36-710: Advanced Statistical Theory **Fall 2018**

Lecture 8: September 26

Lecturer: Alessandro Rinaldo *Scribe: Ron Yurko*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

8.1 Maximal Inequalities

Let $\{X_i, i \in \mathcal{I}\}$ be a Stochastic Process. We want to obtain bounds for,

$$\mathbb{E}\big[\sup_{i\in\mathcal{I}} X_i\big],$$
$$\mathbb{P}\big(\sup_{i\in\mathcal{I}} X_i \geq t\big),\ t \geq 0$$

You may want to replace X_i with $|X_i|$. If $|\mathcal{I}|$ is finite, union bound will typically work!

Last time: $(\mathcal{X}, d)$ is a metric space, d is a distance.

Definition 8.1 *(Covering) Let $\delta > 0$, a δ-cover of $(\mathcal{X}, d)$ is a subset $\{\theta_1, \ldots, \theta_N\} \subset \mathcal{X}$ such that $\forall \varkappa \in \mathcal{X}$ $\exists \theta_i = \theta_i(x)$ such that $d(\varkappa, \theta_i) \leq \delta$.*

The δ-covering number of $(\mathcal{X}, d)$ is the cardinality of a minimal δ-cover. It is denoted by $N(\delta, \varkappa, d)$. Think of it as the number of closed ball you can put around your space to cover it.

Remark

$$\mathcal{X} \subset \bigcup_{i=1}^{N} B(\theta_i, \delta) = \{\varkappa \in \mathcal{X} : d(x, \theta_i) \leq \delta\}$$

We will only consider spaces that are totally bounded : $\forall \delta > 0, N(\delta, \varkappa, d) < \infty$ (in $\mathbb{R}^d$ equivalent to $\mathcal{X}$ being bounded).

Remark $(\delta, \varkappa, d)$ is decreasing in δ and as $\delta \to 0$, it will typically diverge.

Definition 8.2 *$\delta \in \mathbb{R}_+ \to logN(\delta, \varkappa, d)$ is also known as metric entropy of $(\varkappa, d)$.*

Examples: $\varkappa = [-1, 1],\ d(x, y) = |x - y|$, then

$$N(\delta, \varkappa, d) \leq \frac{1}{\delta} + 1,$$

if $\varkappa = [-1, 1]^p,\ d = ||\cdot||_\infty,\ N(\delta, \varkappa, d) \leq (\frac{1}{\delta} + 1)^p$.

In p-dimensional Euclidean spaces, the metric entropy is of the order $p\log(\frac{1}{\delta})$.

For infinite dimensional spaces, the metric entropy has a much worse dependence on δ, let $\mathcal{F} = \{f : [0,1]^p \to \mathbb{R}, \text{L - Lipschitz}\}$ $(|f(x) - f(y)| \leq \text{L}||x - y||$.

$$\log N(\delta, \mathcal{F}, ||\cdot||_\infty) \asymp (\frac{\text{L}}{\delta})^p$$

where $||f - g||_\infty = \sup_{x\in[0,1]^p} |f(x) - g(x)|$ and $f(0) = 0$. If $\mathcal{F}$ consists of L-Lipschitz but also "smooth", where smoothness is controlled by $\alpha > 0$ ($\alpha \uparrow$ means smoother functions),

$$\log(\delta, \mathcal{F}, ||\cdot||_\infty) \asymp (\frac{\text{L}}{\delta})^{p/\alpha}$$

Definition 8.3 *(Packing number) $\delta > 0$, a δ-packing of $(\mathcal{X}, d)$ is a subset of $\{\theta_1, \ldots, \theta_M\} \subset \mathcal{X}$ s.t. $d(\theta_i, \theta_j) > \delta$ $\forall i \neq j$ notice that $M = M(\delta)$. The δ-packing number of $(\mathcal{X}, d)$ is the cardinality of a largest δ-packing*

Lemma 8.4 $\forall\delta > 0,\ M(2\delta, \varkappa, d) \leq N(\delta, \varkappa, d) \leq M(\delta, \varkappa, d)$.

8.2 Covering of Euclidean Spaces

$\mathcal{X} \subsetneq \mathbb{R}^d,\ d = ||\cdot||_p,\ p \geq 1.$

Lemma 8.5 *Let $||\cdot||$ and $||\cdot||'$ be two norms in $\mathbb{R}^d$, let B and B' be corresponding unit balls $[B = \{\in \mathbb{R}^d : ||x|| \leq 1\}]$. Aside: Let K be closed convex symmetric subset of $\mathbb{R}^p$, let $p_k(x) = inf\{t > 0 : x \in tk\}$. Then $p_k(x)$ is a norm in $\mathbb{R}^p$!*

Then

$$(\frac{1}{\delta})^p \frac{Vol(B)}{Vol(B')} \leq N(\delta, B, ||\cdot||') \leq \frac{Vol(\frac{2}{\delta} + B')}{Vol(B')}$$

$\to Vol(B) =$ Lebesgue measure of B, where $B + B' = \{x + x' : x \in B, x' \in B'\}$ (Minkowski sum).

