Maximal inequalities, covering nets, norms of random matrices

Setting: Random process $\{X_t\}_{t \in T}$. A few examples:

- $\{g^T t\}_{t \in T}$, $g \sim N(0, I)$ a Gaussian vector

- $\{u^T X v\}_{\substack{u \in \mathbb{R}^n : \|u\|_2 \le 1 \\ v \in \mathbb{R}^m : \|v\|_2 \le 1}}$, $X \in \mathbb{R}^{n \times m}$ a random matrix

- $\left\{\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}f(X_i)\right\}_{f \in \mathcal{F}}$, $\mathcal{F}$ some function class,

  $X_1, \ldots, X_n$ iid r.v.'s.

Goal: (1) Sharp upper bounds for $\mathbb{E} \sup_{t \in T} X_t$

(2) Sharp tail bounds for $\mathbb{P}\left[\sup_{t \in T} X_t \ge u\right]$

These questions are often equivalent if $\sup_{t \in T} X_t$ concentrates around its mean.

Principle: If $\{X_t\}_{t \in T}$ "sufficiently continuous," then size of $\sup_{t \in T} X_t$ is controlled by "complexity" of T.

Basic idea: If T has finite cardinality and $X_t \ge 0 \ \forall t \in T$, then $\mathbb{E} \sup_{t \in T} X_t \le \sum_{t \in T} \mathbb{E} X_t$.

Example: For any r.v.'s $X_1, \dots, X_n$:

$$\mathbb{E} \sup_{i=1}^{n} X_i \leq \sum_{i=1}^{n} \mathbb{E}|X_i| \leq n \cdot \sup_{i=1}^{n} \mathbb{E}|X_i|$$

$$\mathbb{E} \sup_{i=1}^{n} X_i \overset{\text{(Jensen's)}}{\leq} \left(\mathbb{E} \sup_{i=1}^{n} |X_i|^p\right)^{1/p} \leq n^{1/p} \sup_{i=1}^{n} \left(\mathbb{E}|X_i|^p\right)^{1/p}$$

Lemma (maximal inequality): Suppose $|T|$ is finite and $\log \mathbb{E} e^{\lambda X_t} \leq \psi(\lambda)$ for all $\lambda \geq 0$, $t \in T$, where ψ is convex, $\psi(0) = \psi'(0) = 0$. Set $\psi^*(t) = \sup_{\lambda \geq 0} \lambda t - \psi(\lambda)$.

Then

$$\mathbb{E} \sup_{t \in T} X_t \leq \psi^{*-1}(\log |T|).$$

Cor: If $\mathbb{E} e^{\lambda X_t} \leq \frac{\lambda^2 \sigma^2}{2}$ for all $t \in T$ and $\lambda \geq 0$, then

$$\mathbb{E} \sup_{t \in T} X_t \leq \sqrt{2\sigma^2 \log |T|}$$

(Apply Lemma w/ $\psi(\lambda) = \frac{\lambda^2 \sigma^2}{2}$, $\psi^*(t) = \frac{t_+^2}{2\sigma^2}$.)

Proof: $$\mathbb{E} \sup_{t \in T} X_t \overset{\text{(Jensen's)}}{\leq} \frac{1}{\lambda} \log \mathbb{E} e^{\lambda \sup_{t \in T} X_t} \quad \forall \lambda > 0$$

$$= \frac{1}{\lambda} \log \mathbb{E} \sup_{t \in T} e^{\lambda X_t}$$

$$\leq \frac{1}{\lambda} \log \sum_{t \in T} \mathbb{E}\, e^{\lambda X_t}$$

$$\leq \frac{1}{\lambda} \log \left(|T| \cdot \sup_{t \in T} \mathbb{E} e^{\lambda X_t} \right) \leq \frac{\log |T| + \psi(\lambda)}{\lambda}.$$

$$\Rightarrow \log |T| \geq \sup_{\lambda \geq 0} \left(\lambda \mathbb{E} \sup_{t \in T} X_t - \psi(\lambda) \right) = \psi^*\left(\mathbb{E} \sup_{t \in T} X_t \right).$$

$$\Rightarrow \mathbb{E} \sup_{t \in T} X_t \leq \psi^{*-1}(\log |T|). \quad (\psi^* \text{ convex, increasing on } [0,\infty).) \blacksquare$$

Remark: This is closely related to the <u>union bound</u>

$$\mathbb{P}\left[\sup_{t \in T} X_t \geq u \right] \leq \sum_{t \in T} \mathbb{P}[X_t \geq u] \leq |T| \cdot e^{-\lambda u + \psi(\lambda)} \quad \forall \lambda \geq 0$$

$$\Rightarrow \mathbb{P}\left[\sup_{t \in T} X_t \geq u \right] \leq e^{\log |T| - \psi^*(u)}$$

$$\Rightarrow \mathbb{P}\left[\sup_{t \in T} X_t \geq \psi^{*-1}(\log |T| + s) \right] \leq e^{-s}$$

Lemma can be proven (up to a constant) by integrating this tail.

