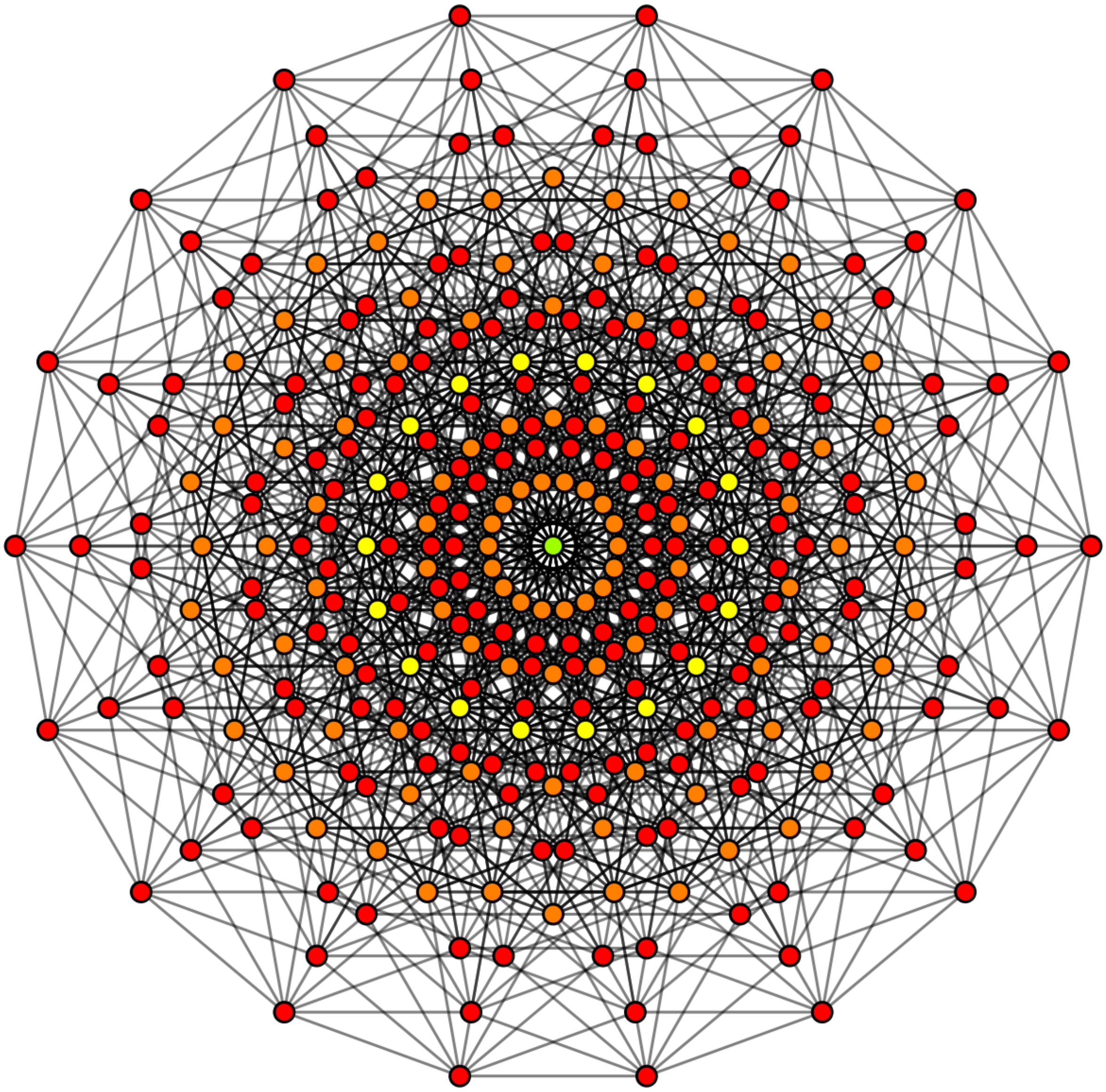
High-Dimensional Probability
Solution Manual

Pingbang Hu

July 26, 2024

Abstract

This is the solution I write when organizing the reading group on Roman Vershynin's *High Dimensional Probability* [Ver24]. While we aim to solve all the exercises, occasionally we omit some due to either 1.) simplicity; 2.) difficulty; or 3.) skipped section. Additionally, it may contain factual and/or typographic errors.



The reading group started from Spring 2024, and the date on the cover page is the last updated time.

Contents

Appetizer: using probability to cover a geometric set

Week 1: Appetizer and Basic Inequalities

19 Jan. 2024

Problem (Exercise 0.0.3). Check the following variance identities that we used in the proof of Theorem 0.0.2.

(a) Let $Z_1, \dots, Z_k$ be independent mean zero random vectors in $\mathbb{R}^n$. Show that

$$\mathbb{E}\left[\left\|\sum_{j=1}^{k} Z_j\right\|_2^2\right] = \sum_{j=1}^{k} \mathbb{E}[\|Z_j\|_2^2].$$

(b) Let Z be a random vector in $\mathbb{R}^n$. Show that

$$\mathbb{E}[\|Z - \mathbb{E}[Z]\|_2^2] = \mathbb{E}[\|Z\|_2^2] - \|\mathbb{E}[Z]\|_2^2.$$

Answer. (a) If $Z_1, \dots, Z_k$ are independent mean zero random vectors in $\mathbb{R}^n$, then

$$\mathbb{E}\left[\left\|\sum_{j=1}^{k} Z_j\right\|_2^2\right] = \mathbb{E}\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{k}(Z_j)_i\right)^2\right] = \sum_{i=1}^{n}\mathbb{E}\left[\left(\sum_{j=1}^{k}(Z_j)_i\right)^2\right].$$

From the assumption, $\mathbb{E}\left[(Z_j)_i (Z_{j'})_i\right] = \mathbb{E}\left[(Z_j)_i\right]\mathbb{E}\left[(Z_{j'})_i\right] = 0$, hence

$$\sum_{i=1}^{n}\mathbb{E}\left[\left(\sum_{j=1}^{k}(Z_j)_i\right)^2\right] = \sum_{i=1}^{n}\mathbb{E}\left[\sum_{j=1}^{k}(Z_j)_i^2\right] = \sum_{j=1}^{k}\mathbb{E}\left[\sum_{i=1}^{n}(Z_j)_i^2\right] = \sum_{j=1}^{k}\mathbb{E}\left[\|Z_j\|_2^2\right],$$

proving the result.

(b) If Z is a random vector in $\mathbb{R}^n$, then

$$\begin{aligned}
\mathbb{E}\left[\|Z - \mathbb{E}[Z]\|_2^2\right] &= \mathbb{E}\left[\sum_{i=1}^{n}(Z_i - \mathbb{E}[Z_i])^2\right] \\
&= \sum_{i=1}^{n}\mathbb{E}\left[Z_i^2 - 2Z_i\mathbb{E}[Z_i] + (\mathbb{E}[Z_i])^2\right] \\
&= \sum_{i=1}^{n}\mathbb{E}\left[Z_i^2\right] - 2\sum_{i=1}^{n}\mathbb{E}[Z_i]\mathbb{E}[Z_i] + \sum_{i=1}^{n}\mathbb{E}[Z_i]^2 \\
&= \mathbb{E}\left[\|Z\|_2^2\right] - \|\mathbb{E}[Z]\|_2^2.
\end{aligned}$$

⊛

Problem (Exercise 0.0.5). Prove the inequalities

$$\left(\frac{n}{m}\right)^m \leq \binom{n}{m} \leq \sum_{k=0}^{m} \binom{n}{k} \leq \left(\frac{en}{m}\right)^m$$

for all integers $m \in [1, n]$.

Answer. Fix some $m \in [1, n]$. We first show $(n/m)^m \leq \binom{n}{m}$. This is because

$$\frac{(n/m)^m}{\binom{n}{m}} = \prod_{j=0}^{m-1} \left(\frac{n}{m}\frac{m-j}{n-j}\right) \leq 1$$

as $\frac{n-j}{m-j} \geq \frac{n}{m}$ for all j. The second inequality $\binom{n}{m} \leq \sum_{k=0}^{m} \binom{n}{k}$ is trivial since $\binom{n}{k} \geq 1$ for all k. The last inequality is due to

$$\frac{\sum_{k=0}^{m} \binom{n}{k}}{\left(\frac{n}{m}\right)^m} \leq \sum_{k=0}^{n} \binom{n}{k} \left(\frac{m}{n}\right)^k = \left(1 + \frac{m}{n}\right)^n \leq e^m.$$

⊛

Problem (Exercise 0.0.6). Check that in Corollary 0.0.4,

$$(C + C\epsilon^2 N)^{\lceil 1/\epsilon^2 \rceil}$$

suffice. Here C is a suitable absolute constant.

Answer. Omit. ⊛

Chapter 1

Preliminaries on random variables

1.1 Basic quantities associated with random variables

No Exercise!

1.2 Some classical inequalities

Problem (Exercise 1.2.2). Prove the following extension of Lemma 1.2.1, which is valid for any random variable X (not necessarily non-negative):

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t)\,\mathrm{d}t - \int_{-\infty}^0 \mathbb{P}(X < t)\,\mathrm{d}t.$$

Answer. Separating X into the plus and minus parts would do the job. Specifically, let $X = X_+ - X_-$ where $X_+ = \max(X, 0)$ and $X_- = \max(-X, 0)$, both are non-negative. Then, we see that by applying Lemma 1.2.1,

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X_+] - \mathbb{E}[X_-] \\
&= \int_0^\infty \Pr(t < X_+)\,\mathrm{d}t - \int_0^\infty \Pr(t < X_-)\,\mathrm{d}t \\
&= \int_0^\infty \Pr(X > t)\,\mathrm{d}t - \int_0^\infty \Pr(X < -t)\,\mathrm{d}t \\
&= \int_0^\infty \Pr(X > t)\,\mathrm{d}t - \int_{-\infty}^0 \Pr(X < t)\,\mathrm{d}t.
\end{aligned}$$

⊛

Problem (Exercise 1.2.3). Let X be a random variable and $p \in (0, \infty)$. Show that

$$\mathbb{E}[|X|^p] = \int_0^\infty p t^{p-1} \mathbb{P}(|X| > t)\,\mathrm{d}t$$

whenever the right-hand side is finite.

Answer. Since $|X|$ is non-negative, from Lemma 1.2.1, we have

$$\mathbb{E}[|X|^p] = \int_0^\infty \Pr(t < |X|^p)\,\mathrm{d}t = \int_0^\infty p t^{p-1} \Pr(|X| > t)\,\mathrm{d}t$$

where we let $t \leftarrow t^p$, hence $\mathrm{d}t \leftarrow p t^{p-1} \mathrm{d}t$. ⊛

Week 2: Basic Inequalities and Limit Theorems

24 Jan. 2024

Problem (Exercise 1.2.6). Deduce Chebyshev's inequality by squaring both sides of the bound $|X - \mu| \geq t$ and applying Markov's inequality.

Answer. From Markov's inequality, for any $t > 0$,

$$\Pr(|X-\mu| \geq t) = \Pr(|X-\mu|^2 \geq t^2) \leq \frac{\mathbb{E}\left[|X-\mu|^2\right]}{t^2} = \frac{\sigma^2}{t^2}.$$

⊛

1.3 Limit theorems

Problem (Exercise 1.3.3). Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and finite variance. Show that

$$\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N X_i - \mu\right|\right] = O\left(\frac{1}{\sqrt{N}}\right) \text{ as } N \to \infty.$$

Answer. We see that

$$\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N X_i - \mu\right|\right] \leq \sqrt{\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N X_i - \mu\right|^2\right]} = \sqrt{\operatorname{Var}\left[\frac{1}{N}\sum_{i=1}^N X_i\right]} = \frac{\sigma}{\sqrt{N}}.$$

As $\sigma < \infty$ is a constant, the rate is exactly $O(1/\sqrt{N})$. ⊛

Chapter 2

Concentration of sums of independent random variables

Week 3: More Powerful Concentration Inequalities

2.1 Why concentration inequalities?

2 Feb. 2024

Problem (Exercise 2.1.4). Let $g \sim \mathcal{N}(0,1)$. Show that for all $t \geq 1$, we have

$$\mathbb{E}[g^2 \mathbb{1}_{g>t}] = t \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} + \mathbb{P}(g > t) \leq \left(t - \frac{1}{t}\right) \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

Answer. Denote the standard normal density as

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Since we have $\Phi'(x) = -x\Phi(x)$, by integration by part,

$$\begin{aligned}
\mathbb{E}\left[g^2 \mathbb{1}_{g>t}\right] &= \int_0^\infty x^2 \mathbb{1}_{x>t} \Phi(x) \,\mathrm{d}x \\
&= -\int_t^\infty x\Phi'(x) \,\mathrm{d}x \\
&= -\left. x\Phi(x)\right|_t^\infty + \int_t^\infty \Phi(x) \,\mathrm{d}x \\
&= t \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} + \mathbb{P}(g > t),
\end{aligned}$$

which gives the first equality. Furthermore, as $t \geq 1$, we trivially have

$$\int_t^\infty \Phi(x) \,\mathrm{d}x \leq \int_t^\infty \frac{x}{t} \Phi(x) \,\mathrm{d}x = \frac{1}{t} \int_t^\infty -\Phi'(x) \,\mathrm{d}x = \frac{\Phi(t)}{t},$$

implying that

$$\mathbb{E}\left[g^2 \mathbb{1}_{g>t}\right] = t \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} + \int_t^\infty \Phi(x) \,\mathrm{d}x \leq \left(t + \frac{1}{t}\right) \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

which gives the second inequality. ⊛

2.2 Hoeffding's inequality

Problem (Exercise 2.2.3)**.** Show that

$$\cosh(x) \leq \exp\left(x^2/2\right) \text{ for all } x \in \mathbb{R}.$$

Answer. Omit. ⊛

The next exercise is to prove Theorem 2.2.5 (Hoeffding's inequality for general bounded random variables), which we restate it for convenience.

Theorem 2.2.1 (Hoeffding's inequality for general bounded random variables)**.** Let $X_1, \ldots, X_N$ be independent random variables. Assume that $X_i \in [m_i, M_i]$ for every i. Then, for any $t > 0$, we have

$$\mathbb{P}\left(\sum_{i=1}^{N}(X_i - \mathbb{E}[X_i]) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{N}(M_i - m_i)^2}\right).$$

Problem (Exercise 2.2.7)**.** Prove the Hoeffding's inequality for general bounded random variables, possibly with some absolute constant instead of 2 in the tail.

Answer. Since raising both sides to p-th power doesn't work since we're now working with sum of random variables, so we instead consider the MGF trick (also known as Crarmer-Chernoff method):

Lemma 2.2.1 (Crarmer-Chernoff method)**.** Given a random variable X,

$$\mathbb{P}(X - \mu \geq t) = \mathbb{P}(e^{\lambda(X-\mu)} \geq e^{\lambda t}) \leq \inf_{\lambda > 0} \frac{\mathbb{E}\left[e^{\lambda(X-\mu)}\right]}{e^{\lambda t}}.$$

Proof. This directly follows from the Markov's inequality. ■

Hence, we see that

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{N}(X_i - \mathbb{E}\left[X_i\right]) \geq t\right) &\leq \inf_{\lambda > 0} e^{-\lambda t}\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{N}(X_i - \mathbb{E}\left[X_i\right])\right)\right] \\ &= \inf_{\lambda > 0} e^{-\lambda t}\prod_{i=1}^{N}\exp(\lambda(X_i - \mathbb{E}\left[X_i\right])).\end{aligned}$$

So now everything left is to bound $\mathbb{E}\left[\exp(\lambda(X_i - \mathbb{E}\left[X_i\right]))\right]$. Before we proceed, we need one lemma.

Lemma 2.2.2. For any bounded random variable $Z \in [a, b]$,

$$\mathrm{Var}\left[Z\right] \leq \frac{(b-a)^2}{4}.$$

Proof. Since

$$\mathrm{Var}\left[Z\right] = \mathrm{Var}\left[Z - \frac{a+b}{2}\right] \leq \mathbb{E}\left[\left(Z - \frac{a+b}{2}\right)^2\right] \leq \frac{(b-a)^2}{4}.$$

■

Claim. Given $X \in [a, b]$ such that $\mathbb{E}\left[X\right] = 0$, for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \exp\left(\lambda^2\frac{(b-a)^2}{8}\right).$$

Proof. We first define $\psi(\lambda) = \ln \mathbb{E}\left[e^{\lambda X}\right]$, and compute

$$\psi'(\lambda) = \frac{\mathbb{E}\left[Xe^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]}, \quad \psi''(\lambda) = \frac{\mathbb{E}\left[X^2 e^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]} - \left(\frac{\mathbb{E}\left[Xe^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]}\right)^2.$$

Now, observe that ψ'' is the variance under the law of X re-weighted by $\frac{e^{\lambda X}}{\mathbb{E}[e^{\lambda X}]}$, i.e., by a change of measure, consider a new distribution $\mathbb{P}_\lambda$ (w.r.t. the original distribution $\mathbb{P}$ of X) as

$$\mathrm{d}\mathbb{P}_\lambda(x) \coloneqq \frac{e^{\lambda X}}{\mathbb{E}_{\mathbb{P}}\left[e^{\lambda X}\right]} \,\mathrm{d}\mathbb{P}(x),$$

then

$$\psi'(\lambda) = \frac{\mathbb{E}_{\mathbb{P}}\left[Xe^{\lambda X}\right]}{\mathbb{E}_{\mathbb{P}}\left[e^{\lambda X}\right]} = \int \frac{xe^{\lambda x}}{\mathbb{E}_{\mathbb{P}}\left[e^{\lambda X}\right]} \,\mathrm{d}\mathbb{P}(x) = \mathbb{E}_{\mathbb{P}_\lambda}\left[X\right]$$

and

$$\psi''(\lambda) = \frac{\mathbb{E}_{\mathbb{P}}\left[X^2 e^{\lambda X}\right]}{\mathbb{E}_{\mathbb{P}}\left[e^{\lambda X}\right]} - \left(\frac{\mathbb{E}_{\mathbb{P}}\left[Xe^{\lambda X}\right]}{\mathbb{E}_{\mathbb{P}}\left[e^{\lambda X}\right]}\right)^2 = \mathbb{E}_{\mathbb{P}_\lambda}\left[X^2\right] - \mathbb{E}_{\mathbb{P}_\lambda}\left[X\right]^2 = \mathrm{Var}_{\mathbb{P}_\lambda}\left[X\right].$$

From Lemma 2.2.2, since X under the new distribution $\mathbb{P}_\lambda$ is still bounded between a and b,

$$\psi''(\lambda) = \mathrm{Var}_{\mathbb{P}_\lambda}\left[X\right] \leq \frac{(b-a)^2}{4}.$$

Then by Taylor's theorem, there exists some $\widetilde{\lambda} \in [0, \lambda]$ such that

$$\psi(\lambda) = \psi(0) + \psi'(0)\lambda + \frac{1}{2}\psi''(\widetilde{\lambda})\lambda^2 = \frac{1}{2}\psi''(\widetilde{\lambda})\lambda^2$$

since $\psi(0) = \psi'(0) = 0$. By bounding $\psi''(\widetilde{\lambda})\lambda^2/2$, we finally have

$$\ln \mathbb{E}\left[e^{\lambda X}\right] = \psi(\lambda) \leq \frac{1}{2} \cdot \frac{(b-a)^2}{4}\lambda^2 = \lambda^2 \frac{(b-a)^2}{8},$$

raising both sides by e shows the desired result. ⊛

Say given $X_i \in [m_i, M_i]$ for every i, then $X_i - \mathbb{E}\left[X_i\right] \in \left[m_i - \mathbb{E}\left[X_i\right], M_i - \mathbb{E}\left[X_i\right]\right]$ with mean 0 for every i. Then given any of the two bounds, for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}\left[e^{\lambda(X_i - \mathbb{E}[X_i])}\right] \leq \exp\left(\lambda^2 \frac{(M_i - m_i)^2}{8}\right).$$

Then we simply recall that

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^N (X_i - \mathbb{E}\left[X_i\right]) \geq t\right) &= \inf_{\lambda > 0} e^{-\lambda t} \prod_{i=1}^N \exp(\lambda(X_i - \mathbb{E}\left[X_i\right])) \\
&\leq \inf_{\lambda > 0} \exp\left(-\lambda t + \sum_{i=1}^N \lambda^2 \frac{(M_i - m_i)^2}{8}\right) \\
&= \exp\left(-\frac{4t^2}{\sum_{i=1}^N (M_i - m_i)^2} + \frac{2t^2}{\sum_{i=1}^N (M_i - m_i)^2}\right) \\
&= \exp\left(-\frac{2t^2}{\sum_{i=1}^N (M_i - m_i)^2}\right)
\end{aligned}$$

since infimum is achieved at $\lambda = 4t/(\sum_{i=1}^N (M_i - m_i)^2)$. ⊛

Problem (Exercise 2.2.8). Imagine we have an algorithm for solving some decision problem (e.g., is a given number p a prime?). Suppose the algorithm makes a decision at random and returns the correct answer with probability $\frac{1}{2}+\delta$ with some $\delta > 0$, which is just a bit better than a random guess. To improve the performance, we run the algorithm N times and take the majority vote. Show that, for any $\epsilon \in (0,1)$, the answer is correct with probability at least $1-\epsilon$, as long as

$$N \geq \frac{1}{2\delta^2}\ln\left(\frac{1}{\epsilon}\right).$$

Answer. Consider $X_1, \dots, X_N \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(\frac{1}{2}+\delta)$, which is a series of indicators indicting whether the random decision is correct or not. Note that $\mathbb{E}[X_i] = \frac{1}{2}+\delta$.

