THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 384 11: Theoretical Statistics

Lecture 16: Uniform Law of Large Numbers- Dudley's chaining Introduction

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

A sub-gaussian process

Definition

A stochastic process $\theta \to X_\theta$ with indexing set T is sub-Gaussian w.r.t a metric d_X if $\forall \theta, \theta' \in T$ and $\lambda \in \mathbb{R}$,

$$E \exp(\lambda(X_\theta - X'_\theta)) \leq \exp\left(\frac{\lambda^2 d_X(\theta, \theta')^2}{2}\right)$$

- This immediately implies the following tail bound.

$$P(|X_\theta - X_{\theta'}| \geq t) \leq 2 \exp\left(-\frac{t^2}{2 d_X(\theta, \theta')^2}\right)$$

Upper bound by 1 step discretization

Theorem

(1-step discretization bound). Let $\{X_\theta, \theta \in \mathcal{T}\}$ be a zero-mean sub-Gaussian process with respect to the metric d_X. Then for any $\delta > 0$, we have

$$E\left[\sup_{\theta,\theta' \in \mathcal{T}} (X_\theta - X_{\theta'})\right] \leq 2E\left[\sup_{\substack{\theta,\theta' \in \mathcal{T} \\ d_X(\theta,\theta') \leq \delta}} (X_\theta - X_{\theta'})\right] + 2D\sqrt{\log N(\delta; \mathcal{T}, d_X)},$$

where $D := \max_{\theta,\theta' \in \Theta} d_X(\theta, \theta')$.

- The mean zero condition gives us: $E[\sup_{\theta \in \mathcal{T}} X_\theta] = E[\sup_{\theta \in \mathcal{T}} (X_\theta - X_{\theta_0})] \leq E[\sup_{\theta,\theta' \in \mathcal{T}} (X_\theta - X_{\theta'})]$

Tradeoff

$$E\left[\sup_{\theta,\theta'\in\mathcal{T}}(X_\theta - X_{\theta'})\right] \leq 2\,\underbrace{E\left[\sup_{\substack{\theta,\theta'\in\mathcal{T}\\ d_X(\theta,\theta')\leq\delta}}(X_\theta - X_{\theta'})\right]}_{\text{Approximation error}} + 4\,\underbrace{\sqrt{D^2 \log N(\delta;\mathcal{T},d_X)}}_{\text{Estimation error}}$$

- As $\delta \to 0$, the cover becomes more refined, and so the approximation error decays to zero.
- But the estimation error grows.
- In practice the δ can be chosen to achieve the optimal trade-off between two terms.

Proof

- Choose a δ cover T.
- For $\theta, \theta' \in \mathcal{T}$, let $\theta^1, \theta^2 \in T$ such that $d_X(\theta, \theta^1) \leq \delta$ and $d_X(\theta', \theta^2) \leq \delta$.

$$\begin{aligned} X_\theta - X_{\theta'} &= (X_\theta - X_{\theta^1}) + (X_{\theta^1} - X_{\theta^2}) + (X_{\theta^2} - X_{\theta'}) \\ &\leq 2 \sup_{\substack{\theta, \theta' \in \mathcal{T} \\ d_X(\theta, \theta') \leq \delta}} (X_\theta - X_{\theta'}) + \sup_{\theta^i, \theta^j \in T} (X_{\theta^1} - X_{\theta^2}) \end{aligned}$$

- But note that $X_{\theta^1} - X_{\theta^2} \sim \mathit{Subgaussian}(d_X(\theta^1, \theta^2))$..

Finite class lemma for subgaussian processes

Theorem

Consider X_θ sub-gaussian w.r.t d on $\mathcal{T}$ and A is a set of pairs from $\mathcal{T}$.

$$E \max_{(\theta,\theta')\in A} (X_\theta - X_{\theta'}) \leq D\sqrt{2\log|A|},$$

where $D := \max_{(\theta,\theta')\in A} d_X(\theta,\theta')$.

Finite class lemma

$$
\begin{aligned}
\exp\left(\lambda E \max_{(\theta,\theta')\in A}(X_\theta - X_{\theta'})\right) &\leq E\exp\left(\lambda \max_{(\theta,\theta')\in A}(X_\theta - X_{\theta'})\right) \\
&= \max_{(\theta,\theta')\in A} E\exp(\lambda(X_\theta - X_{\theta'})) \\
&\leq \sum_{(\theta,\theta')\in A} \exp\left(\frac{\lambda^2 d_X(\theta,\theta')^2}{2}\right) \\
&\leq |A| \exp\left(\frac{\lambda^2 D^2}{2}\right)
\end{aligned}
$$

- Now optimize over λ.