Intuition: If $\{X_t\}_{t \in T}$ are independent and $\sum_{t \in T} \mathbb{P}[X_t \geq u]$ is small,

$$\mathbb{P}\left[\sup_{t \in T} X_t \geq u \right] = 1 - \prod_{t \in T} \left(1 - \mathbb{P}[X_t \geq u] \right)$$

$$\geq 1 - \prod_{t \in T} e^{-\mathbb{P}[X_t \geq u]} \approx \sum_{t \in T} \mathbb{P}[X_t \geq u]$$

so union bound is tight. Bound is loose when $\{X_t\}_{t \in T}$ are dependent.

Covering nets

Idea: Trade off dependence of $\{X_t\}_{t\in T}$ with the continuity

$$X_t \approx X_s \text{ for } t \approx s.$$

Def: Let (T,d) be a metric space. $\mathcal{N} \subseteq T$ is an <u>ε-net</u> of T if, for all $t\in T$, there exists $\pi(t) \in \mathcal{N}$ s.t. $d(t,\pi(t)) \le \varepsilon$. The <u>covering number</u> is

$$N(T,d,\varepsilon) = \inf\{|\mathcal{N}| : \mathcal{N} \text{ is } \varepsilon\text{-net of } T\}.$$

Prop: If $\{X_t\}_{t\in T}$ is L-Lipschitz, i.e. there exists (possibly random) $L \ge 0$ for which $|X_t - X_s| \le L \cdot d(s,t)$ $\forall s,t\in T$, and $\log \mathbb{E} e^{\lambda X_t} \le \frac{\lambda^2\sigma^2}{2}$ $\forall \lambda \ge 0$ and $t\in T$, then

$$\mathbb{E} \sup_{t\in T} X_t \le \inf_{\varepsilon>0} \varepsilon \cdot \mathbb{E} L + \sqrt{2\sigma^2 \log N(T,d,\varepsilon)}$$

Proof: For any ε-net $\mathcal{N}$,

$$\sup_{t\in T} X_t = \sup_{t\in T} (X_t - X_{\pi(t)}) + X_{\pi(t)}$$

$$\le L \cdot d(t,\pi(t)) + \sup_{s\in\mathcal{N}} X_s.$$

Thus $\mathbb{E} \sup_{t \in T} X_t \le \varepsilon \cdot \mathbb{E} L + \sqrt{2\sigma^2 \log |\mathcal{N}|}$

Take inf over $\varepsilon > 0$ and all ε-nets $\mathcal{N}$. $\blacksquare$

Remark: Sometimes $T \subseteq S$ a larger space, and it's simpler to construct an ε-net of T with points in S. Let

$$N^{ext}(T, d, \varepsilon) = \inf \{ |\mathcal{N}| : \mathcal{N} \subseteq S \text{ is } \varepsilon\text{-net of } T \}.$$

Then $N^{ext}(T, d, \varepsilon) \le N(T, d, \varepsilon) \le N^{ext}(T, d, \varepsilon/2)$

(See Homework 7.)

Example: $X_1, \dots, X_n$ iid r.v.'s on $[0,1]$,

$\mathcal{F} = \{ f : [0,1] \to [0,1], \text{ 1-Lipschitz} \}$.

