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

SDS 384 11: Theoretical Statistics

Lecture 15: Uniform Law of Large Numbers- Dudley's chaining Intro

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 if $\forall \theta, \theta' \in T$ and $\lambda \in \mathbb{R}$,

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

- This immediately implies the following tail bound.

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

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. Then for any $\delta > 0$, we have

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

- 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

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

- 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 ϵ 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 $d(\theta, \theta^1) \le \delta$ and $d(\theta', \theta^2) \le \delta$.

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

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

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 $\mathcal{T}$.

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

Finite class lemma

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

- Now optimize over λ.

Finishing the proof

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

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

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