Finishing the proof

$$X_\theta - X_{\theta'} \le 2 \sup_{\substack{\theta,\theta' \in \mathcal{T} \\ d_X(\theta,\theta') \le \delta}} (X_\theta - X_{\theta'}) + \sup_{\theta^i,\theta^j \in T} (X_{\theta 1} - X_{\theta 2})$$

$$E\left[\sup_{\theta,\theta' \in \mathcal{T}} (X_\theta - X_{\theta'})\right] \le 2E\left[\sup_{\substack{\theta,\theta' \in \mathcal{T} \\ d_X(\theta,\theta') \le \delta}} (X_\theta - X_{\theta'})\right] + E\left[\sup_{\theta^i,\theta^j \in T} (X_{\theta 1} - X_{\theta 2})\right]$$

$$\le 2E\left[\sup_{\substack{\theta,\theta' \in \mathcal{T} \\ d_X(\theta,\theta') \le \delta}} (X_\theta - X_{\theta'})\right] + D\sqrt{2 \log N(\delta; \mathcal{T}, d_X)^2}$$

Examples: smoothly parametrized class

Example

Suppose $\mathcal{F}$ is a class parametric functions $\mathcal{F} := \{f(\theta, .) : \theta \in B_2\}$, where B_2 is the unit L_2 ball in $\mathbb{R}^d$. Assume that $\mathcal{F}$ is closed under negation. f is L Lipschitz w.r.t. the Euclidean distance on Θ, i.e. $|f(\theta, .) - f(\theta', .)| \leq L\|\theta - \theta'\|_2$.

$$\mathcal{R}_n(\mathcal{F}) = O\left(L\sqrt{\frac{d\log(Ln)}{n}}\right)$$

Proof

- Denote $f(\theta, X_1^n)$ as the vector $(f(\theta, X_1), \dots, f(\theta, X_n))$.
- Recall that $n\mathcal{R}_n(\mathcal{F}) = E\left[\sup_{f \in \mathcal{F}} \langle \epsilon, f(\theta, X_1^n) \rangle\right] = E\left[\sup_{\theta \in \Theta} \langle \epsilon, f(\theta, X_1^n) \rangle\right]$
- The process $f(\theta, X_1^n) \to \langle \epsilon, f(\theta, X_1^n) \rangle =: Y_\theta$ is mean zero subgaussian.
- Note that $Y_\theta - Y_\theta' \sim Subgaussian(d_X(\theta, \theta'))$
- We have:

$$d_X(\theta, \theta') = \|f(\theta, X_1^n) - f(\theta', X_1^n)\|^2 \leq nL^2 \|\theta - \theta'\|_2^2$$

  - So it is $L\sqrt{n}$ Lipschitz.

Proof

- Also,

$$n\mathcal{R}_n(\mathcal{F}) = E[\sup_{\theta\in\Theta}(Y_\theta - Y_{\theta'})] \le E[\sup_{\theta,\theta'\in\Theta}(Y_\theta - Y_{\theta'})]$$

-

$$n\mathcal{R}_n(\mathcal{F}) \le \underbrace{E \sup_{\substack{\|\theta-\theta'\|_2\le\delta \\ \theta,\theta'\in\Theta}} (Y_\theta - Y'_\theta)}_{A} + 2D\sqrt{\log N(\delta;\mathcal{F},d_X)}$$

- $A \le \delta L\sqrt{n}E\left[\sup_{\|v\|_2=1}\langle\epsilon, v\rangle\right] \le \delta Ln$
- $D = \sup_{\theta,\theta'} d_X(\theta,\theta) = 2L\sqrt{n}$

Proof

- $N(\delta; \mathcal{F}, d_X) \leq N(\delta/L\sqrt{n}, \Theta, \|.\|_2) \leq \left(1 + \frac{L\sqrt{n}}{\delta}\right)^d$
- Finally,

$$\mathcal{R}_n(\mathcal{F}) \leq 2L\delta + 4L\sqrt{\frac{d\log(1 + L\sqrt{n}/\delta)}{n}}$$

- Setting $\delta = 1/\sqrt{n}$ gives:

$$\mathcal{R}_n(\mathcal{F}) \leq \frac{2L}{\sqrt{n}} + 4L\sqrt{\frac{d\log(1 + Ln)}{n}}$$