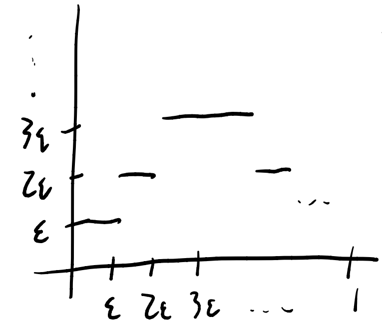
Let $X_f = \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E} f(X_i)$, $W = \sup_{f \in \mathcal{F}} X_f$

(1-Wasserstein distance between $\frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and law of X_i)

- $|X_f - X_g| \le \frac{1}{n} \sum_{i=1}^n | f(X_i) - g(X_i) - \mathbb{E} f(X_i) + \mathbb{E} g(X_i) | \le 2 \| f - g \|_\infty$.
- Each $f(X_i) - \mathbb{E} f(X_i) \in [-\mathbb{E} f(X_i), 1 - \mathbb{E} f(X_i)]$, so

$\log \mathbb{E} e^{\lambda X_f} \le \frac{\lambda^2}{8n}$ $\forall \lambda \ge 0$ by Hoeffding's lemma + Hoeffding's inequality

- Let $\mathcal{N} = \{$piecewise constant functions on $[0,1]$ w/ jumps at $\varepsilon, 2\varepsilon, 3\varepsilon, \ldots$, values in $0, \varepsilon, 2\varepsilon, \ldots$, where value changes by at most ε at each jump$\}$



For any $f \in \mathcal{F}$: If $f(k\varepsilon) \in [j\varepsilon, (j+1)\varepsilon]$, then $\forall x \in [k\varepsilon, (k+1)\varepsilon]$,

$$|f(x) - j\varepsilon| \leq |f(x) - f(k\varepsilon)| + |f(k\varepsilon) - j\varepsilon| \leq 2\varepsilon.$$

Also, $f((k+1)\varepsilon) \in [(j-1)\varepsilon, j\varepsilon] \cup [j\varepsilon, (j+1)\varepsilon] \cup [(j+1)\varepsilon, (j+2)\varepsilon]$

$\Rightarrow \exists g \in \mathcal{N}$ (round value of f at left-endpoint of each interval down to nearest $j\varepsilon$) s.t. $\|f-g\|_\infty \leq 2\varepsilon$.

$|\mathcal{N}| \leq \left(\frac{1}{\varepsilon}\right) \cdot 3^{1/\varepsilon} \Rightarrow N(\mathcal{F}, \|\cdot\|_\infty, \varepsilon) \leq C^{1/\varepsilon} \; \forall \varepsilon \in (0,1]$.

So $\mathbb{E} W \leq \inf_{\varepsilon > 0} \left(2\varepsilon + \sqrt{\frac{1}{2n} \log C^{1/\varepsilon}}\right) \leq C' n^{-1/3}$.

In later lectures, we'll:

(1) Improve this bound to $Cn^{-1/2}$, using the typical size

$$|X_f - X_g| \lesssim \frac{1}{\sqrt{n}} \|f-g\|_\infty \text{ w.h.p.}$$

(2) Extend to non-Lipschitz classes $\mathcal{F}$ for which $N(\mathcal{F}, \|\cdot\|_\infty, \varepsilon)$ is infinite, by using other metrics for $\mathcal{F}$.

Norm of random matrices

Let $X \in \mathbb{R}^{n \times m}$. $\|X\|_{op} = \sup_{u \in B^n, v \in B^m} \underbrace{u^T X v}_{:= X_{u,v}}$,

$$B^n := \{u \in \mathbb{R}^n : \|u\|_2 \leq 1\}.$$

Then for any $u, u' \in B^n$, $v, v' \in B^m$,

$$\begin{aligned} |X_{u,v} - X_{u',v'}| &= |u^T X v - u'^T X v'| \\ &\leq |u^T X (v - v')| + |(u-u')^T X v'| \\ &\leq \|X\|_{op} \cdot (\|v - v'\|_2 + \|u - u'\|_2) \end{aligned}$$

$\Rightarrow \{X_{u,v}\}_{u,v}$ is $\|X\|_{op}$-Lipschitz w.r.t. the metric

$$d((u,v),(u',v')) = \|u-u'\|_2 + \|v-v'\|_2.$$

To bound $N(B^n \times B^m, d, \varepsilon)$:

Def: Let (T,d) be a metric space. $\mathcal{D} \subseteq T$ is an <u>ε-packing</u> of T if $d(s,t) > \varepsilon$ $\forall s \neq t \in \mathcal{D}$.

The <u>packing number</u> is

$$D(T, d, \varepsilon) = \sup\{|\mathcal{D}| : \mathcal{D} \text{ is } \varepsilon\text{-packing of } T\}.$$

Prop: For any $\varepsilon > 0$,

$$D(T, d, 2\varepsilon) \leq N(T, d, \varepsilon) \leq D(T, d, \varepsilon).$$

Proof: Let $\mathcal{D}$ be any 2ε-packing, $\mathcal{N}$ any ε-net.