We see that by taking majority vote over N times, the algorithm makes a mistake if $\sum_{i=1}^N X_i \leq N/2$ (let's not consider tie). This happens with probability

$$\mathbb{P}\left(\sum_{i=1}^N X_i \leq \frac{N}{2}\right) = \mathbb{P}\left(\sum_{i=1}^N (X_i - \mathbb{E}[X_i]) \leq -N\delta\right) \leq \exp\left(-\frac{2(N\delta)^2}{N}\right) = e^{-2N\delta^2}$$

from Hoeffding's inequality.[a] Requiring $e^{-2N\delta^2} \leq \epsilon$ is equivalent to requiring $N \geq \frac{1}{2\delta^2}\ln(1/\epsilon)$. ⊛

[a] Note that the sign is flipped. However, Hoeffding's inequality still holds (why?).

Problem (Exercise 2.2.9). Suppose we want to estimate the mean μ of a random variable X from a sample $X_1, \dots, X_N$ drawn independently from the distribution of X. We want an ϵ-accurate estimate, i.e., one that falls in the interval $(\mu-\epsilon, \mu+\epsilon)$.

(a) Show that a sample of size $N = O(\sigma^2/\epsilon^2)$ is sufficient to compute an ϵ-accurate estimate with probability at least $3/4$, where $s;^2 = \mathrm{Var}[X]$.

(b) Show that a sample of size $N = O(\log(\delta^{-1})\sigma^2/\epsilon^2)$ is sufficient to compute an ϵ-accurate estimate with probability at least $1-\delta$.

Answer. (a) Consider using the sample mean $\hat{\mu} = \frac{1}{N}\sum_{i=1}^N X_i$ as an estimator of μ. From the Chebyshev's inequality,

$$\mathbb{P}(|\hat{\mu}-\mu| > \epsilon) \leq \frac{\sigma^2/N}{\epsilon^2}.$$

By requiring $\sigma^2/(N\epsilon^2) \leq 1/4$, i.e., $N \geq 4\sigma^2/\epsilon^2 = O(\sigma^2/\epsilon^2)$, suffices.

(b) Consider gathering k estimator from the above procedure, i.e., we now have $\hat{\mu}_1, \dots, \hat{\mu}_k$ such that each are an ϵ-accurate mean estimator with probability at least $3/4$. This requires $k \cdot 4\sigma^2/\epsilon^2 = O(k\sigma^2/\epsilon^2)$ samples. We claim that the median $\hat{\mu} \coloneqq \mathrm{median}(\hat{\mu}_1, \dots, \hat{\mu}_k)$ is an ϵ-accurate mean estimator with probability at least $1-\delta$ for some k (depends on δ). Consider a series of indicators $X_i = \mathbb{1}_{|\hat{\mu}_i - \mu| > \epsilon}$, indicating if $\hat{\mu}_i$ is not ϵ-accurate. Then $X_i \sim \mathrm{Ber}(1/4)$. Then, our median estimator $\hat{\mu}$ fails with probability

$$\mathbb{P}(|\hat{\mu}-\mu| > \epsilon) = \mathbb{P}\left(\sum_{i=1}^k X_i > \frac{k}{2}\right) = \mathbb{P}\left(\sum_{i=1}^k (X_i - \mathbb{E}[X_i]) > \frac{k}{4}\right)$$

as $\mathbb{E}[X_i] = 1/4$. From Hoeffding's inequality, the above probability is bounded above by $\exp(-2(k/4)^2/k)$, setting it to be less than δ we have

$$\exp\left(-\frac{2(k/4)^2}{k}\right) \leq \delta \Leftrightarrow \ln\left(\frac{1}{\delta}\right) \geq \frac{k}{8} \Leftrightarrow k = O(\ln(\delta^{-1})),$$

i.e., the total number of samples required is $O(k\sigma^2/\epsilon^2) = O(\ln(\delta^{-1})\sigma^2/\epsilon^2)$.

⊛

Problem (Exercise 2.2.10). Let $X_1, \dots, X_N$ be *non-negative* independent random variables with continuous distributions. Assume that the densities of X_i are uniformly bounded by 1.

(a) Show that the MGF of X_i satisfies

$$\mathbb{E}[\exp(-tX_i)] \leq \frac{1}{t} \text{ for all } t > 0.$$

(b) Deduce that, for any $\epsilon > 0$, we have

$$\mathbb{P}\left(\sum_{i=1}^{N} X_i \leq \epsilon N\right) \leq (e\epsilon)^N.$$

Answer. (a) Since X_i's are non-negative and the densities $f_{X_i} \leq 1$ uniformly, for every $t > 0$,

$$\mathbb{E}\left[\exp(-tX_i)\right] = \int_0^\infty e^{-tx} f_{X_i}(x) \,\mathrm{d}x \leq \int_0^\infty e^{-tx} \,\mathrm{d}x = -\frac{1}{t} e^{-tx} \bigg|_0^\infty = \frac{1}{t}.$$

(b) From Chernoff's inequality, for any $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{N} X_i \leq \epsilon N\right) &= \mathbb{P}\left(\sum_{i=1}^{N} -\frac{X_i}{\epsilon} \geq -N\right) \\
&\leq \inf_{\lambda > 0} e^{\lambda N} \mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{N} -\frac{X_i}{\epsilon}\right)\right] \\
&= \inf_{\lambda > 0} e^{\lambda N} \prod_{i=1}^{N} \mathbb{E}\left[\exp\left(-\lambda \frac{X_i}{\epsilon}\right)\right] \\
&\leq \inf_{\lambda > 0} e^{\lambda N} \prod_{i=1}^{N} \frac{\epsilon}{\lambda} && \text{Part (a) with } t = \lambda/\epsilon \\
&= \inf_{\lambda > 0} \left(e^{\lambda} \frac{\epsilon}{\lambda}\right)^N \\
&= (e\epsilon)^N
\end{aligned}$$

since the infimum is achieved when $\lambda = 1$.

⊛

2.3 Chernoff's inequality

Problem (Exercise 2.3.2). Modify the proof of Theorem 2.3.1 to obtain the following bound on the lower tail. For any $t < \mu$, we have

$$\mathbb{P}(S_N \leq t) \leq e^{-\mu} \left(\frac{e\mu}{t}\right)^t.$$

Answer. A direct modification is that considering for any $\lambda > 0$,

$$\mathbb{P}(S_N \leq t) = \mathbb{P}(-S_N \geq -t) = \mathbb{P}(e^{-\lambda S_n} \geq e^{-\lambda t}) \leq e^{\lambda t} \prod_{i=1}^{N} \mathbb{E}\left[\exp(-\lambda X_i)\right].$$

A direct computation gives

$$\mathbb{E}\left[\exp(-\lambda X_i)\right] = e^{-\lambda} p_i + (1 - p_i) = 1 + (e^{-\lambda} - 1) p_i \leq \exp\big((e^{-\lambda} - 1) p_i\big),$$

hence

$$\mathbb{P}(S_N \leq t) \leq e^{\lambda t} \prod_{i=1}^{N} \exp\big((e^{-\lambda} - 1)p_i\big) = e^{\lambda t} \exp\big((e^{-\lambda} - 1)\mu\big) = \exp\big(\lambda t + (e^{-\lambda} - 1)\mu\big).$$

Minimizing the right-hand side, we see that

$$t + (-\mu e^{-\lambda}) = 0 \Leftrightarrow t = \mu e^{-\lambda} \Leftrightarrow \lambda = \ln \frac{\mu}{t}$$

achieves the infimum. And since $t < \mu$, $\lambda > 0$ as required, which gives

$$\mathbb{P}(S_N \leq t) \leq \exp\left(t \ln \frac{\mu}{t} + \left(\frac{t}{\mu} - 1\right)\mu\right) = \exp\left(t \ln \frac{\mu}{t} + t - \mu\right) = e^{-\mu}\left(\frac{e\mu}{t}\right)^t.$$

⊛

Problem (Exercise 2.3.3)**.** Let $X \sim \mathrm{Pois}(\lambda)$. Show that for any $t > \lambda$, we have

$$\mathbb{P}(X \geq t) \leq e^{-\lambda}\left(\frac{e\lambda}{t}\right)^t.$$

Answer. From Chernoff's inequality, for any $\theta > 0$, we have

$$\mathbb{P}(X \geq t) \leq e^{-\theta t}\mathbb{E}\left[\exp(\theta X)\right].$$

Then the Poisson moment can be calculated as

$$\mathbb{E}\left[\exp(\theta X)\right] = \sum_{k=0}^{\infty} e^{\theta k} \cdot e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(e^\theta \lambda)^k}{k!} = e^{-\lambda}\exp\big(e^\theta\lambda\big) = \exp\big((e^\theta - 1)\lambda\big),$$

hence

$$\mathbb{P}(X \geq t) \leq e^{-\theta t}\exp\big((e^\theta - 1)\lambda\big) = \left(\frac{\lambda}{t}\right)^t \exp(t - \lambda) = e^{-\lambda}\left(\frac{e\lambda}{t}\right)^t$$

where we take the minimizing $\theta = \ln(t/\lambda) > 0$ as $t > \lambda$. ⊛

Alternatively, we can also solve Exercise 2.3.3 directly as follows.

Answer. Consider a series of independent Bernoulli random variables $X_{N,i}$ for a fixed N such that the Poisson limit theorem applies to approximate $X \sim \mathrm{Pois}(\lambda)$, i.e., as $N \to \infty$, $\max_{i \leq N} p_{N,i} \to 0$ and $\lambda_N \coloneqq \mathbb{E}\left[S_N\right] \to \lambda < \infty$, $S_N \to \mathrm{Pois}(\lambda)$. From Chernoff's inequality, for any $t > \lambda_N$,

$$\mathbb{P}(S_N > t) \leq e^{-\lambda_N}\left(\frac{e\lambda_N}{t}\right)^t.$$

We then see that

$$\mathbb{P}(X > t) = \lim_{N\to\infty}\mathbb{P}(S_N > t) \leq \lim_{N\to\infty} e^{-\lambda_N}\left(\frac{e\lambda_N}{t}\right)^t = e^{-\lambda}\left(\frac{e\lambda}{t}\right)^t$$

since $\lambda_N \to \lambda$ as $N \to \infty$. ⊛

Week 4: Chernoff's Inequality and Degree Concentration

7 Feb. 2024

Problem (Exercise 2.3.5)**.** Show that, in the setting of Theorem 2.3.1, for $\delta \in (0, 1]$ we have

$$\mathbb{P}(|S_N - \mu| \geq \delta\mu) \leq 2e^{-c\mu\delta^2}$$

where $c > 0$ is an absolute constant.

Answer. From Chernoff's inequality (right-tail), for $t = (1+\delta)\mu$, we have

$$\begin{aligned}\ln \mathbb{P}(S_N \geq (1+\delta)\mu) &\leq -\mu + (1+\delta)\mu\,(1 + \ln\mu - \ln(1+\delta) - \ln\mu) \\ &= \delta\mu - (1+\delta)\mu(\ln(1+\delta)) \\ &= \mu(\delta - (1+\delta)\ln(1+\delta)).\end{aligned}$$

A classic bound for $\ln(1+\delta)$ is the following.

Claim. For all $x > 0$,

$$\frac{2x}{2+x} \leq \ln(1+x).$$

Proof. As $(1+x/2)^2 = 1 + x + x^2/4 \geq 1 + x$,

$$[\log(1+x)]' = \frac{1}{1+x} \geq \frac{1}{(1+x/2)^2} = \left(\frac{x}{1+x/2}\right)'.$$

Note that $\log(1+x) = x/(1+x/2) = 0$ at $x = 0$, so for all $x > 0$

$$\log(1+x) \geq \frac{x}{1+x/2}.$$

⊛

Hence, as our $\delta \in (0,1]$, we have

$$\ln \mathbb{P}(S_N \geq (1+\delta)\mu) \leq \mu(\delta - (1+\delta)\ln(1+\delta)) \leq \mu\delta - \mu(1+\delta)\frac{2\delta}{2+\delta} = -\frac{\mu\delta^2}{2+\delta} \leq -\frac{\mu\delta^2}{3}.$$

Similarly, from Chernoff's inequality (left-tail), for $t = (1-\delta)\mu$, we have

$$\begin{aligned}\ln \mathbb{P}(S_N \leq (1-\delta)\mu) &\leq -\mu + (1-\delta)\mu(1 + \ln\mu - \ln(1-\delta) - \ln\mu) \\ &= -\delta\mu - (1-\delta)\mu\ln(1-\delta) \\ &= \mu(-\delta - (1-\delta)\ln(1-\delta)).\end{aligned}$$

Another classic bound for $\ln(1-\delta)$ is the following.

Claim. For all $x \in [-1, 1)$,

$$-x - \frac{x^2}{2} \leq \ln(1-x).$$

Proof. This one is even easier: since $\ln(1-x) = -x - x^2/2 - x^3/3 - \dots$. ⊛

Hence, if $\delta \in (0,1]$,[a] we have

$$\ln \mathbb{P}(S_N \leq (1-\delta)\mu) \leq \mu(-\delta - (1-\delta)\ln(1-\delta)) \leq -\mu\delta - \mu(1-\delta)\left(-\delta - \frac{\delta^2}{2}\right) \leq -\frac{\mu\delta^2}{2}.$$

Combining two tails, we then see that

$$\begin{aligned}\mathbb{P}(|S_N - \mu| > \delta\mu) &\leq \mathbb{P}(S_N \geq (1+\delta)\mu) + \mathbb{P}(S_N \leq (1-\delta)\mu) \\ &\leq \exp\left(-\frac{\mu\delta^2}{3}\right) + \exp\left(-\frac{\mu\delta^2}{2}\right) \\ &\leq 2\exp\left(-\frac{\mu\delta^2}{3}\right),\end{aligned}$$

which almost complete the proof for $c = 1/3$. ⊛

[a] When $\delta = 1$, $\ln \mathbb{P}(S_N \leq (1-\delta)\mu) \leq -\frac{\mu\delta^2}{2}$ holds trivially since $\mathbb{P}(S_N = 0) \leq \exp(-\mu/2)$.

Problem (Exercise 2.3.6). Let $X \sim \operatorname{Pois}(\lambda)$. Show that for $t \in (0, \lambda]$, we have

$$\mathbb{P}(|X - \lambda| \geq t) \leq 2\exp\left(-\frac{ct^2}{\lambda}\right).$$

Answer. Fix some $t =: \delta\lambda \in (0, \lambda]$ for some $\delta \in (0, 1]$ first. Consider a series of independent Bernoulli random variables $X_{N,i}$ for a fixed N such that the Poisson limit theorem applies to approximate $X \sim \operatorname{Pois}(\lambda)$, i.e., as $N \to \infty$, $\max_{i \leq N} p_{N,i} \to 0$ and $\lambda_N := \mathbb{E}[S_N] \to \lambda < \infty$, $S_N \to \operatorname{Pois}(\lambda)$. From multiplicative form of Chernoff's inequality, for $t_N := \delta\lambda_N$,

$$\mathbb{P}(|S_N - \lambda_N| \geq t_N = \delta\lambda_N) \leq 2\exp\left(-\frac{ct_N^2}{\lambda_N}\right).$$

It then follows that from the Poisson limit theorem,

$$\mathbb{P}(|X - \lambda| \geq t) = \lim_{N\to\infty} \mathbb{P}(|S_N - \lambda_N| \geq t_N) = \lim_{N\to\infty} 2\exp\left(-\frac{ct_N^2}{\lambda_N}\right) = 2\exp\left(-\frac{ct^2}{\lambda}\right)$$

since $t_N = \delta\lambda_N \to \delta\lambda = t$. ⊛

Problem (Exercise 2.3.8). Let $X \sim \operatorname{Pois}(\lambda)$. Show that, as $\lambda \to \infty$, we have

$$\frac{X - \lambda}{\sqrt{\lambda}} \overset{D}{\to} \mathcal{N}(0, 1).$$

Answer. Since $X := \sum_{i=1}^{\lambda} X_i \sim \operatorname{Pois}(\lambda)$ if $X_i \overset{\text{i.i.d.}}{\sim} \operatorname{Pois}(1)$ for all i, from Lindeberg-Lévy central limit theorem, we have

$$\frac{X - \mathbb{E}[X]}{\sqrt{\operatorname{Var}[X]}} = \frac{X - \lambda}{\sqrt{\lambda}} \overset{d}{\to} \mathcal{N}(0, 1)$$

as $\mathbb{E}[X_i] = \operatorname{Var}[X_i] = 1$. ⊛

2.4 Application: degrees of random graphs

Problem (Exercise 2.4.2). Consider a random graph $G \sim G(n, p)$ with expected degrees $d = O(\log n)$. Show that with high probability (say, 0.9), all vertices of G have degrees $O(\log n)$.

