Stat461 - 2023 Fall
Final Exam

Tuorui Peng[1]

Exercise 1

1.(a)

Since X_i are i.i.d. r.v. with finite variance, using the Central Limit Theorem, we have

$$\mathbb{E}\left[\bar{X}_g\right] = \mu, \quad \mathrm{Var}\left[\bar{X}_g\right] = \frac{\sigma^2}{k}$$

$$\begin{aligned}\mathbb{P}\left(\bar{X}_g - \mu \geq \frac{2\sigma}{\sqrt{k}}\right) \leq& \mathbb{P}\left((\bar{X}_g - \mu)^2 \geq \frac{4\sigma^2}{k}\right)\\ \leq& \frac{var(\bar{X}_g)}{4\sigma^2/k}\\ =& \frac{1}{4}\end{aligned}$$

1.(b)

With $\hat{\mu} := \mathrm{median}(\bar{X}_1, \ldots, \bar{X}_G)$, we have

$$\mathbb{P}\left(\hat{\mu} - \mu \geq 2\sigma/\sqrt{k}\right) = \mathbb{P}\left(\text{less than half of } \bar{X}_g \geq \mu + 2\sigma/\sqrt{k}\right)$$

thus we consider the events $A_g := \{\bar{X}_g \geq 2\sigma/\sqrt{k}\}$ with $\mathbf{1}(A_g) \sim \mathrm{Bin}(p)$ for some $p \in [0,1]$. Using the result from (a), we further have $p \leq 1/4$.

Since $\bar{X}_g$ are i.i.d., we have

$$\begin{aligned}\mathbb{P}\left(\hat{\mu} - \mu \geq 2\sigma/\sqrt{k}\right) \leq& \mathbb{P}\left(\mathrm{Bin}(G,p) \geq G/2\right)\\ \leq& \mathbb{P}\left(\mathrm{Bin}(G,1/4) \geq G/2\right)\end{aligned}$$

1.(c)

Note that $\mathrm{Bin}(G,1/4)$ is sub-gaussian with σ at most $\sqrt{G}/2$, we thus have

$$\begin{aligned}\mathbb{P}\left(\mathrm{Bin}(G,1/4) \geq G/2\right) =& \mathbb{P}\left(\mathrm{Bin}(G,1/4) - G/4 \geq G/4\right)\\ \leq& \exp\left(-\frac{G^2}{8G}\right)\\ =& \exp\left(-\frac{G}{8}\right) = \delta\end{aligned}$$

[1]TuoruiPeng2028@u.northwestern.edu

then we substitute $k = n/G = \dfrac{n}{8\log 1/\delta}$ to result in (b) to obtain

$$\mathbb{P}\left(\hat{\mu} - \mu \geq 4\sigma\sqrt{\frac{2\log(1/\delta)}{n}}\right) \leq \delta$$

1.(d)

Denote $t = 4\sigma\sqrt{\dfrac{2\log(1/\delta)}{n}}$, which yields

$$\mathbb{P}\left(\hat{\mu} - \mu \geq t\right) \leq \delta = e^{-\dfrac{t^2}{2(4\sigma/\sqrt{n})^2}}$$

which means that $\hat{\mu}$ is sub-gaussian with σ at most $4\sigma/\sqrt{n}$.

Exercise 2

The proof is in Vershynin's book, Theorem 0.0.4. Here is a sketch of the proof.

First we work on a Lemma: for any $x \in P$, where $P = \operatorname{conv}(x_1, \ldots, x_N)$ being the polytope, we have

$$\left\| x - \frac{1}{k}\sum_{i=1}^{k} x_i \right\| \leq \frac{1}{\sqrt{k}}, \quad \text{for some } \{x_i\} \subset P,\ \forall k \in [N]$$

The proof is as follows:

Proof. Note that for such polytope, for any $x \in P$ we can always express it as (using Carahtéodory's theorem)

$$x = \sum_{i=1}^{N} \lambda_i x_i, \quad \sum_{i=1}^{N} \lambda_i = 1,\ \lambda_i \geq 0$$

Consider an r.v. $Z = x_i$ w.p. λ_i, we have

$$\mathbb{E}\left[Z\right] = x$$

then we create k copies of Z, denoted as $Z_1, \ldots, Z_k$, and we have

$$\begin{aligned}
\mathbb{E}\left\| x - \frac{1}{k}\sum_{i=1}^{k} Z_i \right\|_2^2 =& var(\frac{1}{k}\sum_{i=1}^{k} Z_i) \\
=& \frac{1}{k} var(Z) \\
\leq& \frac{1}{k}
\end{aligned}$$

since $Z \in P$ is bounded by 1. Then since Z_i takes value in $\{x_1, \ldots, x_N\}$, then there should always exist some realization of Z_i as $Z_i =$ some vertice that satisfies the inequality. □

Using the Lemma, we can just choose $k = \lceil 1/\varepsilon^2 \rceil$ and obtain that for any $x \in P$, we have

$$\exists \{x_i\}_{i=1}^{k} \subset P,\ \left\| x - \frac{1}{k}\sum_{i=1}^{k} x_i \right\| \leq \frac{1}{\sqrt{k}} \leq \varepsilon$$

so such $\{\frac{1}{k}\sum_{i=1}^{k} x_i : x_i \in$ vertices of $P\}$ set is a ε covering of P, which has size at most $N^k = N^{\lceil 1/\varepsilon^2 \rceil}$.