For any $t \in \mathcal{D}$, $\exists \pi(t) \in \mathcal{N}$ s.t. $d(t, \pi(t)) \leq \varepsilon$.

Furthermore $\pi(t) \neq \pi(s)$ for $t \neq s \in \mathcal{D}$, because

$$d(\pi(t), \pi(s)) \geq d(t,s) - d(t, \pi(t)) - d(s, \pi(s)) > 0.$$

So $|\mathcal{D}| \leq |\mathcal{N}|$, implying $D(T, d, 2\varepsilon) \leq N(T, d, \varepsilon)$.

Let $\mathcal{D}$ be any <u>maximal</u> ε-packing, i.e. $\mathcal{D} \cup \{t\}$ is not a ε-packing for any $t \in T \setminus \mathcal{D}$. Then for any $t \in T \setminus \mathcal{D}$, $\exists s \in \mathcal{D}$ s.t. $d(t,s) \leq \varepsilon$. So $\mathcal{D}$ is also a ε-net, implying $N(T, d, \varepsilon) \leq |\mathcal{D}| \leq D(T, d, \varepsilon)$. ∎

Prop: Let $B^n = \{u \in \mathbb{R}^n : \|u\|_2 \leq 1\}$. Then for any $\varepsilon \in (0,1)$,

$$\left(\frac{1}{\varepsilon}\right)^n \leq N(B^n, \|\cdot\|_2, \varepsilon) \leq \left(\frac{2}{\varepsilon} + 1\right)^n < \left(\frac{3}{\varepsilon}\right)^n$$

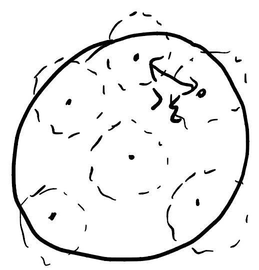
Proof: Let $\mathrm{Vol}(\cdot)$ be the volume in $\mathbb{R}^n$. If $\mathcal{N}$ is an ε-net, then $B^n \equiv B(0,1) \subseteq \bigcup_{t \in \mathcal{N}} B(t, \varepsilon)$

$$\Rightarrow \mathrm{Vol}(B(0,1)) \le \sum_{t \in \mathcal{N}} \mathrm{Vol}(B(t,\varepsilon))$$

$$= |\mathcal{N}| \cdot \varepsilon^n \cdot \mathrm{Vol}(B(0,1))$$

$\Rightarrow |\mathcal{N}| \ge \left(\frac{1}{\varepsilon}\right)^n$, implying $N(B^n, \|\cdot\|_2, \varepsilon) \ge \left(\frac{1}{\varepsilon}\right)^n$.

If $\mathcal{D}$ is an ε-packing, then $\{B(t, \frac{\varepsilon}{2}) : t \in \mathcal{D}\}$ are disjoint balls contained in $B(0, 1+\frac{\varepsilon}{2})$.



$$\Rightarrow \mathrm{Vol}(B(0, 1+\tfrac{\varepsilon}{2})) \ge \sum_{t \in \mathcal{D}} \mathrm{Vol}(B(t, \tfrac{\varepsilon}{2}))$$

$$\Rightarrow (1+\tfrac{\varepsilon}{2})^n \cdot \mathrm{Vol}(B(0,1)) \ge |\mathcal{D}| \left(\tfrac{\varepsilon}{2}\right)^n \cdot \mathrm{Vol}(B(0,1))$$

$$\Rightarrow |\mathcal{D}| \le \left(\tfrac{2}{\varepsilon} + 1\right)^n$$

Thus $N(B, \|\cdot\|_2, \varepsilon) \le D(B, \|\cdot\|_2, \varepsilon) \le \left(\frac{2}{\varepsilon}+1\right)^n$. $\blacksquare$

Thm: Suppose $X \in \mathbb{R}^{n \times m}$ has independent mean-zero σ^2-subgaussian entries.

Then $\mathbb{E}\|X\|_{op} \le C\sigma(\sqrt{n} + \sqrt{m})$ for a universal constant $C > 0$.