Proof. Use the fact that $Vol(\delta B) = \delta^d Vol(B)$. If $\{\theta_1, \ldots, \theta_N\}$ is a δ-cover of B in $||\cdot||'$. Then $B \subset \bigcup_i^N(\theta_i + \delta B') \to \{x : ||x - \theta_i||' \leq \delta\}$.

So,

$$\begin{aligned} Vol(B) &\leq \sum_i^N Vol(\theta_i + \delta B') \\ &= N\delta^p Vol(B') \\ \Rightarrow N(\delta, B, ||\cdot||') &\geq (1/\delta)^p \frac{Vol(B)}{Vol(B')} \end{aligned}$$

To show upper bound, let $\theta_1, \ldots, \theta_M$ be a maximal δ-packing of B in $||\cdot||'$. Then $\{\theta_1, \ldots, \theta_m\}$ is also a δ-cover of B in $||\cdot||'$, [if not, $\in B$ *s.t.* $x + \delta B'$ would not contain any θ_i, then $\{\theta_1, \ldots, \theta_m, x\}$ would also be a δ-packing. Contradiction because is is largest number of packing.]

The balls $\theta_i + \frac{\delta}{2}B'$ are disjoint $i = 1, \ldots, M$ and $\bigcup_i^M(\theta_i + \delta/2B') \subset B + \delta/2B'$. Same proof gives, if $K \subseteq \mathbb{R}^p$,

$$\frac{Vol(K)}{\delta^p V_p} \leq N(\delta, K, ||\cdot||_2) \leq \frac{Vol(K+\delta/2B)}{(\delta/2)^p V_p}$$

Where V_p = volume of unit euclidean ball, $B_p = \{x : ||x||_2 \leq 1\}$.

Corollary 8.6 *If* $||\cdot|| = ||\cdot||$,

1. $logN(\delta, B, ||\cdot||) \leq plog(1+2/\delta) \leq plog(3/\delta)$
2. *If* $K = S^{p-1} = \{x \in \mathbb{R}^p : ||x||_2 = 1\}$
   $N(\delta, S^{p-1}, ||\cdot||_2) \leq (1+2/\delta)^p$

8.3 Application to estimating the Euclidean Norm of Random Vector

Definition 8.7 *A vector* $X \in \mathbb{R}^d$ *is said to be* $SG(\sigma^2)$ *when* $v^T X \in SG(\sigma^2)$ *for each* $v \in S^{d-1}$ *(unit sphere). If coordinates of* X *are independent* $SG(\sigma^2)$ *this is true.*

Theorem 8.8 *Let* X *be a random vector* $X \in \mathbb{R}^d$ *be* $SG_d(\sigma^2)$ *then* $\mathbb{E}[||X||] \leq 4\sigma\sqrt{d}$ *and* $||X|| \leq 4\sigma\sqrt{d} + 2\sigma\sqrt{log(1/\delta)}$ *with probability* $\geq 1-\delta,\ \delta \in (0,1)$.

Proof. Use variational characterization of $||X||$: $||X|| = \max_{v\in B_d} v^T X$. Then $||X||$ is the max of RVs indexed by B_d. $||X|| = \max_{v\in B_d} Y_v$. Let $\mathcal{N}_{1/2}$ be a 1/2 cover B_d in Euclidean norm $\Rightarrow |\mathcal{N}_{1/2}| \leq 5^d$. Next $\forall v \in B_d, \exists z = z(v) \in \mathcal{N}_{1/2}$ *s.t.* $v = z + w, w \in 1/2B_d$. So,

$$\max_{v\in B_d} v^T X \leq \max_{z\in\mathcal{N}_{1/2}} z^T X + \underset{w\in 1/2B_d}{w^T X}$$
$$\Rightarrow 1/2 \max_{v\in B_d} v^T X \leq 2 \max_{z\in\mathcal{N}_{1/2}} z^T X$$

Use same argument for other way.

<u>Remark</u>
For $\epsilon \in (0,1)$ the same argument gives you

$$\underset{v\in B_d}{v^T X} \leq \frac{1}{1-\epsilon} \max_{z\in\mathcal{N}_\epsilon} z^T X$$

where $\epsilon-$cover of B_d in $||\cdot||_2$.