Exercise 3

3.(a)

Note that we have

$$\left\|y - X\hat{\theta}\right\|_2^2 \leq \|y - X\theta^*\|_2^2$$

which gives

$$\left\|X(\hat{\theta} - \theta^*)\right\|_2^2 \leq 2w'\left(X(\hat{\theta} - \theta^*)\right) \Rightarrow \left\|X(\hat{\theta} - \theta^*)\right\|_2 \leq 2w'\frac{X(\hat{\theta} - \theta^*)}{\left\|X(\hat{\theta} - \theta^*)\right\|_2}$$

3.(b)

Note that $\hat{\theta} - \theta^*$ should be $2s$-sparse, and $X(\hat{\theta} - \theta^*) \in \mathrm{col}(X_{\mathrm{span}(\hat{\theta}\cup\theta^*)})$, we thus have

$$w'\frac{X(\hat{\theta} - \theta^*)}{\left\|X(\hat{\theta} - \theta^*)\right\|_2} \leq \sup_{|S|\leq 2s} \sup_{v_S\in\mathbb{S}^{n-1}, v_S\in\mathrm{col}(X_S)} w'v_S$$

3.(c)

We have for $w \mapsto \sup_{v_S\in\mathbb{S}^{n-1}, v_S\in\mathrm{col}(X_S)} w'v_S := g(w)$:

$$g(w) = g(w_S + w_{S^\complement}) = g(w_S) = \|w_S\|_2$$

in which w_S is the projection of w onto $\mathrm{col}(X_S)$, and $w_{S^\complement}$ is its orthogonal complement. Further since $w \sim N(0, \sigma^2 I)$, we can consider

$$w = P\tilde{w} = P_S\tilde{w}_S + P_{S^\complement}\tilde{w}_{S^\complement}, \quad P \in \mathrm{SO}(n)$$

i.e. decompose w into w_S and $w_{S^\complement}$, with each following gaussian distribution.

Then for $w \sim N(0, \sigma^2 I)$ we have

$$\mathbb{E}\left[g(w)\right] = \mathbb{E}\left[\|w_S\|_2\right] = \mathbb{E}\left[\|P_S\tilde{w}_S\|_2\right] = \sigma\mathbb{E}\left[\sqrt{\chi^2_{2s}}\right] \leq \sigma\sqrt{\mathbb{E}\chi^2_{2s}} = \sigma\sqrt{2s}$$

Then we note that for any $w, y \in \mathbb{R}^d$:

$$|g(w) - g(y)| = \|w_S - y_S\|_2 \leq \|w - y\|_2$$

which means that $g(w)$ is 1-Lipschitz. Then we can apply the concentration inequality for Lipschitz function to obtain that

$$\begin{aligned}
&\mathbb{P}\left(|g(w) - \mathbb{E}g(w)| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right) \\
\Rightarrow &\mathbb{P}\left(\sup_{v_S\in\mathbb{S}^{n-1}, v_S\in\mathrm{col}(X_S)} w'v_S \geq \sigma\sqrt{2s} + t\right) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right)
\end{aligned}$$

letting $\delta = 2\exp\left(-\dfrac{t^2}{2\sigma^2}\right)$ we have: w.p. at least $1-\delta$,

$$\sup_{v_S\in\mathbb{S}^{n-1}, v_S\in\mathrm{col}(X_S)} w'v_S \leq \sigma\big(\sqrt{2s}+\sqrt{2\log(2/\delta)}\big)$$

3.(d)

Using union bound on $\{S\subset[d]:|S|\leq 2s\}$ we have

$$\sup_{|S|\leq 2s}\sup_{v_S\in\mathbb{S}^{n-1}, v_S\in\mathrm{col}(X_S)} w'v_S \leq \sigma\big(\sqrt{2s}+\sqrt{2\log(2/\delta)}\big)$$

with probability at least

$$1-\delta\sum_{j=1}^{2s}\binom{d}{j}\geq 1-\delta\big(\frac{ed}{2s}\big)^{2s}$$

then we obtain the bound

$$\left\|X(\hat\theta-\theta^*)\right\|_2 \leq 2w'\frac{X(\hat\theta-\theta^*)}{\left\|X(\hat\theta-\theta^*)\right\|_2}\leq 2\sigma\big(\sqrt{2s}+\sqrt{2\log(2/\delta)}\big),\quad \text{w.p. at least } 1-\delta\big(\frac{ed}{2s}\big)^{2s}$$