Proof: Recall $\|X\|_{op} = \sup_{u \in B^n, v \in B^m} X_{u,v} = \sup_{u \in B^n, v \in B^m} u^T X v$

where $X_{u,v}$ is $\|X\|_{op}$-Lipschitz w.r.t. $d((u,v),(u',v')) = \|u-u'\|_2 + \|v-v'\|_2$.

If $\mathcal{N}_u$ is an ε-net of B^n, $\mathcal{N}_v$ an ε-net of B^m

(in $\|\cdot\|_2$), then $\mathcal{N}_u \times \mathcal{N}_v$ is 2ε-net of $B^n \times B^m$ in d.

$$\Rightarrow N(B^n \times B^m, d, \varepsilon) \leq N(B^n, \|\cdot\|_2, \tfrac{\varepsilon}{2}) \cdot N(B^m, \|\cdot\|_2, \tfrac{\varepsilon}{2})$$
$$\leq \left(\tfrac{4}{\varepsilon}+1\right)^{n+m}$$

For each $(u,v) \in B^n \times B^m$, $u^T X v$ is σ^2-subgaussian by Hoeffding's inequality. So

$$\mathbb{E}\|X\|_{op} \leq \varepsilon \cdot \underbrace{\mathbb{E}\|X\|_{op}}_{\text{Lipschitz constant of } \{X_{u,v}\}} + \sqrt{2\sigma^2 \cdot \log \left(\tfrac{4}{\varepsilon}+1\right)^{n+m}} \quad \forall \varepsilon \geq 0.$$

Pick $\varepsilon = 1/2$ and rearrange: $\mathbb{E}\|X\|_{op} \leq C\sigma(\sqrt{n}+\sqrt{m})$. ∎

Remark: $\|X\|_{op}$ is a convex, 1-Lipschitz function of its entries. If the entries are bounded $X_{ij} \in [-C\sigma, C\sigma]$, or satisfy a LSI $\operatorname{Ent} f(X_{ij})^2 \leq C\sigma^2\, \mathbb{E} f'(X_{ij})^2$, then by concentration

$$\mathbb{P}\left[\|X\|_{op} \geq \mathbb{E}\|X\|_{op} + \sigma \cdot t\right] \leq e^{-ct^2}.$$

Alternatively, assuming only X_{ij} are σ^2-subgaussian, the above argument may be adapted into a union bound:

$$\mathbb{P}\left[\|X\|_{op} \geq \varepsilon \cdot \|X\|_{op} + C\sigma(\sqrt{n}+\sqrt{m}) + \sigma \cdot t\right]$$

$$\leq \mathbb{P}\left[\exists (u,v) \in \mathcal{N}_u \times \mathcal{N}_v : X_{u,v} \geq C\sigma(\sqrt{n}+\sqrt{m}) + \sigma \cdot t\right]$$

$$\leq \left(\frac{4}{\varepsilon}+1\right)^{n+m} \cdot e^{-\frac{1}{2}\left(C(\sqrt{n}+\sqrt{m})+t\right)^2}$$

$$\leq \left(\frac{4}{\varepsilon}+1\right)^{n+m} \cdot e^{-\frac{C^2}{2}(n+m) - \frac{t^2}{2}} \quad \forall t \geq 0.$$

Take $\varepsilon = 1/2$, $C^2/2 = \log 9$

$$\Rightarrow \mathbb{P}\left[\|X\|_{op} \geq 2C\sigma(\sqrt{n}+\sqrt{m}) + 2\sigma \cdot t\right] \leq e^{-t^2/2}.$$

Thm: Let $X \in \mathbb{R}^{n \times m}$ have independent rows $x_1, \ldots, x_n \in \mathbb{R}^m$ s.t. $\mathbb{E} x_i = 0$, $\mathbb{E} x_i x_i^T = I$, $u^T x_i$ is σ^2-subgaussian for every (fixed) unit vector $u \in \mathbb{R}^m$. Let $s_{min}(X)$ and $s_{max}(X) \equiv \|X\|_{op}$ be the smallest/largest singular values. Then for a universal constant $C > 0$,

$$\mathbb{P}\left[\sqrt{n} - C\sigma^2(\sqrt{m}+t) \leq s_{min}(X) \leq s_{max}(X) \leq \sqrt{n} + C\sigma^2(\sqrt{m}+t)\right]$$
$$\geq 1 - 2e^{-t^2}.$$

Interpretation: With probability $1-e^{-cm}$, $\forall u \in \mathbb{R}^m$ w/ $\|u\|_2 = 1$: $\|Xu\|_2 \in \sqrt{n} \pm C\sqrt{m}$, i.e. $\frac{1}{\sqrt{n}} X$ is near isometry if $m \ll n$.