Answer. Since $d = O(\log n)$, there exists an absolute constant $M > 0$ such that $d = (n-1)p \leq M \log n$ for all large enough n. Now, consider some $C > 0$ such that $eM/C =: \alpha < 1$. From Chernoff's inequality,

$$\mathbb{P}(d_i \geq C\log n) \leq e^{-d}\left(\frac{ed}{C\log n}\right)^{C\log n} \leq e^{-d}\left(\frac{eM}{C}\right)^{C\log n} \leq \alpha^{C\log n}.$$

Hence, from union bound, we have

$$\mathbb{P}(\forall i\colon d_i \leq C\log n) \geq 1 - n\alpha^{C\log n},$$

which can be arbitrarily close to 1 as C is sufficiently large. ⊛

Problem (Exercise 2.4.3). Consider a random graph $G \sim G(n, p)$ with expected degrees $d = O(1)$. Show that with high probability (say, 0.9), all vertices of G have degrees

$$O\left(\frac{\log n}{\log\log n}\right).$$

Answer. Since now $d = (n-1)p \leq M$ for some absolute constant $M > 0$ for all large n, from Chernoff's inequality,

$$\mathbb{P}\left(d_i \geq C\frac{\log n}{\log\log n}\right) \leq e^{-d}\left(\frac{ed}{C\frac{\log n}{\log\log n}}\right)^{C\frac{\log n}{\log\log n}} \leq e^{-d}\left(\frac{eM\log\log n}{C\log n}\right)^{C\frac{\log n}{\log\log n}}$$

for some $C > 0$. This implies that

$$\mathbb{P}\left(\forall i\colon d_i \leq C\frac{\log n}{\log\log n}\right) \geq 1 - ne^{-d}\left(\frac{eM\log\log n}{C\log n}\right)^{C\frac{\log n}{\log\log n}}.$$

Now, considering $C = M$, we have

$$ne^{-d}\left(\frac{eM\log\log n}{C\log n}\right)^{C\frac{\log n}{\log\log n}} \leq ne^{-d}\left(\frac{e\log\log n}{\log n}\right)^{M\frac{\log n}{\log\log n}}.$$

Taking logarithm, we observe that

$$\begin{aligned}
&\log n - d + M\frac{\log n}{\log\log n}(1 + \log\log\log n - \log\log n) \\
&= (1-M)\log n - d + M\frac{\log n}{\log\log n}(1 + \log\log\log n) \\
&= \left[1 - M\left(1 + \frac{1}{\log\log n} + \frac{\log\log\log n}{\log\log n}\right)\right]\log n - d \to -\infty
\end{aligned}$$

as $n \to \infty$, i.e.,

$$ne^{-d}\left(\frac{eM\log\log n}{C\log n}\right)^{C\frac{\log n}{\log\log n}} \to 0,$$

which is what we want to prove. ⊛

Problem (Exercise 2.4.4)**.** Consider a random graph $G \sim G(n,p)$ with expected degrees $d = o(\log n)$. Show that with high probability, (say, 0.9), G has a vertex with degree $10d$.

Answer. Omit. ⊛

Problem (Exercise 2.4.5)**.** Consider a random graph $G \sim G(n,p)$ with expected degrees $d = O(1)$. Show that with high probability, (say, 0.9), G has a vertex with degree

$$\Omega\left(\frac{\log n}{\log\log n}\right).$$

Answer. Firstly, note that the question is ill-defined in the sense that if $d = (n-1)p = O(1)$, it can be $d = 0$ (with $p = 0$), which is impossible to prove the claim. Hence, consider the non-degenerate case, i.e., $d = \Theta(1)$.

We want to prove that there exists some absolute constant $C > 0$ such that with high probability G has a vertex with degree at least $C\log n/\log\log n$. First, consider separate the graph randomly into two parts A, B, each of size $n/2$. It's then easy to see by dropping every inner edge in A and B, the graph becomes bipartite such that now A and B forms independent sets. Consider working on this new graph (with degree denoted as d'), we have

$$\begin{aligned}
\mathbb{P}(d_i' = k) = \binom{n/2}{k}\left(\frac{d}{n-1}\right)^k\left(1 - \frac{d}{n-1}\right)^{n/2-k} &\geq \left(\frac{n}{2k}\right)^k \cdot \frac{d^k}{n^k} \cdot e^{-d} \\
&= d^k n^{-k}\left(\frac{n}{2k}\right)^k e^{-d} = \left(\frac{d}{2k}\right)^k e^{-d}.
\end{aligned}$$

Let $k = C\log n/\log\log n$ such that $d/2k > 1/\log n$ for large enough n,[a] we have

$$\mathbb{P}\left(d_i' = \frac{C\log n}{\log\log n}\right) \geq e^{-d}\left(\frac{d}{2k}\right)^k \geq e^{-d}(\log n)^{-k} = \exp(-d - k\log\log n)$$
$$= \exp(-d - C\log n) = e^{-d}n^{-C}.$$

Let this probability be q, and focus on A. We can then define $X_i = \mathbb{1}_{d_i'=k}$ for $i \in A$, and note that X_i are all independent as A being an independent set. Then, the number of vertices in A, denoted as X, with degree exactly k follows $\mathrm{Bin}(n/2, q)$ with $X = \sum_{i\in A} X_i$ and mean $nq/2$, variance $nq(1-q)/2$. From Chebyshev's inequality,

$$\mathbb{P}(X=0) \leq \mathbb{P}(|X-\mu| \geq \mu) \leq \frac{\sigma^2}{\mu^2} = \frac{nq(1-q)/2}{(nq/2)^2} = 2\frac{1-q}{nq} \leq \frac{2}{nq} \leq \frac{2}{ne^{-d}n^{-C}} = \frac{2e^d}{n^{1-C}}.$$

Now, by setting $C < 1$, say $1/2$, then

$$\mathbb{P}(X=0) \leq 2e^d n^{-1/2} \to 0$$

as $n \to \infty$, which means $\mathbb{P}(X \geq 1) \to 1$, i.e., with probability 1, there are at least one point with degree $\log n/2\log\log n$. Now, by considering the deleting edges in the beginning, we conclude that there will be a vertex with degree

$$\Omega\left(\frac{\log n}{\log\log n}\right)$$

with overwhelming probability. ⊛

[a] Since this is equivalent as $k < d\log n/2$. As k has a $\log\log n \to \infty$ factor in the denominator, the claim holds.

Week 5: Sub-Gaussian Random Variables

16 Feb. 2024

2.5 Sub-gaussian distributions

Problem (Exercise 2.5.1). Show that for each $p \geq 1$, the random variable $X \sim \mathcal{N}(0,1)$ satisfies

$$\|X\|_{L^p} = (\mathbb{E}[|X|^p])^{1/p} = \sqrt{2}\left(\frac{\Gamma((1+p)/2)}{\Gamma(1/2)}\right)^{1/p}.$$

Deduce that

$$\|X\|_{L^p} = O(\sqrt{p}) \text{ as } p \to \infty.$$

Answer. We see that for $p \geq 1$, we have

$$(\mathbb{E}[|X|^p])^{1/p} = \left(\int_{-\infty}^{\infty} |x|^p \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,\mathrm{d}x\right)^{1/p} = \left(2\int_0^{\infty} |x|^p \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,\mathrm{d}x\right)^{1/p}$$

from the symmetry around 0. Next, consider a change of variable $x^2 =: u$, we have

$$= \left(2\frac{1}{\sqrt{2\pi}}\int_0^{\infty} u^{p/2}e^{-u/2}\frac{1}{2\sqrt{u}}\,\mathrm{d}u\right)^{1/p} = \left(\frac{1}{\sqrt{2\pi}}\int_0^{\infty} u^{(p-1)/2}e^{-u/2}\,\mathrm{d}u\right)^{1/p}$$

with another change of variable $u/2 =: t$,

$$= \left(\frac{1}{\sqrt{2\pi}}\int_0^{\infty} (2t)^{(p-1)/2}e^{-t}2\,\mathrm{d}t\right)^{1/p} = \left(\frac{1}{\sqrt{2\pi}} \cdot 2^{(p-1)/2} \cdot 2\int_0^{\infty} t^{(p-1)/2}e^{-t}\,\mathrm{d}t\right)^{1/p}$$
$$= \left(\frac{1}{\sqrt{2\pi}}2^{(p+1)/2}\Gamma\left(\frac{p+1}{2}\right)\right)^{1/p} = \left(\frac{1}{\sqrt{2}}\sqrt{2}^{p+1}\frac{\Gamma((p+1)/2)}{\Gamma(1/2)}\right)^{1/p}$$

as $\Gamma(1/2) = \sqrt{\pi}$, we finally have

$$= \sqrt{2}\left(\frac{\Gamma((p+1)/2)}{\Gamma(1/2)}\right)^{1/p},$$

where we recall that
$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,\mathrm{d}t.$$

To show that $\|X\|_{L^p} = O(\sqrt{p})$ as $p\to\infty$, we first note the following.

Lemma 2.5.1. We have that for $p \geq 1$,
$$\Gamma\left(\frac{1+p}{2}\right) = \begin{cases} 2^{-p/2}\sqrt{\pi}(p-1)!!, & \text{if } p \text{ is even;} \\ 2^{-(p-1)/2}(p-1)!!, & \text{if } p \text{ is odd.} \end{cases}$$

Proof. Consider the Legendre duplication formula, i.e.,
$$\Gamma(z)\Gamma(z+1/2) = 2^{1-2z}\sqrt{\pi}\Gamma(2z).$$

We see that for p being even, $(1+p)/2 = p/2 + 1/2$, by letting $z \coloneqq p/2 \in \mathbb{N}$,
$$\begin{aligned}\Gamma((1+p)/2) = \frac{2^{1-p}\sqrt{\pi}\Gamma(p)}{\Gamma(p/2)} &= 2^{1-p}\sqrt{\pi}\frac{(p-1)!}{(p/2-1)!} \\ &= 2^{1-p}\sqrt{\pi}\frac{(p-1)!}{(1/2)^{p/2-1}(p-2)!!} = 2^{-p/2}\sqrt{\pi}(p-1)!!.\end{aligned}$$

For odd p, recall the identity $\Gamma(z+1) = z\Gamma(z)$. We then have
$$\begin{aligned}\Gamma((1+p)/2) =& \frac{p-1}{2}\cdot\Gamma((p-1)/2) \\ =& \frac{(p-1)(p-3)}{2^2}\cdot\Gamma((p-3)/2) \\ &\vdots \\ =& \frac{(p-1)(p-3)\dots(p-(p-2))}{2^{(p-1)/2}}\cdot\Gamma(1) && 2 = (p-(p-2)) \\ =& 2^{-(p-1)/2}(p-1)(p-3)\dots(2) \\ =& 2^{-(p-1)/2}(p-1)!!.\end{aligned}$$

■

We then see that as $p\to\infty$,
$$\|X\|_{L^p} = \sqrt{2}\left(\frac{\Gamma((1+p)/2)}{\Gamma(1/2)}\right)^{1/p} \lesssim ((p-1)!!)^{1/p} = O(\sqrt{p!}^{1/p}) = O(\sqrt{p}).$$

⊛

Problem (Exercise 2.5.4). Show that the condition $\mathbb{E}[X] = 0$ is necessary for property v to hold.

Answer. Since if $\mathbb{E}[\exp(\lambda X)] \leq \exp(K_5^2\lambda^2)$ for all $\lambda\in\mathbb{R}$, we see that from Jensen's inequality,
$$\exp(\mathbb{E}[\lambda X]) \leq \mathbb{E}[\exp(\lambda X)] \leq \exp(K_5^2\lambda^2),$$
i.e.,
$$\lambda\mathbb{E}[X] \leq K_5^2\lambda^2.$$
Since this holds for every $\lambda\in\mathbb{R}$, if $\lambda > 0$, $\mathbb{E}[X] \leq K_5^2\lambda$; on the other hand, if $\lambda < 0$, $\mathbb{E}[X] \geq K_5^2\lambda$. In either case, as $\lambda \to 0$ (from both sides, respectively), $0 \leq \mathbb{E}[X] \leq 0$, hence $\mathbb{E}[X] = 0$. ⊛

Problem (Exercise 2.5.5). (a) Show that if $X \sim \mathcal{N}(0,1)$, the function $\lambda \mapsto \mathbb{E}[\exp(\lambda^2 X^2)]$ is only finite in some bounded neighborhood of zero.

(b) Suppose that some random variable X satisfies $\mathbb{E}[\exp(\lambda^2 X^2)] \leq \exp(K\lambda^2)$ for all $\lambda \in \mathbb{R}$ and some constant K. Show that X is a bounded random variable, i.e., $\|X\|_\infty < \infty$.

Answer. (a) If $X \sim \mathcal{N}(0,1)$, we see that

$$\mathbb{E}[\exp(\lambda^2 X^2)] = \int_{-\infty}^{\infty} \exp(\lambda^2 x^2) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left((\lambda^2 - 1/2)x^2\right) \mathrm{d}x.$$

It's obvious that if $\lambda^2 - 1/2 \geq 0$, the above integral doesn't converge simply because $e^{\epsilon x^2}$ for any $\epsilon \geq 0$ is unbounded. On the other hand, if $\lambda^2 - 1/2 < 0$, then this is just a (scaled) Gaussian integral, which converges. Hence, this function is only finite in $\lambda \in (-1/\sqrt{2}, 1/\sqrt{2})$.

(b) Simply because that for any t, we have that for any λ,

$$\mathbb{P}(|X| > t) \leq \frac{\mathbb{E}[\exp(\lambda^2 X^2)]}{\exp(\lambda^2 t^2)} \leq \frac{\exp(K\lambda^2)}{\exp(\lambda^2 t^2)} = \exp\left(\lambda^2 (K - t^2)\right).$$

Now, let's pick $t > \sqrt{K}$ (as K being a constant, t can be any constant greater than $t > \sqrt{K}$), so $\lambda^2(K - t^2) < 0$. By letting $\lambda \to \infty$, we see that $\mathbb{P}(|X| > t) = 0$, i.e., $\mathbb{P}(|X| \leq t) = 1$. Since we're in one-dimensional, $|X| = \|X\|_\infty$, hence we're done.

⊛

Problem (Exercise 2.5.7). Check that $\|\cdot\|_{\psi_2}$ is indeed a norm on the space of sub-gaussian random variables.

Answer. It's clear that $\|X\|_{\psi_2} = 0$ if and only if $X = 0$. Also, for any $\lambda > 0$, $\|\lambda X\|_{\psi_2} = \lambda \|X\|_{\psi_2}$ is obvious. Hence, we only need to verify triangle inequality, i.e., for any sub-gaussian random variables X and Y,

$$\|X + Y\|_{\psi_2} \leq \|X\|_{\psi_2} + \|Y\|_{\psi_2}.$$

Firstly, we observe that since $\exp(x)$ and x^2 are both convex (hence their composition),

$$\begin{aligned}\exp\left(\left(\frac{X+Y}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}}\right)^2\right) \leq & \frac{\|X\|_{\psi_2}}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}} \exp\left((X/\|X\|_{\psi_2})^2\right) \\ & + \frac{\|Y\|_{\psi_2}}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}} \exp\left((Y/\|Y\|_{\psi_2})^2\right).\end{aligned}$$

Then, by taking expectation on both sides,

$$\mathbb{E}\left[\exp\left(\left(\frac{X+Y}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}}\right)^2\right)\right] \leq 2\frac{\|X\|_{\psi_2}}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}} + 2\frac{\|Y\|_{\psi_2}}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}} = 2.$$

Now, we see that from the definition of $\|X+Y\|_{\psi_2}$ and $t \coloneqq \|X\|_{\psi_2} + \|Y\|_{\psi_2}$, the above implies

$$\|X + Y\|_{\psi_2} \leq \|X\|_{\psi_2} + \|Y\|_{\psi_2},$$

hence the triangle inequality is verified. ⊛

Problem (Exercise 2.5.9). Check that Poisson, exponential, Pareto and Cauchy distributions are not sub-gaussian.

Answer. Omit. ⊛

Problem (Exercise 2.5.10). Let $X_1, X_2, \ldots,$ be a sequence of sub-gaussian random variables, which are not necessarily independent. Show that

$$\mathbb{E}\left[\max_i \frac{|X_i|}{\sqrt{1+\log i}}\right] \leq CK,$$

where $K = \max_i \|X_i\|_{\psi_2}$. Deduce that for every $N \geq 2$ we have

$$\mathbb{E}\left[\max_{i \leq N} |X_i|\right] \leq CK\sqrt{\log N}.$$

Answer. Let $Y_i := |X_i|/K\sqrt{1+\log i}$ (which is always positive) for all $i \geq 1$. Then for all $t \geq 0$,

$$\begin{aligned}\mathbb{P}(Y_i \geq t) &= \mathbb{P}\left(\frac{|X_i|}{K\sqrt{1+\log i}} \geq t\right) \\ &= \mathbb{P}\left(|X_i| \geq tK\sqrt{1+\log i}\right) \\ &\leq 2\exp\left(-\frac{ct^2K^2(1+\log i)}{\|X_i\|_{\psi_2}^2}\right) \leq 2\exp\left(-ct^2(1+\log i)\right) = 2(ei)^{-ct^2}\end{aligned}$$

as $K := \max_i \|X_i\|_{\psi_2}^2$. Then, our goal now is to show that $\mathbb{E}[\max_i Y_i] \leq C$ for some absolute constant C. Consider $t_0 := \sqrt{1/c}$, then we have

$$\begin{aligned}\mathbb{E}\left[\max_i Y_i\right] &= \int_0^\infty \mathbb{P}\left(\max_i Y_i \geq t\right) \mathrm{d}t \\ &\leq \int_0^{t_0} \mathbb{P}\left(\max_i Y_i \geq t\right) \mathrm{d}t + \int_{t_0}^\infty \sum_{i=1}^\infty \mathbb{P}(Y_i \geq t)\,\mathrm{d}t && \text{union bound} \\ &\leq t_0 + \int_{t_0}^\infty \sum_{i=1}^\infty 2(ei)^{-ct^2}\,\mathrm{d}t \\ &\leq \sqrt{1/c} + 2\int_{t_0}^\infty e^{-ct^2} \sum_{i=1}^\infty i^{-2}\,\mathrm{d}t \\ &\leq \sqrt{1/c} + 2\cdot\frac{\pi^2}{6}\int_0^\infty e^{-ct^2}\,\mathrm{d}t = \sqrt{1/c} + \frac{\pi^2}{3}\cdot\frac{\sqrt{\pi}}{2\sqrt{c}} = \frac{1+\frac{\pi^{5/2}}{6}}{\sqrt{c}} =: C.\end{aligned}$$

Finally, for every $N \geq 2$,

$$\mathbb{E}\left[\max_{i \leq N} \frac{|X_i|}{\sqrt{1+\log N}}\right] \leq \mathbb{E}\left[\max_{i \leq N} \frac{|X_i|}{\sqrt{1+\log i}}\right] \leq \mathbb{E}\left[\max_i \frac{|X_i|}{\sqrt{1+\log i}}\right] \leq CK,$$

i.e., $\mathbb{E}[\max_{i \leq N} |X_i|] \leq CK\sqrt{1+\log N} \leq CK\sqrt{2\log N}$ for all $N \geq 2$. By letting $C' := \sqrt{2}C$,

$$\mathbb{E}\left[\max_{i \leq N} |X_i|\right] \leq C'K\sqrt{\log N},$$

which is exactly what we want. ⊛

Problem (Exercise 2.5.11)**.** Show that the bound in Exercise 2.5.10 is sharp. Let $X_1, X_2, \ldots, X_N$ be independent $\mathcal{N}(0,1)$ random variables. Prove that

$$\mathbb{E}\left[\max_{i \leq N} X_i\right] \geq c\sqrt{\log N}.$$

Answer. Again, let's first write

$$\mathbb{E}\left[\max_{i \leq N} X_i\right] = \int_0^\infty \mathbb{P}\left(\max_{i \leq N} X_i \geq t\right) \mathrm{d}t,$$

and observe that for any $t \geq 0$,

$$\begin{aligned}\mathbb{P}(X_i \geq t) &= \int_t^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left(-\frac{(x+t)^2}{2}\right) \mathrm{d}x && x \leftarrow x+t \\ &\geq \frac{1}{\sqrt{2\pi}} \int_0^1 \exp\left(-\frac{(x+t)^2}{2}\right) \mathrm{d}x \\ &\geq Ce^{-t^2}\end{aligned}$$

for some constant $C > 0$. Since X_i's are i.i.d.,

$$\mathbb{P}\left(\max_{i \leq N} X_i \geq t\right) = 1 - \left(\mathbb{P}(X_1 < t)\right)^N = 1 - \left(1 - \mathbb{P}(X_1 \geq t)\right)^N,$$

so

$$\begin{aligned}\mathbb{E}\left[\max_{i \leq N} X_i\right] &= \int_0^\infty 1 - \left(1 - \mathbb{P}(X_1 \geq t)\right)^N \mathrm{d}t \\ &\geq \int_0^\infty 1 - (1 - Ce^{-t^2})^N \mathrm{d}t \\ &= \sqrt{\log N} \int_0^\infty 1 - \left(1 - \frac{C}{N^{u^2}}\right)^N \mathrm{d}u. && t =: \sqrt{\log N} u\end{aligned}$$

Finally, as the final integral can be further bounded below by some absolute constant c depending only on C, hence we obtain the desired result. ⊛

Week 6: Hoeffding's and Khintchine's Inequalities

2.6 General Hoeffding's and Khintchine's inequalities

21 Feb. 2024

Problem (Exercise 2.6.4). Deduce Hoeffding's inequality for bounded random variables (Theorem 2.2.6) from Theorem 2.6.3, possibly with some absolute constant instead of 2 in the exponent.

