^2 \le \frac{2}{n}\sum_{i=1}^n w_i(\hat{g}(x_i) - g^+(x_i))$$

where $w_i = y_i - g^+(x_i)$. We'll come back to this, will lead to faster rates.

§2 One-Step Discretization

Proposition 2.1

For any $\tilde{t} \in \Pi_2$, define $D(\tilde{t}) = \sup_{t\in\Pi} \phi_x(t, \tilde{t})$. Assume $\{X_t, t \in \Pi\}$ centered, ϕ_x-subgaussian. Then,

$$\mathbb{E}\left[\sup_{t\in\Pi} X_t\right] \le \mathbb{E}\left[\sup_{\phi_x(s,s')\le\delta} (X_s - X_{\tilde{s}}) + D(\tilde{t})\sqrt{2\log N(\delta;\Pi,\phi_x)}\right]$$

The idea is we choose δ to make the two terms of roughly equal magnitude.

Proof. We have that

$$\mathbb{E}\left[\sup_{t\in\Pi} X_t\right] = \mathbb{E}\left[\sup_{t\in\Pi}(X_t - X_{\tilde{t}})\right]$$

Can find $\{t^1, \dots t^N\}$ such that $\forall t \in \Pi$, there exists t^j with $\phi_x(t, t^j) \le \delta$. We write

$$X_t - X_{\tilde{t}} = (X_t - X_{t^j}) + (X_{t^j} - X_{\tilde{t}}) \le \mathbb{E}\left[\sup_{\phi_x(s,s')\le\delta} (X_s - X_{\tilde{s}}) + \max_{j=1,..N}(X_{t^j} - X_{\tilde{t}})\right]$$

We want to bound $\max_{j=1,..N}(X_{t^j} - X_{\tilde{t}})$. We know that $X_{t^j} - X_{\tilde{t}}$ is sub-Gaussian with parameter $\phi_x(t^j, \tilde{t}) \le D(\tilde{t})$. Therefore, we get

$$\mathbb{E}[\max_{j=1,..N}(X_{t^j} - X_{\tilde{t}})] \le D(\tilde{t})\sqrt{2\log N}$$

§3 Examples

Example 3.1

Prove that $\mathbb{E}[||Z||_{op}||] \le c(\sqrt{d} + \sqrt{n})$ using method from today's lecture.

Proof. We have that

$$X_{u,v} = u^T Z v = \text{trace}(Z^T u v^T)$$

We look at

$$X_{u,v} - X_{\tilde{u},\tilde{v}} = \text{trace}(Z^T(uv^T - \tilde{u}\tilde{v}^T))$$

Our norm is Frobenius norm and we know that $||uv^T - \tilde{u}\tilde{v}^T||_F \leq \delta$. Thus, we have

$$\begin{aligned} X_{u,v} - X_{\tilde{u},\tilde{v}} &= \text{trace}(Z^T(uv^T - \tilde{u}\tilde{v}^T)) \\ &\leq ||Z||_{op}||uv^T - \tilde{u}\tilde{v}^T||_1 \\ &\leq ||Z||_{op}\sqrt{2}||uv^T - \tilde{u}\tilde{v}^T||_F \\ &\leq \sqrt{2}\delta||Z||_{op} \end{aligned}$$

We basically get itself! Choose δ small, move the term over. So, we've shown

$$\mathbb{E}[||Z||_{op}] \leq \sqrt{2}\delta\mathbb{E}[||Z||_{op}] + \text{finitemaximum}$$

We have

$$D(\tilde{u}, \tilde{v}) \leq \max ||uv^T - \tilde{u}\tilde{v}^T||_F \leq \sqrt{2}$$

We bound $N(\delta; \Pi, \phi_x)$ next. We have

$$\phi_x((u,v),(\tilde{u},\tilde{v})) = ||uv^T - \tilde{u}\tilde{v}^T||_F \leq ||u - \tilde{u}||_2 + ||v - \tilde{v}||_2$$

which is good because we've reduced the problem to covering two spaces independently. Consider the cover $(u^{(i)}, v^{(j)})$ where u and v are existing covers. Therefore,

$$\log N(\delta; \Pi, \phi_x) \lesssim (d+n)\log(1/\delta)$$