Proof: Consider $\|\frac{1}{n}X^TX - I\|_{op} = \sup_{\substack{u \in \mathbb{R}^m: \\ \|u\|_2 = 1}} |u^T(\frac{1}{n}X^TX - I)u|$

$$= \sup_{\substack{u \in \mathbb{R}^m: \\ \|u\|_2 = 1}} \underbrace{\left|\frac{1}{n}\|Xu\|_2^2 - 1\right|}_{:= X_u}$$

X_u is $2\|\frac{1}{n}X^TX - I\|_{op}$ - Lipschitz wrt $\|\cdot\|_2$. Let $\mathcal{N}$ be $\frac{1}{4}$-net:

$$\|\tfrac{1}{n}X^TX - I\|_{op} \leq \tfrac{1}{2} \cdot \|\tfrac{1}{n}X^TX - I\|_{op} + \sup_{u \in \mathcal{N}} X_u$$

$$\Rightarrow \|\tfrac{1}{n}X^TX - I\|_{op} \leq 2 \sup_{u \in \mathcal{N}} X_u$$

$$\Rightarrow \mathbb{P}\left[\|\tfrac{1}{n}X^TX - I\|_{op} \geq \delta\right] \leq \mathbb{P}\left[2 \sup_{u \in \mathcal{N}} X_u \geq \delta\right] \leq \underbrace{|\mathcal{N}|}_{\leq 9^m} \cdot \sup_{u \in \mathcal{N}} \mathbb{P}\left[X_u \geq \tfrac{\delta}{2}\right]$$

Here: $\frac{1}{n}\|Xu\|_2^2 - 1 = \frac{1}{n}\sum_{i=1}^n (x_i^Tu)^2 - \mathbb{E}(x_i^Tu)^2$, where (by assumption)

$$\|x_i^Tu\|_{\psi_2} \leq C\sigma \Rightarrow \|(x_i^Tu)^2 - \mathbb{E}(x_i^Tu)^2\|_{\psi_1} \leq C'\sigma^2$$

$$\Rightarrow \mathbb{P}\left[\left|\tfrac{1}{n}\|Xu\|_2^2 - 1\right| \geq \tfrac{\delta}{2}\right] \leq 2e^{-c_1 n \cdot \min\left(\frac{\delta^2}{\sigma^4}, \frac{\delta}{\sigma^2}\right)}$$

(Bernstein's ineq.)

Set $\delta = \sigma^2 \max(s, s^2)$

$$\Rightarrow \mathbb{P}\left[\|\tfrac{1}{n}X^TX - I\|_{op} \geq \sigma^2 \max(s, s^2)\right] \leq 9^m \cdot 2e^{-c_1 n s^2}$$

Note we must have $\sigma^2 \geq 1$, and for any unit vector $u \in \mathbb{R}^m$,

$$\left|\tfrac{1}{\sqrt{n}}\|Xu\|_2 - 1\right| \geq \sigma^2 s \Rightarrow \left|\tfrac{1}{n}\|Xu\|_2^2 - 1\right| \geq \max(\sigma^2 s, \sigma^4 s^2) \geq \sigma^2 \max(s, s^2).$$

(See Lecture 2.) Thus

$$\mathbb{P}\left[\sup_{u:\|u\|_2=1} \left|\frac{1}{\sqrt{n}}\|Xu\|_2 - 1\right| \geq \sigma^2 s\right]$$

$$\leq \mathbb{P}\left[\left\|\frac{1}{n}X^T X - I\right\|_{op} \geq \sigma^2 \max(s, s^2)\right] \leq 9^m \cdot 2e^{-c_1 n s^2}$$

$$\Rightarrow \mathbb{P}\left[\sqrt{n} - C\sigma^2(\sqrt{m}+t) \leq s_{min}(X) \leq s_{max}(X) \leq \sqrt{n} + C\sigma^2(\sqrt{m}+t)\right]$$

$$= \mathbb{P}\left[\sup_{u:\|u\|_2=1} \left|\frac{1}{\sqrt{n}}\|Xu\|_2 - 1\right| \geq \frac{C\sigma^2}{\sqrt{n}}(\sqrt{m}+t)\right] \leq 2e^{-t^2}$$

for $C>0$ large enough. $\blacksquare$