Answer. Omit. ⊛

Problem (Exercise 2.6.5). Let $X_1, \ldots, X_N$ be independent sub-gaussian random variables with zero means and unit variances, and let $a = (a_1, \ldots, a_N) \in \mathbb{R}^N$. Prove that for every $p \in [2, \infty)$ we have

$$\left(\sum_{i=1}^N a_i^2\right)^{1/2} \leq \left\lVert \sum_{i=1}^N a_i X_i \right\rVert_{L^p} \leq CK\sqrt{p} \left(\sum_{i=1}^N a_i^2\right)^{1/2}$$

where $K = \max_i \lVert X_i \rVert_{\psi_2}$ and C is an absolute constant.

Answer. From Jensen's inequality,

$$\left\lVert \sum_{i=1}^N a_i X_i \right\rVert_{L^p} \geq \left\lVert \sum_{i=1}^N a_i X_i \right\rVert_{L^2} = \left[\mathbb{E}\left[\left(\sum_{i=1}^N a_i X_i\right)^2\right]\right]^{1/2}.$$

Then, observe that since $\mathbb{E}[X_i] = 0$,

$$\operatorname{Var}\left[\sum_{i=1}^N a_i X_i\right] = \mathbb{E}\left[\left(\sum_{i=1}^N a_i X_i\right)^2\right] - \left(\mathbb{E}\left[\sum_{i=1}^N a_i X_i\right]\right)^2 = \mathbb{E}\left[\left(\sum_{i=1}^N a_i X_i\right)^2\right],$$

and at the same time, as $\operatorname{Var}[X_i] = 1$, $\operatorname{Var}\left[\sum_{i=1}^N a_i X_i\right] = \sum_{i=1}^N a_i^2 \operatorname{Var}[X_i] = \sum_{i=1}^N a_i^2 = \|a\|^2$, hence we have

$$\left\| \sum_{i=1}^N a_i X_i \right\|_{L^p} \geq \left[\|a\|^2\right]^{1/2} = \|a\|,$$

which is the desired lower-bound. For the upper-bound, we see that

$$\begin{aligned}\left\| \sum_{i=1}^N a_i X_i \right\|_{L_p}^2 &\leq C^2 \sqrt{p}^2 \left\| \sum_{i=1}^N a_i X_i \right\|_{\psi_2}^2 \\ &\leq C' p \sum_{i=1}^N \|a_i X_i\|_{\psi_2}^2 = C'' p \sum_{i=1}^N a_i^2 \|X_i\|_{\psi^2}^2 \leq C'' K^2 p \|a\|^2,\end{aligned}$$

where C, C', C'' are all absolute constant (might depend on each other). Taking square root on both sides, we obtain the desired result. ⊛

Problem (Exercise 2.6.6)**.** Show that in the setting of Exercise 2.6.5, we have

$$c(K) \left(\sum_{i=1}^N a_i^2 \right)^{1/2} \leq \left\| \sum_{i=1}^N a_i X_i \right\|_{L^1} \leq \left(\sum_{i=1}^N a_i^2 \right)^{1/2}.$$

Here $K g \max_i \|X_i\|_{\psi_2}$ and $c(K) > 0$ is a quantity which may depend only on K.

Answer. Skip, as this is a special case of Exercise 2.6.7. ⊛

Problem (Exercise 2.6.7)**.** State and prove a version of Khintchine's inequality for $p \in (0, 2)$.

Answer. The Khintchine's inequality for $p \in (0, 2)$ can be stated as

$$c(K, p) \left(\sum_{i=1}^N a_i^2 \right)^{1/2} \leq \left\| \sum_{i=1}^N a_i X_i \right\|_{L^p} \leq \left(\sum_{i=1}^N a_i^2 \right)^{1/2}.$$

Here $K = \max_i \|X_i\|_{\psi_2}$ and $c(K, p) > 0$ is a quantity which depends on K and p. We first recall the generalized Hölder inequality.

Theorem 2.6.1 (Generalized Hölder inequality)**.** For $1/p + 1/q = 1/r$ where $p, q \in (0, \infty]$,

$$\|fg\|_{L^r} \leq \|f\|_{L^p} \|g\|_{L^q}.$$

Proof. The classical case is when $r = 1$. By considering $|f|^r \in L^{p/r}$ and $|g|^r \in L^{q/r}$, $r/p + r/q = 1$. Then the standard Hölder inequality implies

$$\begin{aligned}\|fg\|_{L^r}^r = \int |fg|^r = \||fg|^r\|_{L^1} &\leq \||f|^r\|_{L^{p/r}} \||g|^r\|_{L^{q/r}} \\ &= \left(\int (|f|^r)^{p/r} \right)^{r/p} \left(\int (|g|^r)^{q/r} \right)^{r/q} = \|f\|_{L^p}^r \|g\|_{L^q}^r,\end{aligned}$$

implying the result. ■

Now, take $r = 2$, $p = q = 4$, we get

$$\|XY\|_{L^2} \leq \|X\|_{L^4} \|Y\|_{L^4} = \left(\mathbb{E}[|X|^4]\right)^{1/4} \left(\mathbb{E}[|Y|^4]\right)^{1/4}.$$

Let $X = |Z|^{p/4}$ and $Y = |Z|^{(4-p)/4}$, we see that

$$\|Z\|_{L^2} \leq (\mathbb{E}[|Z|^p])^{1/4} \left(\mathbb{E}[|Z|^{4-p}]\right)^{1/4} = \|Z\|_{L^p}^{p/4} \|Z\|_{L^{4-p}}^{(4-p)/4},$$

implying

$$\|Z\|_{L^p} \geq \left(\frac{\|Z\|_{L^2}}{\|Z\|_{L^{4-p}}^{(4-p)/4}} \right)^{4/p} = \frac{\|Z\|_{L^2}^{4/p}}{\|Z\|_{L^{4-p}}^{(4-p)/p}}.$$

Finally, by letting $Z = \sum_{i=1}^N a_i X_i$,

$$\left\| \sum_{i=1}^N a_i X_i \right\|_{L^p} \geq \left\| \sum_{i=1}^N a_i X_i \right\|_{L^2}^{4/p} \Bigg/ \left\| \sum_{i=1}^N a_i X_i \right\|_{L^{4-p}}^{(4-p)/p}.$$

Observe that from Exercise 2.6.5:

- $\|\sum_{i=1}^N a_i X_i\|_{L^2} = \|a\|$;
- $\|\sum_{i=1}^N a_i X_i\|_{L^{4-p}} \leq CK\sqrt{4-p}\|a\|$ (as $4-p > 2$ from $p \in (0,2)$),

hence

$$\left\| \sum_{i=1}^N a_i X_i \right\|_{L^p} \geq \|a\|^{4/p} \Big/ \left(CK\sqrt{4-p}\|a\|\right)^{(4-p)/p} = \left(CK\sqrt{4-p}\right)^{-\frac{p}{4-p}} \|a\|.$$

Hence, we see that by letting $c(K,p) := (CK\sqrt{4-p})^{-p/(4-p)}$, the lower-bound is established. The upper-bound is essentially the same as Exercise 2.6.5 (in there we use have the lower-bound since $p \geq 2$), where this time we use $\|\cdot\|_{L^p} \leq \|\cdot\|_{L^2}$ since $p \leq 2$.[a] Hence, we're done. ⊛

[a] Note that although $\|\cdot\|_{L^p}$ for $p \in [0,1)$ is not a norm, this inequality still holds.

Remark. Exercise 2.6.6 is just a special case with $c(K,1) = (CK\sqrt{3})^{-1/3}$.

Problem (Exercise 2.6.9). Show that unlike (2.19), the centering inequality in Lemma 2.6.8 does not hold with $C = 1$.

Answer. Consider the random variable $X := \sqrt{\log 2} \cdot \epsilon$ where ϵ is a Rademacher random variable with parameter p, i.e.,

$$X = \begin{cases} \sqrt{\log 2}, & \text{w.p. } p; \\ -\sqrt{\log 2}, & \text{w.p. } 1-p. \end{cases}$$

Since $\mathbb{E}[\exp(X^2)] = 2$, we know that $\|X\|_{\psi_2}$ is exactly 1. We now want to show that $\|X - \mathbb{E}[X]\|_{\psi_2} > \|X\|_{\psi_2} = 1$ for some p. It amounts to show that $\mathbb{E}[\exp(|X - \mathbb{E}[X]|^2)] > 2$. Now, we know that $\mathbb{E}[X] = \sqrt{\log 2}(2p-1)$, and hence

$$X - \mathbb{E}[X] = \begin{cases} 2(1-p)\sqrt{\log 2}, & \text{w.p. } p; \\ -2p\sqrt{\log 2}, & \text{w.p. } 1-p. \end{cases}$$

Hence, we have that

$$\mathbb{E}[\exp(|X - \mathbb{E}[X]|^2)] = p \cdot 2^{4(1-p)^2} + (1-p)2^{4p^2}.$$

A quick numerical optimization gives the desired result with $p \approx 0.236$. ⊛

Week 7: Sub-Exponential Random Variables

2.7 Sub-exponential distributions

1 Mar. 2024

Problem (Exercise 2.7.2). Prove the equivalence of properties a-d in Proposition 2.7.1 by modifying the proof of Proposition 2.5.2.

Answer. This is a special case of Exercise 2.7.3 with $\alpha = 1$. ⊛

Problem (Exercise 2.7.3). More generally, consider the class of distributions whose tail decay is of the type $\exp(-ct^\alpha)$ or faster. Here $\alpha = 2$ corresponds to sub-gaussian distributions, and $\alpha = 1$, to sub-exponential. State and prove a version of Proposition 2.7.1 for such distributions.

Answer. The generalized version of Proposition 2.7.1 is known to be the so-called *Sub-Weibull distributions* [Vla+20]: Let X be a random variable. Then the following properties are equivalent; the parameters $K_i > 0$ appearing in these properties differ from each other by at most an absolute constant factor.

(a) The tails of X satisfy

$$\mathbb{P}(|X| \geq t) \leq 2\exp(-t^\alpha / K_1) \text{ for all } t \geq 0.$$

(b) The moments of X satisfy

$$\|X\|_{L^p} = (\mathbb{E}[|X|^p])^{1/p} \leq K_2 p^{1/\alpha} \text{ for all } p \geq 1.$$

(c) The MGF of $|X|$ satisfies

$$\mathbb{E}[\exp(\lambda^\alpha |X|^\alpha)] \leq \exp(\lambda^\alpha K_3^\alpha) \text{ for all } \lambda \text{ such that } 0 \leq \lambda \leq \frac{1}{K_3}.$$

(d) The MGF of $|X|$ is bounded at some point, namely

$$\mathbb{E}[\exp(|X|^\alpha / K_4^\alpha)] \leq 2.$$

Claim. (a) ⇒ (b)

Proof. Without loss of generality, let $K_1 = 1$. Then, we have

$$\begin{aligned}
\|X\|_{L^p}^p &= \int_0^\infty \mathbb{P}(|X|^p \geq t)\,\mathrm{d}t \\
&= \int_0^\infty p u^{p-1} \mathbb{P}(|X| \geq u)\,\mathrm{d}u && u := t^{1/p} \\
&\leq 2p \int_0^\infty u^{p-1} e^{-u^\alpha}\,\mathrm{d}u && \text{from our assumption} \\
&= \frac{2p}{\alpha} \int_0^\infty t^{p/\alpha - 1} e^{-t}\,\mathrm{d}t && t := u^\alpha \\
&= 2\frac{p}{\alpha}\Gamma(p/\alpha) = 2\Gamma(p/\alpha + 1) \lesssim (p/\alpha + 1)^{p/\alpha + 1}
\end{aligned}$$

for some constant C from Stirling's approximation. Hence,

$$\|X\|_{L^p} \lesssim \left(\frac{p}{\alpha} + 1\right)^{\frac{1}{\alpha} + \frac{1}{p}} = \left(\frac{p}{\alpha} + 1\right)^{\frac{1}{\alpha}} \left(\frac{p}{\alpha} + 1\right)^{\frac{1}{p}} \lesssim p^{1/\alpha}$$

as we desired. ⊛

Claim. (b) ⇒ (c)

Proof. Firstly, from Taylor's expansion, we have

$$\mathbb{E}[\exp(\lambda^\alpha |X|^\alpha)] = 1 + \sum_{k=1}^{\infty} \frac{\lambda^{\alpha k}\mathbb{E}[|X|^{\alpha k}]}{k!} \leq 1 + \sum_{k=1}^{\infty} \frac{\lambda^{\alpha k}\mathbb{E}[|X|^{\alpha k}]}{k!}.$$

From (b), when $\alpha k \geq 1$, we have $\mathbb{E}[|X|^{\alpha k}] \leq (K_2(\alpha k)^{1/\alpha})^{\alpha k} = K_2^{\alpha k}(\alpha k)^k$. On the other hand, for any given $\alpha > 0$, there are only finitely many $k \geq 1$ such that $\alpha k < 1$. Hence, there exists some $\widetilde{K}_2$ such that

$$\mathbb{E}[|X|^{\alpha k}] \leq \widetilde{K}_2^{\alpha k}(\alpha k)^k$$

for all $k \geq 1$. With $k! \geq (k/e)^k$ from Stirling's approximation, we further have

$$1 + \sum_{k=1}^{\infty} \frac{\lambda^{\alpha k}\mathbb{E}[|X|^{\alpha k}]}{k!} \leq 1 + \sum_{k=1}^{\infty} \frac{\lambda^{\alpha k}\widetilde{K}_2^{\alpha k}(\alpha k)^k}{(k/e)^k} = 1 + \sum_{k=1}^{\infty} \lambda^{\alpha k}\widetilde{K}_2^{\alpha k}(\alpha e)^k = 1 + \sum_{k=1}^{\infty} (\widetilde{K}_2^\alpha \lambda^\alpha \alpha e)^k.$$

Observe that if $0 < \widetilde{K}_2^\alpha \lambda^\alpha \alpha e < 1$, we then have

$$\mathbb{E}[\exp(\lambda^\alpha |X|^\alpha)] \leq 1 + \sum_{k=1}^{\infty} (\widetilde{K}_2^\alpha \lambda^\alpha \alpha e)^k = \frac{1}{1 - \widetilde{K}_2^\alpha \lambda^\alpha \alpha e}.$$

As $(1-x)e^{2x} \geq 1$ for all $x \in [0, 1/2]$, the above is further less than

$$\exp\left(2(\widetilde{K}_2\lambda)^\alpha \alpha e\right) = \exp\left(\left[(2\alpha e)^{1/\alpha}\widetilde{K}_2\right]^\alpha \lambda^\alpha\right).$$

By letting $K_3 \coloneqq (2\alpha e)^{1/\alpha}\widetilde{K}_2$, we have the desired result whenever $\widetilde{K}_2^\alpha \lambda^\alpha \alpha e < 1$, or equivalently,

$$0 < \lambda^\alpha < \frac{1}{\widetilde{K}_2^\alpha \alpha e} \Leftrightarrow 0 < \lambda < \frac{1}{\widetilde{K}_2(\alpha e)^{1/\alpha}}.$$

Hence, if $0 < \lambda \leq \frac{1}{\widetilde{K}_2(2\alpha e)^{1/\alpha}} = \frac{1}{K_3}$, the above is satisfied. ⊛

Claim. (c) ⇒ (d)

Proof. Assuming (c) holds, then (d) is obtained by taking $\lambda \coloneqq 1/K_4$ where $K_4 \coloneqq K_3(\ln 2)^{-1/\alpha}$. In this case, $\lambda = 1/K_3 \cdot (\ln 2)^{1/\alpha}$, hence

$$\mathbb{E}[\exp(\lambda^\alpha |X|^\alpha)] = \mathbb{E}[\exp(|X|^\alpha / K_4^\alpha)] \leq \exp(\lambda^\alpha K_3^\alpha)$$

for all $0 \leq \lambda = 1/K_4 \leq 1/K_3$ from (d) gives

$$\mathbb{E}[\exp(|X|^\alpha / K_4^\alpha)] \leq \exp\left(\ln 2 \cdot \frac{1}{K_3^\alpha} \cdot K_3^\alpha\right) = 2.$$

⊛

Claim. (d) ⇒ (a)

Proof. Let $K_4 = 1$ without loss of generality. Then, we have

$$\mathbb{P}(|X| \geq t) = \mathbb{P}(\exp(|X|^\alpha) \geq \exp(t^\alpha)) \leq \frac{\mathbb{E}[\exp(|X|^\alpha)]}{\exp(t^\alpha)} \leq 2\exp(-t^\alpha),$$

hence $K_1 \coloneqq 1$ proves the result. ⊛

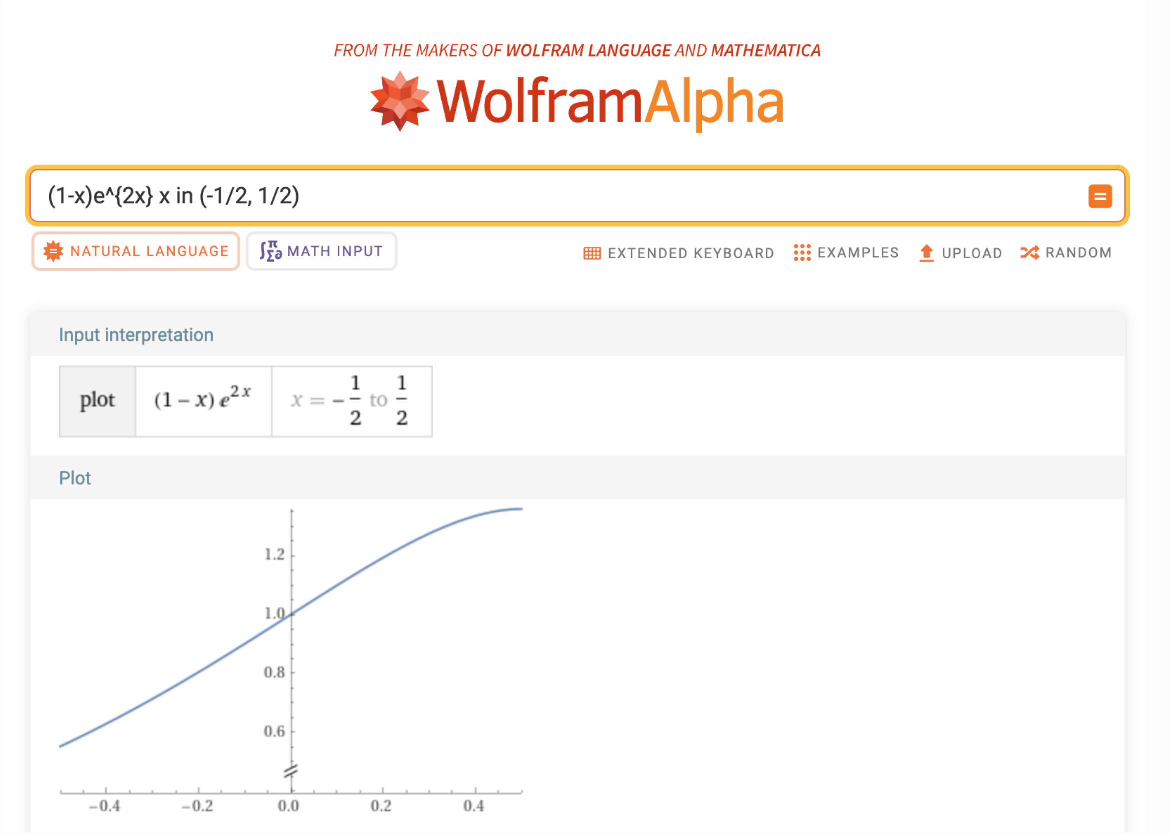
⊛

Problem (Exercise 2.7.4). Argue that the bound in property c can not be extended for all λ such that $|\lambda| \leq 1/K_3$.

Answer. It's easy to see that in the proof of Exercise 2.7.3, when we prove (b) $\Rightarrow$ (c), the condition for λ essentially comes from:

- whether $1+\sum_{k=1}^{\infty}(\widetilde{K}_2^{\alpha}\lambda^{\alpha}\alpha e)^k = 1+\sum_{k=1}^{\infty}(\widetilde{K}_2\lambda e)^k$ as $\alpha = 1$ converges; and
- the numerical inequality $(1-x)e^{2x} \geq 1$ for $x \in [0, 1/2]$ such that $x := \widetilde{K}_2\lambda e$.

For the first condition, we only need $|\widetilde{K}_2\lambda e| < 1$, hence we don't need positivity for λ at first; however, the second condition indeed requires $\lambda \geq 0$, and it's impossible to remove as this is tight.



$\circledast$

Problem (Exercise 2.7.10). Prove an analog of the Centering Lemma 2.6.8 for sub-exponential random variables X:

$$\lVert X - \mathbb{E}[X]\rVert_{\psi_1} \leq C\lVert X\rVert_{\psi_1}.$$

Answer. Since $\lVert\cdot\rVert_{\psi_2}$ is a norm, we have $\lVert X-\mathbb{E}[X]\rVert_{\psi_1} \leq \lVert X\rVert_{\psi_1} + \lVert\mathbb{E}[X]\rVert_{\psi_1}$ such that

$$\begin{aligned} \lVert\mathbb{E}[X]\rVert_{\psi_1} &\lesssim |\mathbb{E}[X]| && \lVert a\rVert_{\psi_1} = \inf_{t>0}\{\mathbb{E}[e^{|a|/t}] \leq 2\} \lesssim |a| \\ &\leq \mathbb{E}[|X|] && \text{Jensen's inequality} \\ &= \lVert X\rVert_{L^1} \lesssim \lVert X\rVert_{\psi_1} \end{aligned}$$

from Proposition 2.7.1 (b) with $p=1$, i.e.,

$$\lVert X\rVert_{L^1} \leq K_2 \cong \lVert X\rVert_{\psi_1}$$

since $K_i \cong \lVert X\rVert_{\psi_1} = K_4$. $\circledast$

Week 8: Bernstein's Inequality

6 Mar. 2024

Problem (Exercise 2.7.11). Show that $\|X\|_\psi$ is indeed a norm on the space L_ψ.

Answer. Clearly, $\|X\|_\psi \geq 0$. To check $\|X\|_\psi = 0$ if and only if $X = 0$ a.s., we first see that $\|0\|_\psi = 0$ as $\psi(0) = 0$. On the other hand, if $\|X\|_\psi = 0$, then by the monotone convergence theorem, we have

$$\begin{aligned}
1 \geq \lim_{t\to 0} \mathbb{E}[\psi(|X|/t)] &= \mathbb{E}\left[\lim_{t\to 0} \psi(|X|/t)\right] \\
&= \int_0^\infty \mathbb{P}\left(\lim_{t\to 0} \psi(|X|/t) > u\right) \mathrm{d}u \\
&= \mathbb{P}(|X| > 0) \int_0^\infty \mathbb{P}\left(\lim_{t\to 0} \psi(|X|/t) > u \mid |X| > 0\right) \mathrm{d}u \\
&= \mathbb{P}(|X| > 0) \int_0^\infty \mathrm{d}u \\
&= \infty \cdot \mathbb{P}(|X| > 0),
\end{aligned}$$

since if $|X| = 0$, $\psi(|X|/t) = \psi(0) = 0$ for all $t > 0$, and

$$\mathbb{P}\left(\lim_{t\to 0} \psi(|X|/t) > u \mid |X| > 0\right) = 1$$

since $\psi(x) \to \infty$ for $x \to \infty$, and in this case, $x = |X|/t$, which indeed goes to ∞ as $t \to 0$. Overall, this implies $\mathbb{P}(|X| > 0) = 0$, i.e., $X = 0$ almost surely, hence we conclude that $\|X\|_\psi = 0$ if and only if $X = 0$ a.s. The other two properties follows the same proof of Exercise 2.5.7. ⊛

2.8 Bernstein's inequality

Problem (Exercise 2.8.5). Let X be a mean-zero random variable such that $|X| \leq K$. Prove the following bound on the MGF of X:

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\big(g(\lambda)\mathbb{E}[X^2]\big) \text{ where } g(\lambda) = \frac{\lambda^2/2}{1 - |\lambda|K/3},$$

provided that $|\lambda| < 3/K$.

Answer. From the hint, we first check the following.

Claim. For all $|x| < 3$,

$$e^x \leq 1 + x + \frac{x^2/2}{1 - |x|/3}.$$

Proof. From Taylor's expansion,

$$e^x = 1 + x + \frac{x^2}{2}\sum_{k=0}^\infty \frac{x^k}{(2+k)!/2} \leq 1 + x + \frac{x^2}{2}\sum_{k=0}^\infty \frac{x^k}{3^k} = 1 + x + \frac{x^2/2}{1 - |x|/3}$$

where the last equality follows for all $|x| < 3$. ⊛

Now, for a random variable X such that $|X| \leq K$ and $|\lambda| < 3/K$, we have

$$\mathbb{E}[\exp(\lambda X)] \leq \mathbb{E}\left[1 + \lambda X + \frac{\lambda^2 X^2/2}{1 - |\lambda X|/3}\right] = 1 + \frac{\lambda^2 \mathbb{E}[X^2]/2}{1 - |\lambda|K/3} \leq \exp\left(\frac{\lambda^2 \mathbb{E}[X^2]/2}{1 - |\lambda|K/3}\right),$$

where we let $x := \lambda X$ and apply the claim. Finally, note that the right-hand side is exactly $\exp\big(g(\lambda)\mathbb{E}[X^2]\big)$, we're done. ⊛

Problem (Exercise 2.8.6). Deduce Theorem 2.8.4 from the bound in Exercise 2.8.5.

Answer. From Markov's inequality, for every $t \geq 0$,

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^N X_i \geq t\right) &\leq \inf_{\lambda>0} \frac{\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^N X_i\right)\right]}{\exp(\lambda t)} \\ &= \inf_{\lambda>0} e^{-\lambda t} \prod_{i=1}^N \mathbb{E}[\exp(\lambda X_i)] \leq \inf_{\lambda>0} e^{-\lambda t} \exp\left(g(\lambda) \sum_{i=1}^N \mathbb{E}[X_i^2]\right)\end{aligned}$$

from Exercise 2.8.5, if $|\lambda| < 3/K$. Denote $\sigma^2 = \sum_{i=1}^N \mathbb{E}[X_i^2]$, we further have

$$\mathbb{P}\left(\sum_{i=1}^N X_i \geq t\right) \leq \inf_{\lambda>0} \exp\left(-\lambda t + g(\lambda)\sigma^2\right).$$

Let $0 \leq \lambda = \frac{t}{\sigma^2 + tK/3} < 3/K$, we see that

$$\mathbb{P}\left(\sum_{i=1}^N X_i \geq t\right) \leq \exp\left(-\frac{t^2}{\sigma^2 + tK/3} + \frac{\sigma^2\lambda^2/2}{1 - |\lambda|K/3}\right) = \exp\left(-\frac{t^2/2}{\sigma^2 + tK/3}\right).$$

Applying the same argument for $-X_i$, we get

$$\mathbb{P}\left(\left|\sum_{i=1}^N X_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right).$$

⊛

Chapter 3

Random vectors in high dimensions

Week 9: Concentration Inequalities of Random Vectors

15 Mar. 2024

3.1 Concentration of the norm

Problem (Exercise 3.1.4). (a) Deduce from Theorem 3.1.1 that

$$\sqrt{n} - CK^2 \leq \mathbb{E}[\|X\|_2] \leq \sqrt{n} + CK^2.$$

(b) Can CK^2 be replaced by $o(1)$, a quantity that vanishes as $n \to \infty$?

Answer. (a) From Jensen's inequality, we have

$$|\mathbb{E}[\|X\|_2 - \sqrt{n}]| \leq \mathbb{E}[|\|X\|_2 - \sqrt{n}|] \leq \|\|X\|_2 - \sqrt{n}\|_{\psi_2} \leq CK^2$$

from Theorem 3.1.1 and

$$\|Z\|_{\psi_2} = \inf\{t > 0 \colon \mathbb{E}[\exp(Z^2/t^2)] \leq 2\} \geq \|Z\|_{L^1}$$

as $\mathbb{E}[\exp(Z^2/(\mathbb{E}[|Z|]^2))] \geq 1 + \mathbb{E}[Z^2]/(\mathbb{E}[|Z|]^2) \geq 2$, again from Jensen's inequality.

(b) We first observe that $\mathbb{E}[\|X\|_2] \leq \sqrt{\mathbb{E}[\|X\|_2^2]} = \sqrt{n}$, hence we only need to deal with lower-bound. Consider the following non-negative function

$$f(x) = \sqrt{x} - \frac{1}{2}(1 + x - (x-1)^2) \geq 0$$

for $x \geq 0$. Then, for $x = \|X\|_2^2/n \geq 0$, we have

$$\begin{aligned}
\sqrt{\frac{\|X\|_2^2}{n}} &\geq \frac{1}{2}\left(1 + \frac{\|X\|_2^2}{n} - \left(\frac{\|X\|_2^2}{n} - 1\right)^2\right) \\
\Rightarrow \|X\|_2 &\geq \frac{\sqrt{n}}{2}\left(1 + \frac{\|X\|_2^2}{n} - \left(\frac{\|X\|_2^2}{n} - 1\right)^2\right) \\
\Rightarrow \mathbb{E}[\|X\|_2] &\geq \frac{\sqrt{n}}{2}\left(1 + \frac{n}{n}\right) - \frac{\sqrt{n}}{2}\mathbb{E}\left[\left(\frac{\|X\|_2^2 - \mathbb{E}[\|X\|_2^2]}{n}\right)^2\right] \\
\Rightarrow \mathbb{E}[\|X\|_2] &\geq \sqrt{n} - \frac{1}{2n^{3/2}}\operatorname{Var}[\|X\|_2^2].
\end{aligned}$$

Expanding the variance, we see that

$$\operatorname{Var}[\|X\|_2^2] = \sum_{i=1}^n \operatorname{Var}\left[X_i^2\right] = \sum_{i=1}^n \left(\mathbb{E}[X_i^4] - \mathbb{E}[X_i^2]^2\right) \leq n \cdot \max_{1 \leq i \leq n} \mathbb{E}[X_i^4] = n \cdot \max_{1 \leq i \leq n} \|X_i\|_{L^4}^4,$$

and from the sub-gaussian property, this is $\lesssim n \cdot \max_{1\leq i\leq n} \|X_i\|_{\psi_2}^4 = nK^4$. Overall,

$$\mathbb{E}[\|X\|_2] \gtrsim \sqrt{n} - \frac{1}{2n^{3/2}} nK^4 = \sqrt{n} - \frac{K^4}{\sqrt{n}} = \sqrt{n} + o(1),$$

if $K \geq 1$. Otherwise, when $K < 1$, we replace K^4 by 1, the result holds still.

⊛

Problem (Exercise 3.1.5). Deduce from Theorem 3.1.1 that

$$\operatorname{Var}[\|X\|_2] \leq CK^4.$$

Answer. From the definition and the fact that the mean minimizes the MSE,

$$\operatorname{Var}[\|X\|_2] = \mathbb{E}[(\|X\|_2 - \mathbb{E}[\|X\|_2])^2] \leq \mathbb{E}[(\|X\|_2 - \sqrt{n})^2],$$

then from the proof of Exercise 3.1.4, as $\mathbb{E}[|\|X\|_2 - \sqrt{n}|] \leq cK^2$ for some c,

$$\operatorname{Var}[\|X\|_2] \leq \mathbb{E}[(\|X\|_2 - \sqrt{n})^2] \leq c^2 K^4,$$

and by letting $c^2 =: C$, we're done.

⊛

Problem (Exercise 3.1.6). Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector with independent coordinates X_i that satisfy $\mathbb{E}[X_i^2] = 1$ and $\mathbb{E}[X_i^4] \leq K^4$. Show that

$$\operatorname{Var}[\|X\|_2] \leq CK^4.$$

Answer. Firstly, observe that with our new assumption, Exercise 3.1.4 (b) again gives $\mathbb{E}[\|X\|_2] \gtrsim \sqrt{n} - K^4/\sqrt{n}$. Then from the same reason as stated in Exercise 3.1.5,

$$\operatorname{Var}[\|X\|_2] \leq \mathbb{E}[(\|X\|_2 - \sqrt{n})^2] = 2n - 2\sqrt{n}\mathbb{E}[\|X\|_2] \lesssim 2n - 2\sqrt{n}\left(\sqrt{n} - \frac{K^4}{\sqrt{n}}\right) = 2K^4,$$

proving the result. ⊛

Problem (Exercise 3.1.7). Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector with independent coordinates X_i with continuous distributions. Assume that the densities of X_i are uniformly bounded by 1. Show that, for any $\epsilon > 0$, we have

$$\mathbb{P}(\|X\|_2 \leq \epsilon\sqrt{n}) \leq (C\epsilon)^n.$$

Answer. We want to bound

$$\mathbb{P}\left(\|X\|_2 \leq \epsilon\sqrt{n}\right) = \mathbb{P}(\|X\|_2^2 \leq \epsilon^2 n) = \mathbb{P}\left(\sum_{i=1}^n X_i^2 \leq \epsilon^2 n\right).$$

Follow the same argument as Exercise 2.2.10,[a] i.e., first we bound $\mathbb{E}[\exp(-tX_i^2)]$ for all $t > 0$. We have

$$\mathbb{E}[\exp(-tX_i^2)] = \int_0^\infty e^{-tx^2} f_{X_i}(x) \,\mathrm{d}x \leq \int_0^\infty e^{-tx^2} \,\mathrm{d}x = \frac{1}{2}\sqrt{\frac{\pi}{t}}$$

from the Gaussian integral. Then, from the MGF trick, we have

$$\mathbb{P}(\|X\|_2 \leq \epsilon\sqrt{n}) = \mathbb{P}(-\|X\|_2^2 \geq -\epsilon^2 n) \leq \inf_{t>0} \frac{\mathbb{E}[\exp(-t\|X\|_2^2)]}{\exp(-t\epsilon^2 n)} \leq \inf_{t>0} \left(\frac{1}{2}\sqrt{\frac{\pi}{t}}\right)^n e^{t\epsilon^2 n}.$$

Let $t = \epsilon^{-2}$, we have

$$\mathbb{P}(\|X\|_2 \le \epsilon\sqrt{n}) \le \left(\frac{\sqrt{\pi}}{2}\epsilon \cdot e\right)^n =: (C\epsilon)^n$$

by letting $C := \sqrt{\pi}e/2$. ⊛

[a] The result does not directly follow from this because ϵ is replaced by ϵ^2, and a bound on the density of X_i doesn't give a bound on the density of X_i^2.

3.2 Covariance matrices and principal component analysis

Problem (Exercise 3.2.2)**.** (a) Let Z be a mean zero, isotropic random vector in $\mathbb{R}^n$. Let $\mu \in \mathbb{R}^n$ be a fixed vector and Σ be a fixed $n \times n$ symmetric positive semidefinite matrix. Check that the random vector

$$X := \mu + \Sigma^{1/2} Z$$

has mean μ and covariance matrix $\operatorname{Cov}[X] = \Sigma$.

(b) Let X be a random vector with mean μ and invertible covariance matrix $\Sigma = \operatorname{Cov}[X]$. Check that the random vector

$$Z := \Sigma^{-1/2}(X - \mu)$$

is an isotropic, mean zero random vector.

Answer. (a) Firstly,

$$\mathbb{E}[X] = \mathbb{E}[\mu] + \mathbb{E}[\Sigma^{1/2} Z] = \mu + \Sigma^{1/2}\mathbb{E}[Z] = \mu$$

Moreover,

$$\begin{aligned}
\operatorname{Cov}[X] &= \operatorname{Cov}[\mu + \Sigma^{1/2} Z] \\
&= \mathbb{E}[(\mu + \Sigma^{1/2} Z)(\mu + \Sigma^{1/2} Z)^\top] - \mu\mu^\top \\
&= \mathbb{E}[(\mu + \Sigma^{1/2} Z) Z^\top (\Sigma^{1/2})^\top] \\
&= \mathbb{E}[\mu Z^\top (\Sigma^{1/2})^\top] + \mathbb{E}[\Sigma^{1/2} Z Z^\top (\Sigma^{1/2})^\top] \\
&= 0 + \Sigma^{1/2}\mathbb{E}[Z Z^\top](\Sigma^{1/2})^\top \\
&= \Sigma^{1/2} I_n (\Sigma^{1/2})^\top \\
&= \Sigma
\end{aligned}$$

as Σ is positive-semidefinite.

(b) Similarly,

$$\mathbb{E}[Z] = \Sigma^{-1/2}\mathbb{E}[X - \mu] = \Sigma^{-1/2}(\mu - \mu) = 0,$$

and moreover,

$$\begin{aligned}
\operatorname{Cov}[Z] &= \operatorname{Cov}[\Sigma^{-1/2}(X - \mu)] \\
&= \mathbb{E}\left[(\Sigma^{-1/2}(X - \mu))(\Sigma^{-1/2}(X - \mu))^\top\right] \\
&= \Sigma^{-1/2}\mathbb{E}[(X - \mu)(X - \mu)^\top](\Sigma^{-1/2})^\top \\
&= \Sigma^{-1/2}\Sigma(\Sigma^{-1/2})^\top \\
&= I_n,
\end{aligned}$$

hence Z is also isotropic.

⊛

Problem (Exercise 3.2.6)**.** Let X and Y be independent, mean zero, isotropic random vectors in $\mathbb{R}^n$.

Check that

$$\mathbb{E}[\|X - Y\|_2^2] = 2n.$$

Answer. This directly follows from

$$\mathbb{E}[\|X - Y\|_2^2] = \mathbb{E}[\langle X - Y, X - Y\rangle] = \mathbb{E}[\langle X, X\rangle] - 2\mathbb{E}[\langle X, Y\rangle] + \mathbb{E}[\langle Y, Y\rangle] = n - 0 + n = 2n.$$

⊛

Week 10: Common High-Dimensional Distributions

3.3 Examples of high-dimensional distributions

20 Mar. 2024

Problem (Exercise 3.3.1)**.** Show that the spherically distributed random vector X is isotropic. Argue that the coordinates of X are not independent.

Answer. Firstly, from the spherical symmetry of X, for any $x \in \mathbb{R}^n$, $\langle X, x\rangle \overset{D}{=} \langle X, \|x\|_2 e\rangle$ for all $e \in S^{n-1}$. Hence, to show X is isotropic, from Lemma 3.2.3, it suffices to show that for any $x \in \mathbb{R}^n$,

$$\mathbb{E}[\langle X, x\rangle^2] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\langle X, \|x\|_2 e_i\rangle^2] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} (\|x\|_2 X_i)^2\right] = \|x\|_2^2 \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] = \|x\|_2^2,$$

where e_i denotes the i^{th} standard unit vector. The last equality holds from the fact that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] = \frac{1}{n}\mathbb{E}[\|X\|_2^2] = \frac{1}{n} n = 1$$

as $X \sim \mathcal{U}(\sqrt{n} S^{n-1})$. On the other hand, clearly X_i's can't be independent since the first $n-1$ coordinates determines the last coordinate. ⊛

Problem (Exercise 3.3.3)**.** Deduce the following properties from the rotation invariance of the normal distribution.

(a) Consider a random vector $g \sim \mathcal{N}(0, I_n)$ and a fixed vector $u \in \mathbb{R}^n$. Then

$$\langle g, u\rangle \sim \mathcal{N}(0, \|u\|_2^2).$$

(b) Consider independent random variables $X_i \sim \mathcal{N}(0, \sigma_i^2)$. Then

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(0, \sigma^2) \text{ where } \sigma^2 = \sum_{i=1}^{n} \sigma_i^2.$$

(c) Let G be an $m \times n$ Gaussian random matrix, i.e., the entries of G are independent $\mathcal{N}(0, 1)$ random variables. Let $u \in \mathbb{R}^n$ be a fixed unit vector. Then

$$Gu \sim \mathcal{N}(0, I_m).$$

Answer. (a) Without loss of generality, we may assume $\|u\|_2 = 1$ and prove

$$\langle g, u\rangle \sim \mathcal{N}(0, 1)$$

for any fixed unit vector $u \in \mathbb{R}^n$. But this is clear as there must exist $u_1, \ldots, u_{n-1}$ such that $\{u, u_1, \ldots, u_{n-1}\}$ forms an orthonormal basis of $\mathbb{R}^n$, and $U := (u, u_1, \ldots, u_{n-1})^\top$ is orthonormal. From Proposition 3.3.2, we have

$$Ug \sim \mathcal{N}(0, I_n),$$

which implies $(Ug)_1 \sim \mathcal{N}(0,1)$. With $(Ug)_1 = u^\top g = \langle g, u\rangle$, we're done.

(b) For independent $X_i \sim \mathcal{N}(0, \sigma_i^2)$, we have $X_i/\sigma_i \sim \mathcal{N}(0,1)$. We want to show

$$\sum_{i=1}^n X_i \sim \mathcal{N}(0, \sigma^2)$$

where $\sigma^2 = \sum_{i=1}^n \sigma_i^2$. Firstly, we have $g := (X_1/\sigma_1, \dots, X_n/\sigma_n) \sim \mathcal{N}(0, I_n)$, then by considering $u := (\sigma_1, \dots, \sigma_n) \in \mathbb{R}^n$, we have

$$\langle g, u\rangle = \sum_{i=1}^n X_i \sim \mathcal{N}(0, \|u\|_2^2) = \mathcal{N}\left(0, \sum_{i=1}^n \sigma_i^2\right) = \mathcal{N}(0, \sigma^2)$$

from (a).

(c) For any fixed unit vector u, $(Gu)_i = \sum_{j=1}^n g_{ij}u_j = \langle g_i, u\rangle$ where $g_i = (g_{i1}, g_{i2}, \dots, g_{in})$ for all $i \in [m]$. It's clear that $g_i \sim \mathcal{N}(0, I_n)$, and from (a), $\langle g_i, u\rangle \sim \mathcal{N}(0,1)$. This implies

$$Gu = (\langle g_1, u\rangle, \dots, \langle g_m, u\rangle) \sim \mathcal{N}(0, I_m)$$

as desired.

⊛

Problem (Exercise 3.3.4). Let X be a random vector in $\mathbb{R}^n$. Show that X has a multivariate normal distribution if and only if every one-dimensional marginal $\langle X, \theta\rangle$, $\theta \in \mathbb{R}^n$, has a (univariate) normal distribution.

Answer. This is an application of Cramér-Wold device and Exercise 3.3.3 (a). Omit the details. ⊛

Problem (Exercise 3.3.5). Let $X \sim \mathcal{N}(0, I_n)$.

(a) Show that, for any fixed vectors $u, v \in \mathbb{R}^n$, we have

$$\mathbb{E}[\langle X, u\rangle\langle X, v\rangle] = \langle u, v\rangle.$$

(b) Given a vector $u \in \mathbb{R}^n$, consider the random variable $X_u := \langle X, u\rangle$. From Exercise 3.3.3 we know that $X_u \sim \mathcal{N}(0, \|u\|_2^2)$. Check that

$$\|X_u - X_v\|_{L^2} = \|u - v\|_2$$

for any fixed vectors $u, v \in \mathbb{R}^n$.

Answer. (a) It's because

$$\mathbb{E}[\langle X, u\rangle\langle X, v\rangle] = \mathbb{E}[(u^\top X)(X^\top v)] = u^\top \mathbb{E}[XX^\top]v = u^\top I_n v = \langle u, v\rangle$$

from the fact that X is isotropic.

(b) Since $X_u - X_v = \langle X, u\rangle - \langle X, v\rangle = \langle X, u - v\rangle = X_{u-v}$ from linearity of inner product. Hence,

$$\|X_u - X_v\|_{L^2} = \sqrt{\langle X_{u-v}, X_{u-v}\rangle} = \sqrt{\mathbb{E}[X_{u-v}^2]} = \sqrt{\mathbb{E}[\langle X, u-v\rangle^2]}.$$

From (a), $\mathbb{E}[\langle X, u-v\rangle^2] = \langle u-v, u-v\rangle = \|u-v\|_2^2$, hence

$$\|X_u - X_v\|_{L^2} = \sqrt{\|u-v\|_2^2} = \|u-v\|_2.$$

⊛

Problem (Exercise 3.3.6)**.** h Let G be an $m\times n$ Gaussian random matrix, i.e., the entries of G are independent $\mathcal{N}(0,1)$ random variables. Let $u, v \in \mathbb{R}^n$ be unit orthogonal vectors. Prove that Gu and Gv are independent $\mathcal{N}(0, I_m)$ random vectors.

Answer. It's clear that Gu and Gv are both $\mathcal{N}(0, I_m)$ random vectors from Exercise 3.3.3 (c). It remains to show that Gu and Gv are independent, i.e., $(Gu)_i$ and $(Gv)_j$ are independent random variables.

For $i \neq j$, this is clear since $(Gu)_i = e_i^\top (Gu)$ and $(Gv)_j = e_j^\top (Gv)$, and $e_i^\top G$ gives the i^{th} row of G, while $e_j^\top G$ gives the j^{th} row of G. The fact that G has independent rows proves the result for the case of $i \neq j$.

For $i = j$, let $e_i^\top G =: g^\top$ where $g \sim \mathcal{N}(0, I_n)$, and we want to show independence of $(Gu)_i = g^\top u$ and $(Gv)_j = g^\top v$. This is still easy since

$$\begin{pmatrix} g^\top u \\ g^\top v \end{pmatrix} = (u, v)^\top g \sim \mathcal{N}(0, (u,v)^\top I_n (u,v)) = \mathcal{N}(0, I_2)$$

as u, v are unit orthogonal vectors. ⊛

Problem (Exercise 3.3.7)**.** Let us represent $g \sim \mathcal{N}(0, I_n)$ in polar form as

$$g = r\theta$$

where $r = \|g\|_2$ is the length and $\theta = g/\|g\|_2$ is the direction of g. Prove the following:

(a) The length r and direction θ are independent random variables.

(b) The direction θ is uniformly distributed on the unit sphere S^{n-1}.

Answer. For any measurable $M \subseteq \mathbb{R}^n$, given the normal density $f_G(g)$ of g, some elementary calculus gives the polar coordinate transformation $\mathrm{d}g = r^{n-1}\,\mathrm{d}r\,\mathrm{d}\sigma(\theta)$, hence

$$\begin{aligned}\mathbb{P}(g\in M) = \int_M f_G(g)\,\mathrm{d}g &= \int_A \int_B f_G(r\theta)\,\mathrm{d}\sigma(\theta) r^{n-1}\,\mathrm{d}r \\ &= \frac{\omega_{n-1}}{(2\pi)^{n/2}} \int_A r^{n-1} e^{-r^2/2}\,\mathrm{d}r \int_B \mathrm{d}\sigma(\theta) = \mathbb{P}(r\in A, \theta\in B)\end{aligned} \tag{3.1}$$

for some $A \subseteq [0,\infty)$ and $B \subseteq S^{n-1}$ generating M, where σ is the surface area element on S^{n-1} such that $\int_{S^{n-1}} \mathrm{d}\sigma = \omega_{n-1}$, i.e., ω_{n-1} is the surface area of the unit sphere S^{n-1}.

(a) From Equation 3.1, it's possible to write

$$\mathbb{P}(g\in M) = \mathbb{P}(r\in A, \theta\in B) =: f(A)g(B)$$

such that $g(S^{n-1}) = 1$ with appropriate constant manipulation. Hence, with $B = S^{n-1}$,

$$\mathbb{P}(r\in A, \theta \in S^{n-1}) = \mathbb{P}(r\in A) = f(A),$$

implying $f([0,\infty)) = 1$ as well. This further shows that by considering $A = [0,\infty)$,

$$\mathbb{P}(r\in[0,\infty), \theta\in B) = \mathbb{P}(\theta\in B) = g(B).$$

Such a separation of probability proves the independence.

(b) From Equation 3.1, we see that for any $B \subseteq S^{n-1}$, the density is uniform among $\mathrm{d}\sigma(\theta)$, hence θ is uniformly distributed on S^{n-1}.

⊛

Problem (Exercise 3.3.9). Show that $\{u_i\}_{i=1}^N$ is a tight frame in $\mathbb{R}^n$ with bound A if and only if

$$\sum_{i=1}^{N} u_i u_i^\top = AI_n.$$

Answer. Recall that for two symmetric matrices $A, B \in \mathbb{R}^{n\times n}$, $A = B$ if and only if $x^\top Ax = x^\top Bx$ for all $x \in \mathbb{R}^n$. Hence,

$$\sum_{i=1}^{N} u_i u_i^\top = AI_n \Leftrightarrow x^\top \left(\sum_{i=1}^{N} u_i u_i^\top\right) x = x^\top (AI_n) x$$

for all $x \in \mathbb{R}^n$. We see that

- The left-hand side:

$$x^\top \left(\sum_{i=1}^{N} u_i u_i^\top\right) x = \sum_{i=1}^{N} (x^\top u_i)(u_i^\top x) = \sum_{i=1}^{N} \langle u_i, x\rangle^2,$$

- The right-hand side:

$$x^\top AI_n x = Ax^\top x = A\|x\|_2^2.$$

Hence, $\sum_{i=1}^N u_i u_i^\top = AI_n$ if and only if $\sum_{i=1}^N \langle u_i, x\rangle^2 = A\|x\|_2^2$, i.e., $\{u_i\}_{i=1}^N$ being a tight frame. ⊛

Week 11: High-Dimensional Sub-Gaussian Distributions

3.4 Sub-gaussian distributions in higher dimensions

29 Mar. 2024

Problem (Exercise 3.4.3). This exercise clarifies the role of independence of coordinates in Lemma 3.4.2.

1. Let $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ be a random vector with sub-gaussian coordinates X_i. Show that X is a sub-gaussian random vector.
2. Nevertheless, find an example of a random vector X with

$$\|X\|_{\psi_2} \gg \max_{i\leq n} \|X_i\|_{\psi_2}.$$

Answer.

1. We see that

$$\|X\|_{\psi_2} = \sup_{x\in S^{n-1}} \|\langle X, x\rangle\|_{\psi_2} \leq \sup_{x\in S^{n-1}} \sum_{i=1}^{n} \|x_i X_i\|_{\psi_2} \leq \sup_{x\in S^{n-1}} \|X_i\|_{\psi_2} < \infty.$$

2. Just consider $X_i = Z$ are the same where $Z \sim \mathcal{N}(0, 1)$. Then, we see that

$$\max_i \|X_i\|_{\psi_2} = \|Z\|_{\psi_2} = \sqrt{8/3}$$

as $\mathbb{E}[\exp(Z^2/t^2)] = 1/\sqrt{1-2/t^2}$. On the other hand,

$$\|X\|_{\psi_2} \geq \left\|\langle X, \mathbb{1}_n/\sqrt{n}\rangle\right\|_{\psi_2} = \|\sqrt{n}Z\|_{\psi_2} = \sqrt{8n/3}.$$

⊛

Problem (Exercise 3.4.4)**.** Show that

$$\|X\|_{\psi_2} \asymp \sqrt{\frac{n}{\log n}}.$$

Answer. Since we not only want an upper-bound, but a tight, non-asymptotic behavior, we need to calculate $\|X\|_{\psi_2}$ as precise as possible. We note that

$$\|X\|_{\psi_2} = \sup_{x\in S^{n-1}} \|\langle X, x\rangle\|_{\psi_2} = \sup_{x\in S^{n-1}} \inf\{t>0\colon \mathbb{E}[\exp(\langle X, x\rangle^2/t^2)] \leq 2\},$$

and clearly the supremum is attained when $x = e_i$ for some i. In this case,

$$\|X\|_{\psi_2} = \inf\{t>0\colon \mathbb{E}[\exp(X_i^2/t^2)] \leq 2\}.$$

Note that since $X \sim \mathcal{U}(\{\sqrt{n}e_i\}_i)$, we see if we focus on a particular coordinate i,

$$X_i = \begin{cases} 0, & \text{w.p. } \dfrac{n-1}{n}; \\ \sqrt{n}, & \text{w.p. } \dfrac{1}{n}. \end{cases}$$

Hence, for any $t > 0$,

$$\mathbb{E}[\exp(X_i^2/t^2)] = \frac{n-1}{n} + \frac{1}{n}\exp\left(\frac{n}{t^2}\right).$$

Equating the above to be exactly 2 and solve it w.r.t. t, we have

$$\frac{n-1+e^{n/t^2}}{n} = 2 \Leftrightarrow n-1+e^{n/t^2} = 2n \Leftrightarrow \ln(n+1) = \frac{n}{t^2} \Leftrightarrow t = \sqrt{\frac{n}{\ln(n+1)}},$$

meaning that

$$\|X\|_{\psi_2} = \inf\{t>0\colon \mathbb{E}[\exp(X_i^2/t^2)] \leq 2\} = \sqrt{\frac{n}{\ln(n+1)}} \asymp \sqrt{\frac{n}{\log n}}.$$

⊛

Problem (Exercise 3.4.5)**.** Let X be an isotropic random vector supported in a finite set $T \subseteq \mathbb{R}^n$. Show that in order for x to be sub-gaussian with $\|X\|_{\psi_2} = O(1)$, the cardinality of the set must be exponentially large in n:

$$|T| \geq e^{cn}.$$

Answer. This is a hard one. See here for details. ⊛

Problem (Exercise 3.4.7)**.** Extend Theorem 3.4.6 for the uniform distribution on the Euclidean ball $B(0,\sqrt{n})$ in $\mathbb{R}^n$ centered at the origin and with radius $\sqrt{n}$. Namely, show that a random vector

$$X \sim \mathcal{U}(B(0,\sqrt{n}))$$

is sub-gaussian, and

$$\|X\|_{\psi_2} \leq C.$$

Answer. For $X \sim \mathcal{U}(B(0, \sqrt{n}))$, consider $R := \|X\|_2/\sqrt{n}$ and $Y := X/R = \sqrt{n}X/\|X\|_2 \sim \mathcal{U}(\sqrt{n}S^{n-1})$. From Theorem 3.4.6, $\|Y\|_{\psi_2} \leq C$. It's clear that $R \leq 1$, hence for any $x \in S^{n-1}$,

$$\mathbb{E}[\exp(\langle X, x\rangle^2/t^2)] = \mathbb{E}[\exp(R^2\langle Y, x\rangle^2/t^2)] \leq \mathbb{E}[\exp(\langle Y, x\rangle^2/t^2)],$$

which implies $\|\langle X, x\rangle\|_{\psi_2} \leq \|\langle Y, x\rangle\|_{\psi_2}$. Hence, $\|X\|_{\psi_2} \leq \|Y\|_{\psi_2} \leq C$. ⊛

Problem (Exercise 3.4.9)**.** Consider a ball of the ℓ_1 norm in $\mathbb{R}^n$:

$$K := \{x \in \mathbb{R}^n : \|x\|_1 \leq r\}.$$

(a) Show that the uniform distribution on K is isotopic for some $r \asymp n$.

(b) Show that the subgaussian norm of this distribution is *not* bounded by an absolute constant as the dimension n grows.

Answer. (a) Observe that for $i \neq j$, $(X_i, X_j) \overset{D}{=} (X_i, -X_j)$, hence $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_iX_j] = 0$ for $i \neq j$. Hence, for X to be isotropic, we need $\mathbb{E}[X_i^2] = 1$. Now, we note that $\mathbb{P}(|X_i| > x) = (r-x)^n/r^n = (1-x/r)^n$ for $x \in [0, r]$, hence

$$\mathbb{E}[X_i^2] = \int_0^\infty 2x\mathbb{P}(|X_i| > x)\,\mathrm{d}x = 2r^2\int_0^r \frac{x}{r}\left(1 - \frac{x}{r}\right)^n \frac{\mathrm{d}x}{r} = 2r^2\int_0^1 t(1-t)^n\,\mathrm{d}t,$$

which with some calculation is $2r^2/(n^2 + 3n + 2)$. Equating this with 1 gives $r \asymp n$.

(b) It suffices to show that $\|X_i\|_{L^p} > C\sqrt{p}$, which in turns blow up the sub-Gaussian property in terms of L^p norm. We see that

$$\begin{aligned}\|X_i\|_{L^p}^p &= \int_0^\infty px^{p-1}\mathbb{P}(|X_i| > x)\,\mathrm{d}x\\ &= pr^p\int_0^r \left(\frac{x}{r}\right)^{p-1}\left(1 - \frac{x}{r}\right)^n \frac{\mathrm{d}x}{r} = pr^p\int_0^1 t^{p-1}(1-t)^n\,\mathrm{d}t = pr^p \cdot B(p, n+1),\end{aligned}$$

where B is the Beta function. From the Beta function,

$$\|X_i\|_{L^p}^p = pr^p \cdot \frac{\Gamma(p)\Gamma(n+1)}{\Gamma(p+n+1)},$$

hence $\|X_i\|_{L^p} > C\sqrt{p}$ is evident from the Stirling's formula.

⊛

Problem (Exercise 3.4.10)**.** Show that the concentration inequality in Theorem 3.1.1 may not hold for a general isotropic sub-gaussian random vector X. Thus, independence of the coordinates of X is an essential requirement in that result.

Answer. We want to show that $\|\|X\|_2 - \sqrt{n}\|_{\psi_2} \leq C\max\|X_i\|_{\psi_2}^2$ does not hold for a general isotropic sub-Gaussian random vector X with $\mathbb{E}[X_i^2] = 1$. Let $0 < a < 1 < b$ such that $a^2 + b^2 = 2$, and define

$$X := (aZ)^\epsilon(bZ)^{1-\epsilon},$$

where $\epsilon \sim \mathrm{Bern}(1/2)$ and $Z \sim \mathcal{N}(0, I_n)$. In human language, consider X has a distribution

$$F_X := \frac{1}{2}F_{aZ} + \frac{1}{2}F_{bZ}.$$

With this construction, X is isotropic since

$$\begin{aligned}\mathbb{E}[XX^\top] &= \frac{1}{2}\mathbb{E}[(aZ)(aZ)^\top] + \frac{1}{2}\mathbb{E}[(bZ)(bZ)^\top] \\ &= \frac{1}{2}a^2\mathbb{E}[ZZ^\top] + \frac{1}{2}b^2\mathbb{E}[ZZ^\top] = \left(\frac{a^2}{2} + \frac{b^2}{2}\right) I_n = I_n,\end{aligned}$$

and $\mathbb{E}[X_i^2] = 1$ with a similar calculation. Moreover, for any vector $x \in S^{n-1}$,

$$\mathbb{E}[\exp(\langle X, x\rangle^2/t^2)] = \frac{1}{2\sqrt{1-2a^2/t^2}} + \frac{1}{2\sqrt{1-2b^2/t^2}} < 2$$

when t is large enough (compared to a, b). This shows $\|\langle X, x\rangle\|_{\psi_2} \leq t$, and since a, b is taken to be constants, X is indeed a sub-Gaussian random vector.

Now, we show that the norm of X actually deviates away from $\sqrt{n}$ at a non-vanishing rate of n. In particular, conciser $t = (b-1)\sqrt{n}/2$, then

$$\begin{aligned}2\mathbb{E}[\exp\big(\|X\|_2 - \sqrt{n}\big)^2/t^2] &> \mathbb{E}[\exp\big((\|bZ\|_2 - \sqrt{n})^2/t^2\big)] \\ &> \mathbb{E}[\exp\big((\|bZ\|_2 - \sqrt{n})^2/t^2\big)\mathbb{1}_{\|Z\|_2^2 > n}] \\ &> \exp\big((b\sqrt{n} - \sqrt{n})^2/t^2\big)\mathbb{P}(\|Z\|_2^2 > n) && \text{since } b > 1 \\ &= e^4\mathbb{P}(\|Z\|_2^2 > n) \\ &\to e^4/2 > 4\end{aligned}$$

since $\mathbb{P}(\|Z\|_2^2 > n) = \mathbb{P}\left(\sum_{i=1}^n Z_i^2 > n\right)$, and with $\mathbb{E}[Z_i^2] = \operatorname{Var}[Z_i] = 1$, and $\operatorname{Var}[Z_i^2] = \mathbb{E}[Z_i^4] - \mathbb{E}[Z_i]^2 = 3 - 1 = 2 < \infty$,

$$\frac{\frac{1}{n}\sum_{i=1}^n Z_i^2 - 1}{\sqrt{2}/\sqrt{n}} = \frac{1}{\sqrt{2n}}\left(\sum_{i=1}^n Z_i^2 - n\right) \overset{D}{\to} \mathcal{N}(0,1)$$

by the central limit theorem, hence, the asymptotic distribution of $\sum_{i=1}^n Z_i^2 - n$ is symmetric around 0, meaning that $\mathbb{P}(\sum_{i=1}^n Z_i^2 > n) = \mathbb{P}(\sum_{i=1}^n Z_i^2 - n > 0) = 1/2$. This implies that for all large enough n,

$$\|\|X\|_2 - \sqrt{n}\|_{\psi_2} \geq t = (b-1)\frac{\sqrt{n}}{2} \to \infty.$$

⊛

Week 12: High-Dimensional Sub-Gaussian Distributions

3.5 Application: Grothendieck's inequality and semidefinite programming

3 Apr. 2024

Problem (Exercise 3.5.2).

1. Check that the assumption of Grothendieck's inequality can be equivalently stated as follows:

$$\left|\sum_{i,j} a_{ij}x_iy_i\right| \leq \max_i|x_i| \cdot \max_j|y_j|$$

for any real numbers x_i and y_j.

2. Show that the conclusion of Grothendieck's inequality can be equivalently stated as follows:

$$\left|\sum_{i,j} a_{ij}\langle u_i, v_j\rangle\right| \leq K\max_i\|u_i\| \cdot \max_j\|v_j\|$$

for any Hilbert space H and any vectors $u_i, v_j \in H$.

Answer. Omit. ⊛

Problem (Exercise 3.5.3). Deduce the following version of Grothendieck's inequality for symmetric $n \times n$ matrices $A = (a_{ij})$ with real entries. Suppose that A is either positive semidefinie or has zero diagonal. Assume that, for any numbers $x_i \in \{-1, 1\}$, we have

$$\left| \sum_{i,j} a_{ij} x_i x_j \right| \leq 1.$$

Then, for any Hilbert space H and any vectors $u_i, v_j \in H$ satisfying $\|u_i\| = \|v_j\| = 1$, we have

$$\left| \sum_{i,j} a_{ij} \langle u_i, v_j \rangle \right| \leq 2K,$$

where K is the absolute constant from Grothendieck's inequality.

Answer. Omit. ⊛

Problem (Exercise 3.5.5). Show that the optimization (3.21) is equivalent to the following semidefinite program:

$$\max \langle A, X \rangle \colon X \succeq 0, \quad X_{ii} = 1 \text{ for } i = 1, \ldots, n.$$

Answer. Omit. ⊛

Problem (Exercise 3.5.7). Let A be an $m \times n$ matrix. Consider the optimization problem

$$\max \sum_{i,j} A_{ij} \langle X_i, Y_j \rangle \colon \|X_i\|_2 = \|Y_j\|_2 = 1 \text{ for all } i, j$$

over $X_i, Y_j \in \mathbb{R}^k$ and $k \in \mathbb{N}$. Formulate this problem as a semidefinite program.

Answer. Omit. ⊛

3.6 Application: Maximum cut for graphs

Problem (Exercise 3.6.4). For any $\epsilon > 0$, given an $(0.5 - \epsilon)$-approximation algorithm for maximum cut, which is always *guaranteed* to give a suitable cut, but may have a random running time. Give a bound on the expected running time.

Answer. Omit. ⊛

Problem (Exercise 3.6.7). Prove Grothendieck's identity.

Answer. Omit. ⊛

3.7 Kernel trick, and tightening of Grothendieck's inequality

Problem (Exercise 3.7.4). Show that for any vectors $u, v \in \mathbb{R}^n$ and $k \in \mathbb{N}$, we have

$$\langle u^{\otimes k}, v^{\otimes k} \rangle = \langle u, v \rangle^k.$$

Answer. This is immediate from the definition, i.e.,

$$\langle u^{\otimes k}, v^{\otimes k}\rangle = \sum_{i_1,\dots,i_k} u_{i_1\dots i_k} v_{i_1\dots i_k} = \sum_{i_1,\dots,i_k} u_{i_1}\dots u_{i_k} v_{i_1}\dots v_{i_k} = \left(\sum_{i=1}^n u_i v_i\right)^k$$

by observation (and probably term-matching). ⊛

Problem (Exercise 3.7.5). (a) Show that there exist a Hilbert space H and a transformation $\Phi\colon \mathbb{R}^n \to H$ such that

$$\langle \Phi(u), \Phi(v)\rangle = 2\langle u, v\rangle^2 + 5\langle u, v\rangle^3 \text{ for all } u, v \in \mathbb{R}^n.$$

(b) More generally, consider a polynomial $f\colon \mathbb{R} \to \mathbb{R}$ with non-negative coefficients, and construct H and Φ such that

$$\langle \Phi(u), \Phi(v)\rangle = f(\langle u, v\rangle) \text{ for all } u, v \in \mathbb{R}^n.$$

(c) Show the same for any *real analytic function* $f\colon \mathbb{R} \to \mathbb{R}$ with non-negative coefficients, i.e., for any function that can be represented as a convergent series

$$f(x) = \sum_{k=0}^\infty a_k x^k, \quad x \in \mathbb{R} \tag{3.2}$$

and such that $a_k \geq 0$ for all k.

Answer. (a) Consider $H = \mathbb{R}^{n\times n} \oplus \mathbb{R}^{n\times n\times n}$. Then, consider $\Phi(x) \coloneqq (\sqrt{2}x^{\otimes 2}, \sqrt{5}x^{\otimes 3})$, and we have

$$\begin{aligned}\langle \Phi(u), \Phi(v)\rangle &= \langle(\sqrt{2}u^{\otimes 2}, \sqrt{5}u^{\otimes 3}), (\sqrt{2}v^{\otimes 2}, \sqrt{5}v^{\otimes 3})\rangle \\ &= 2\langle u^{\otimes 2}, v^{\otimes 2}\rangle + 5\langle u^{\otimes 3}, v^{\otimes 3}\rangle = 2\langle u, v\rangle^2 + 5\langle u, v\rangle^3,\end{aligned}$$

where the last equality follows from Exercise 3.7.4.

(b) Consider an m-order polynomial of $\langle u, v\rangle$, which we write $f(\langle u, v\rangle) =: \sum_{k=0}^m a_k \langle u, v\rangle^k$. Then, by noting that $a_k \geq 0$, we may define

$$H \coloneqq \bigoplus_{k=0}^m \mathbb{R}^{n^k}, \text{ and } \Phi(x) \coloneqq \bigoplus_{k=0}^m \sqrt{a_k} x^{\otimes k} = (\sqrt{a_0}, \sqrt{a_1}x, \sqrt{a_2}x^{\otimes 2}, \dots, \sqrt{a_m}x^{\otimes m}).$$

Then by a similar calculation as (a), we have $\langle \Phi(u), \Phi(v)\rangle = f(\langle u, v\rangle)$ for all $u, v \in \mathbb{R}^n$.

(c) In this case, we just let $m = \infty$ in (b), i.e., consider

$$H \coloneqq \bigoplus_{k=0}^\infty \mathbb{R}^{n^k}, \text{ and } \Phi(x) \coloneqq \bigoplus_{k=0}^\infty \sqrt{a_k} x^{\otimes k},$$

where the limit is allowed as f converges everywhere. Note that $a_k \geq 0$, hence $\sqrt{a_k}$ is also well-defined.

⊛

Problem (Exercise 3.7.6). Let $f\colon \mathbb{R} \to \mathbb{R}$ be any real analytic function (with possibly negative coefficients in Equation 3.2). Show that there exist a Hilbert space H and transformation $\Phi, \Psi\colon \mathbb{R}^n \to H$ such that

$$\langle \Phi(u), \Psi(v)\rangle = f(\langle u, v\rangle) \text{ for all } u, v \in \mathbb{R}^n.$$

Moreover, check that

$$\|\Phi(u)\|^2 = \|\Psi(u)\|^2 = \sum_{k=0}^{\infty} |a_k| \|u\|_2^{2k}.$$

Answer. Again, similar to Exercise 3.7.5 (c), we construct

$$H := \bigoplus_{k=0}^{\infty} \mathbb{R}^{n^k}, \text{ and } \Phi(x) := \bigoplus_{k=0}^{\infty} \sqrt{a_k} x^{\otimes k}, \text{ and } \Psi(x) := \bigoplus_{k=0}^{\infty} \operatorname{sgn}(a_k) \sqrt{|a_k|} x^{\otimes k}.$$

Then, $\langle \Phi(u), \Psi_v \rangle = f(\langle u, v \rangle)$ since the sign of a_k is now taking care by Ψ. The norm can be calculated as

$$\begin{aligned} \|\Phi(u)\|^2 = \langle \Phi(u), \Phi_u \rangle &= \sum_{k=0}^{\infty} \langle \sqrt{|a_k|} u^{\otimes k}, \sqrt{|a_k|} u^{\otimes k} \rangle \\ &= \sum_{k=0}^{\infty} |a_k| \langle u^{\otimes k}, u^{\otimes k} \rangle = \sum_{k=0}^{\infty} |a_k| \langle u, u \rangle^k = \sum_{k=0}^{\infty} |a_k| \|u\|_2^{2k}, \end{aligned}$$

where the last equality follows from Exercise 3.7.4. A similar calculation can be carried out for $\|\Psi(u)\|^2$. ⊛

Chapter 4

Random matrices

4.1 Preliminaries on matrices

Problem (Exercise 4.1.1)**.** Suppose A is an invertible matrix with singular value decomposition

$$A = \sum_{i=1}^{n} s_i u_i v_i^\top .$$

Check that

$$A^{-1} = \sum_{i=1}^{n} \frac{1}{s_i} v_i u_i^\top .$$

Answer. Let $A = U\Sigma V^*$, and it suffices to check that

$$A\left(\sum_{i=1}^{n} \frac{1}{s_i} v_i u_i^\top\right) = I_n .$$

Indeed, by plugging A, we have

$$\left(\sum_{i=1}^{n} s_i u_i v_i^\top\right)\left(\sum_{i=1}^{n} \frac{1}{s_i} v_i u_i^\top\right) = \sum_{i=1}^{n} \frac{s_i}{s_i} u_i v_i^\top v_i u_i^\top = \sum_{i=1}^{n} u_i u_i^\top = UU^\top = I_n,$$

where all the cross-terms vanish since $v_i^\top v_j = 0$ as V is orthonormal, and $\sum_{i=1}^{n} u_i u_i^\top = UU^\top = I_n$ since U is again orthonormal. ⊛

Problem (Exercise 4.1.2)**.** Prove the following bound on the singular values s_i of any matrix A:

$$s_i \leq \frac{1}{\sqrt{i}} \|A\|_F .$$

Answer. We have seen that $\|A\|_F = \|s\|_2 = \sqrt{\sum_k s_k^2}$, hence

$$\|A\|_F^2 = \sum_{k=1}^{r} s_i^2 \geq \sum_{k\leq i} s_k^2 \geq i s_i^2$$

since we arrange s_k's in the decreasing order. This proves the result. ⊛

Problem (Exercise 4.1.3)**.** Let A_k be the best rank k approximation of a matrix A. Express $\|A - A_k\|^2$ and $\|A - A_k\|_F^2$ in terms of the singular values s_i of A.

Answer. From Eckart-Young-Mirsky theorem, we have

$$A_k = \sum_{i=1}^{k} s_i u_i v_i^\top ,$$

hence

$$A - A_k = \sum_{i=k+1}^{n} s_i u_i v_i^\top .$$

This implies, the singular values of the matrix $A - A_k$ are just $s_{k+1}, \ldots, s_n$,[a] implying

$$\|A - A_k\|^2 = s_{k+1}^2,$$

and

$$\|A - A_k\|_F^2 = \sum_{i=k+1}^{n} s_i^2.$$

⊛

[a]This can be seen from the fact that the same U and V still work, but now $s_i = 0$ for all $1 \leq i \leq k$.

Problem (Exercise 4.1.4). Let A be an $m \times n$ matrix with $m \geq n$. Prove that the following statements are equivalent.

(a) $A^\top A = I_n$.

(b) $P := AA^\top$ is an *orthogonal projection*[a] in $\mathbb{R}^m$ onto a subspace of dimension n.

(c) A is an *isometry*, or isometric embedding of $\mathbb{R}^n$ into $\mathbb{R}^m$, which means that

$$\|Ax\|_2 = \|x\|_2 \text{ for all } x \in \mathbb{R}^n.$$

(d) All singular values of A equal 1; equivalently

$$s_n(A) = s_1(A) = 1.$$

[a]Recall that P is a projection if $P^2 = P$, and P is called orthogonal if the image and kernel of P are orthogonal subspaces.

Answer. It's easy to see that (a), (c), and (d) are all equivalent. Indeed, for (a) and (c), we want $\|Ax\|_2^2 = (Ax)^\top (Ax) = xA^\top Ax = x^\top x = \|x\|_2^2$, and the equivalency lies in the equality $xA^\top Ax = x^\top x$. If $\|Ax\|_2 = \|x\|_2$ holds for all x, since $A^\top A$ is a symmetric matrix, we know that this means $A^\top A = I_n$. On the other hand, if $A^\top A = I_n$, then we clearly have the equality. For (c) and (d), noting the Equation 4.5 suffices. Now, we focus on proving the equivalence between (a) and (b).

- (a)⇒(b): Suppose $A^\top A = I_n$. Then $P = AA^\top$ is a projection since $P^2 = AA^\top AA^\top = AI_n A^\top = AA^\top = P$. Moreover, observe that $P^\top = P$, hence P is also an orthogonal projection.[a]

  Finally, we need to show that $\operatorname{rank}(P) = \operatorname{rank}(AA^\top) = n$. But since $A^\top A = I_n$,

  $$n = \operatorname{rank}(I_n) = \operatorname{rank}(A^\top A) \leq \operatorname{rank}(A) \leq n$$

  as matrix multiplication can only reduce the rank, hence $\operatorname{rank}(A) = n$. This also implies $\operatorname{rank}(A^\top) = n$, hence we're left to check whether $\operatorname{Im} A^\top \cap \ker A = \varnothing$. If this is true, then $\operatorname{rank}(AA^\top) = n$ as well, and we're done. But it's well-known that $\operatorname{Im} A^\top = (\ker A)^\top$, which completes the proof.

- (b)⇒(a): We want to show that if $P = AA^\top$ is an orthogonal projection on a subspace of dimension n, then $A^\top A = I_n$. Observe that since $P^2 = P$,

$$(AA^\top)(AA^\top) = AA^\top \Leftrightarrow A(A^\top A - I_n)A^\top = 0.$$

Now, we use the fact that $\operatorname{rank}(P) = \operatorname{rank}(AA^\top) = n$. From the previous argument, we know that $\operatorname{rank}(A) = \operatorname{rank}(A^\top) = n$, and hence

$$A(A^\top A - I_n)A^\top = 0 \Rightarrow A(A^\top A - I_n) = 0$$

as $A^\top$ spans all $\mathbb{R}^n$. Taking the transpose, we again have

$$(A^\top A - I_n)^\top A^\top = 0 \Rightarrow (A^\top A - I_n)^\top = 0$$

since again, $A^\top$ spans all $\mathbb{R}^n$. We hence have $A^\top A = I_n$ as desired.

⊛

[a] Note that such a characterization is standard. See here for example.

Problem (Exercise 4.1.6). Prove the following converse to Lemma 4.1.5: if (4.7) holds, then

$$\|A^\top A - I_n\| \leq 3\max(\delta, \delta^2).$$

Answer. Firstly, by the quadratic maximizing characterization, we have

$$\begin{aligned}\|A^\top A - I_n\| &= \max_{x\in S^{n-1}, y\in S^{n-1}} \langle (A^\top A - I_n)x, y\rangle \\ &\leq \max_{x\in S^{n-1}} |x^\top (A^\top A - I_n)x| = \max_{x\in S^{n-1}} |\|Ax\|_2^2 - 1|.\end{aligned}$$

Since we assume that $\|Ax\|_2 \in [1-\delta, 1+\delta]$ (with $x \in S^{n-1}$ now),

$$\|A^\top A - I_n\| \leq \max|(1\pm\delta)^2 - 1| = \max|\delta^2 \pm 2\delta| \leq 3\max(\delta, \delta^2).$$

⊛

Problem (Exercise 4.1.8). Canonical examples of isometries and projections can be constructed from a fixed unitary matrix U. Check that any sub-matrix of U obtained by selecting a subset of columns is an isometry, and any sub-matrix obtained by selecting a subset of rows is a projection.

Answer. Consider a tall sub-matrix $A_{n\times k}$ of $U_{n\times n}$ for some $k < n$. We know that A is an isometry if and only if $A^\top$ is a projection. From Remark 4.1.7, it suffices to check $A^\top A = I_k$. But this is trivial since U is unitary, and we're basically computing pair-wise inner products between some columns (selected in A) of U.

On the other hand, consider a fat sub-matrix $B_{k\times n}$ of $U_{n\times n}$ for some $k < n$. We want to show that $B^\top B$ is an orthogonal projection (of dimension k). From Exercise 4.1.4, it's equivalent to showing $B^\top$ is an isometry, and from the above, it reduces to show that $U^\top$ is also unitary since $B^\top$ can be viewed as a tall sub-matrix of $U^\top$. But this is true by definition. ⊛

Week 13: Covering and Packing Numbers

4.2 Nets, covering numbers and packing numbers

12 Apr. 2024

Problem (Exercise 4.2.5). (a) Suppose T is a normed space. Prove that $\mathcal{P}(K, d, \epsilon)$ is the largest number of closed disjoint balls with centers in K and radii $\epsilon/2$.

(b) Show by example that the previous statement may be false for a general metric space T.

Answer. (a) Consider any ϵ-separated subset of K. Then, $\overline{B}(x_i, \epsilon/2)$'s are disjoint since if not, then there exists $y \in \overline{B}(x_i, \epsilon/2) \cap \overline{B}(x_j, \epsilon/2)$ such that

$$\epsilon < d(x_i, x_j) \leq d(x_i, y) + d(x_j, y) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

a contradiction. On the other hand, if $d(x_i, x_j) \leq \epsilon$ then

$$\frac{x_i + x_j}{2} \in \overline{B}(x_i, \epsilon/2) \cap \overline{B}(x_j, \epsilon/2),$$

hence, there is a one-to-one correspondence between ϵ-separated subset of K and families of closed disjoint balls with centers in K and radii $\epsilon/2$, proving the result.

(b) Let $T = \mathbb{Z}$ and $d(x, y) = \mathbb{1}_{x \neq y}$. For $K = \{0, 1\}$ and $\epsilon = 1$, we have $\mathcal{P}(K, d, 1) = 1$. On the other hand, $\overline{B}(0, 1/2) = \{0\}$ and $\overline{B}(1, 1/2) = \{1\}$ are disjoint. If the result of (a) holds, then at least $\mathcal{P}(K, d, 1) = 2$ as there are exactly two such disjoint closed balls.

⊛

Problem (Exercise 4.2.9). In our definition of the covering numbers of K, we required that the centers x_i of the balls $B(x_i, \epsilon)$ that form a covering lie in K. Relaxing this condition, define the *exterior covering number* $\mathcal{N}^{\text{ext}}(K, d, \epsilon)$ similarly but without requiring that $x_i \in K$. Prove that

$$\mathcal{N}^{\text{ext}}(K, d, \epsilon) \leq \mathcal{N}(K, d, \epsilon) \leq \mathcal{N}^{\text{ext}}(K, d, \epsilon/2).$$

Answer. The lower bound is trivial. We focus on the upper bound. Consider an exterior cover $\{\overline{B}(x_i, \epsilon/2)\}$ of K where x_i might not lie in K. Now, for every i, choose exactly one y_i from $\overline{B}(x_i, \epsilon/2) \cap K$ is it's non-empty. Then, $\{\overline{B}(y_i, \epsilon)\}$ covers K since

$$\overline{B}(x_i, \epsilon/2) \cap K \subseteq \overline{B}(y_i, \epsilon)$$

from $d(x, y_i) \leq d(x, x_i) + d(x_i, y_i) \leq \epsilon/2 + \epsilon/2 = \epsilon$ for any $x \in \overline{B}(x_i, \epsilon/2)$. Hence, by taking the union over i, $\{\overline{B}(y_i, \epsilon)\}$ indeed cover K, so the upper bound is proved. ⊛

Problem (Exercise 4.2.10). Give a counterexample to the following monotonicity property:

$$L \subseteq K \text{ implies } \mathcal{N}(L, d, \epsilon) \leq \mathcal{N}(K, d, \epsilon).$$

Prove an approximate version of monotonicity:

$$L \subseteq K \text{ implies } \mathcal{N}(L, d, \epsilon) \leq \mathcal{N}(K, d, \epsilon/2).$$

Answer. The problem lies in the fact that we're not allowing exterior covering. Consider $K = [-1, 1]$ and $L = \{-1, 1\}$. Then, $\mathcal{N}(L, d, 1) = 2 > 1 = \mathcal{N}(K, d, 1)$ for $d(x, y) = |x - y|$.

The approximate version of monotonicity can be proved with a similar argument as Exercise 4.2.9: specifically, consider an $\epsilon/2$-covering $\{x_i\}$ of K with size exactly $\mathcal{N}(K, d, \epsilon/2)$. Now, for every i, choose one $y_i \in \overline{B}(x_i, \epsilon/2) \cap L$ if the latter is non-empty. It turns out that $\{\overline{B}(y_i, \epsilon)\}$ covers L. Indeed, $\overline{B}(x_i, \epsilon/2) \cap L \subseteq \overline{B}(y_i, \epsilon)$ since

$$d(x, y_i) \leq d(x, x_i) + d(x_i, y_i) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all $x \in \overline{B}(x_i, \epsilon/2)$. ⊛

Intuition. The fundamental idea is just every such $\overline{B}(y_i, \epsilon)$ can cover $\overline{B}(x_i, \epsilon/2)$.

Problem (Exercise 4.2.15). Check that d_H is indeed a metric.

Answer. We check the following.

- $d_H(x,x)=0$ for all x and $d_H(x,y)>0$ for all $x\neq y$: Trivial.
- $d_H(x,y)=d_H(y,x)$ for all x,y: Trivial.
- $d_H(x,y)\leq d_H(x,z)+d_H(y,z)$ for all x,y,z: Suppose x and y initially disagrees at $d_H(x,y)$ places, and denote the set of those disagreeing indices as I. Then for any z, as long as z and x (hence y) disagrees at an index outside I, $d_H(x,z)+d_H(y,z)$ increases by 2. There's no way to exist a z such that $d_H(x,z)+d_H(y,z)$ can decrease, at best z and x (or y) disagrees at an index in I, then it'll coincide with y (or x), contributing the same amount to $d_H(x,y)$.

⊛

Problem (Exercise 4.2.16). Let $K=\{0,1\}^n$. Prove that for every integer $m\in[0,n]$, we have

$$\frac{2^n}{\sum_{k=0}^{m}\binom{n}{k}}\leq\mathcal{N}(K,d_H,m)\leq\mathcal{P}(K,d_H,m)\leq\frac{2^n}{\sum_{k=0}^{\lfloor m/2\rfloor}\binom{n}{k}}.$$

Answer. The middle inequality follows from Lemma 4.2.8. Now, for $K=\{0,1\}^n$, we first note that we have $|K|=2^n$. Furthermore, observe the following.

Claim. For any $x\in K$, we have

$$|\{y\in K\colon d_H(x,y)\leq m\}|=\sum_{k=0}^{m}|\{y\in K\colon d_H(x,y)=k\}|=\sum_{k=0}^{m}\binom{n}{k}.$$

We then see the following.

- Lower bound: observe that $|K|\leq\mathcal{N}(K,d_H,m)|\{y\in K\colon d_H(x_i,y)\leq m\}|$ where $\{x_i\}$ is an m-net of K of size $\mathcal{N}(K,d_H,m)$.
- Upper bound: observe that $|K|\geq\mathcal{P}(K,d_H,m)|\{y\in K\colon d_H(x_i,y)\leq\lfloor m/2\rfloor\}|$ where $\{x_i\}$ is m-packing of size $\mathcal{P}(K,d_H,m)$.

Plugging the above calculation complete the proof of both bounds. ⊛

Remark. Unlike Proposition 4.2.12, we don't have the issue of "going outside K" since we're working with a hamming cube, i.e., the entire universe is exactly the collection of n-bits string. Moreover, for the upper bound, we use $\lfloor m/2\rfloor$ since $m\in\mathbb{N}$, and taking the floor makes sure that $\{y\in K\colon d_H(x,y)\leq\lfloor m/2\rfloor\}$'s are disjoint for $\{x_i\}$ being m-separated. Hence, the total cardinality is upper bounded by $|K|$.

Week 14: Random Sub-Gaussian Matrices

4.3 Application: error correcting codes

17 Apr. 2024

Problem (Exercise 4.3.7). (a) Prove the converse to the statement of Lemma 4.3.4.

(b) Deduce a converse to Theorem 4.3.5. Conclude that for any error correcting code that encodes k-bit strings into n-bit strings and can correct r errors, the rate must be

$$R\leq 1-f(\delta)$$

where $f(t)=t\log_2(1/t)$ as before.

Answer. Omit. ⊛

4.4 Upper bounds on random sub-gaussian matrices

Problem (Exercise 4.4.2). Let $x \in \mathbb{R}^n$ and $\mathcal{N}$ be an ϵ-net of the sphere S^{n-1}. Show that

$$\sup_{y \in \mathcal{N}} \langle x, y \rangle \leq \|x\|_2 \leq \frac{1}{1-\epsilon} \sup_{y \in \mathcal{N}} \langle x, y \rangle.$$

Answer. The lower bound is again trivial. On the other hand, for any $x \in \mathbb{R}^n$, consider an $x_0 \in \mathcal{N}$ such that $\|x_0 - x/\|x\|_2\|_2 \leq \epsilon$ (normalization is necessary since $\mathcal{N}$ is an ϵ-net of S^{n-1}, while $x \in \mathbb{R}^n$). Now, observe that from the Cauchy-Schwarz inequality, we have

$$\|x\|_2 - \langle x, x_0 \rangle = \left\langle x, \frac{x}{\|x\|_2} - x_0 \right\rangle \leq \|x\|_2 \left\| \frac{x}{\|x\|_2} - x_0 \right\| \leq \epsilon \|x\|_2,$$

which implies $\langle x, x_0 \rangle \geq (1-\epsilon)\|x\|_2$. This proves the upper bound. ⊛

Problem (Exercise 4.4.3). Let A be an $m \times n$ matrix and $\epsilon \in [0, 1/2)$.

(a) Show that for any ϵ-net $\mathcal{N}$ of the sphere S^{n-1} and any ϵ-net $\mathcal{M}$ of the sphere S^{m-1} we have

$$\sup_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y \rangle \leq \|A\| \leq \frac{1}{1-2\epsilon} \cdot \sup_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y \rangle.$$

(b) Moreover, if $m = n$ and A is symmetric, show that

$$\sup_{x \in \mathcal{N}} |\langle Ax, x \rangle| \leq \|A\| \leq \frac{1}{1-2\epsilon} \cdot \sup_{x \in \mathcal{N}} |\langle Ax, x \rangle|.$$

Answer. (a) The lower bound is again trivial. On the other hand, denote $x^* \in S^{n-1}$ and $y^* \in S^{m-1}$ such that $\|A\| = \langle Ax^*, y^* \rangle$. Pick $x_0 \in \mathcal{N}$ and $y_0 \in \mathcal{M}$ such that $\|x^* - x_0\|_2, \|y^* - y_0\|_2 \leq \epsilon$. We then have

$$\begin{aligned} \langle Ax^*, y^* \rangle - \langle Ax_0, y_0 \rangle &= \langle A(x^* - x_0), y^* \rangle + \langle Ax_0, y^* - y_0 \rangle \\ &\leq \|A\|(\|x^* - x_0\|_2 \|y^*\|_2 + \|x_0\|_2 \|y^* - y_0\|_2) \leq 2\epsilon \|A\| \end{aligned}$$

as $\|y^*\| = \|x_0\|_2 = 1$. Rewrite the above, we have

$$\|A\| - \langle Ax_0, y_0 \rangle \leq 2\epsilon \|A\| \Rightarrow \|A\| \leq \frac{1}{1-2\epsilon} \langle Ax_0, y_0 \rangle \leq \frac{1}{1-2\epsilon} \sup_{x \in \mathcal{N}, y \in \mathcal{N}} \langle Ax, y \rangle.$$

(b) Following the same argument as (a), with $y^* := x^*$ and $y_0 := x_0$. To be explicit to handle the absolute value, we see that

$$|\langle Ax^*, x^* \rangle| - |\langle Ax_0, x_0 \rangle| \leq |\langle Ax^*, x^* \rangle - \langle Ax_0, x_0 \rangle| \leq 2\epsilon \|A\|,$$

from the same argument. The result follows immediately.

⊛

Problem (Exercise 4.4.4). Let A be an $m \times n$ matrix, $\mu \in \mathbb{R}$ and $\epsilon \in [0, 1/2)$. Show that for any ϵ-net $\mathcal{N}$ of the sphere S^{n-1}, we have

$$\sup_{x \in S^{n-1}} |\|Ax\|_2 - \mu| \leq \frac{C}{1-2\epsilon} \cdot \sup_{x \in \mathcal{N}} |\|Ax\|_2 - \mu|.$$

Answer. Let $\mu = 1$. Firstly, for $x \in S^{n-1}$, observe that we can write

$$\left| \|Ax\|_2^2 - 1 \right| = \langle Rx, x \rangle$$

for a symmetric $R = A^\top A - I_n$. Secondly, there exists x^* such that $\|R\| = \langle Rx^*, x^* \rangle$, consider $x_0 \in \mathcal{N}$ such that $\|x_0 - x^*\| \leq \epsilon$. Now, from a numerical inequality $|z-1| \leq |z^2 - 1|$ for $z > 0$, we have

$$\begin{aligned} \sup_{x \in S^{n-1}} |\|Ax\|_2 - 1| &\leq \sup_{x \in S^{n-1}} \left| \|Ax\|_2^2 - 1 \right| = \|R\| \\ &\leq \frac{1}{1-2\epsilon} \sup_{x \in \mathcal{N}} |\langle Rx, x \rangle| = \frac{1}{1-2\epsilon} \sup_{x \in \mathcal{N}} \left| \|Ax\|_2^2 - 1 \right|, \end{aligned}$$

where the last inequality follows from Exercise 4.4.3. Further, factoring $|\|Ax\|_2^2 - 1|$ get

$$\sup_{x \in S^{n-1}} |\|Ax\|_2 - 1| \leq \frac{1}{1-2\epsilon} \sup_{x \in \mathcal{N}} |\|Ax\|_2 - 1| \left(\|Ax\|_2 + 1 \right).$$

If $\|A\| \leq 2$, then $\|Ax\|_2 + 1 \leq 3$, and $C = 3$ suffices.

On the other hand, if $\|A\| > 2$, consider directly computing the left-hand side

$$\sup_{x \in S^{n-1}} |\|Ax\|_2 - 1| = \|A\| - 1$$

where the maximum is attained at some $x' \in S^{n-1}$. With the existence of $x'' \in \mathcal{N} \cap \{x \colon \|x - x'\|_2 \leq \epsilon\}$, the supremum over $\mathcal{N}$ can be lower bounded as

$$\sup_{x \in \mathcal{N}} |\|Ax\|_2 - 1| \geq \|Ax''\|_2 - 1 \geq \|Ax'\|_2 - \|A(x'' - x')\|_2 - 1 \geq \|A\|(1-\epsilon) - 1 > 1 - 2\epsilon.$$

The above implies the following.

- $\|A\| \leq \frac{1}{1-\epsilon}(\sup_{x \in \mathcal{N}} |\|Ax\|_2 - 1| + 1)$.
- $\sup_{x \in \mathcal{N}} |\|Ax\|_2 - 1| > 1 - 2\epsilon$.

This allows us to conclude that

$$\begin{aligned} \sup_{x \in S^{n-1}} |\|Ax\|_2 - 1| = \|A\| - 1 &\leq \frac{1}{1-\epsilon} \left(\sup_{x \in \mathcal{N}} |\|Ax\|_2 - 1| + 1 \right) - 1 \\ &= \frac{1}{1-\epsilon} \left(\sup_{x \in \mathcal{N}} |\|Ax\|_2 - 1| + \epsilon \right) \leq \frac{3}{1-2\epsilon} \sup_{x \in \mathcal{N}} |\|Ax\|_2 - 1|, \end{aligned}$$

provided that

$$C \coloneqq 3 \geq \sup_{d > 1-2\epsilon} \frac{1-2\epsilon}{1-\epsilon} \left(1 + \frac{\epsilon}{d} \right) \geq \frac{1-2\epsilon}{1-\epsilon} \frac{\sup_{x \in \mathcal{N}} |\|Ax\|_2 - 1| + \epsilon}{\sup_{x \in \mathcal{N}} |\|Ax\|_2 - 1|},$$

which is true since the middle supremum is just 1. The case that $\mu \neq 1$ can be easily generalized by considering $R = A^\top A - \mu I_n$. ⊛

Problem (Exercise 4.4.6)**.** Deduce from Theorem 4.4.5 that

$$\mathbb{E}[\|A\|] \leq CK(\sqrt{m} + \sqrt{n}).$$

Answer. From Theorem 4.4.5, for any $t > 0$, we have

$$\mathbb{P}(\|A\| - CK(\sqrt{m} + \sqrt{n}) > CKt) \leq 2\exp(-t^2).$$

Then we immediately have

$$\begin{aligned}\mathbb{E}[\|A\| - CK(\sqrt{m}+\sqrt{n})] &= \mathbb{E}[\|A\|] - CK(\sqrt{m}+\sqrt{n}) \\ &= \int_0^\infty \mathbb{P}(\|A\| - CK(\sqrt{m}+\sqrt{n}) > CKt) CK \,\mathrm{d}t \\ &\leq 2CK \int_0^\infty \exp(-t^2)\,\mathrm{d}t = CK\sqrt{\pi},\end{aligned}$$

hence $\mathbb{E}[\|A\|] \leq CK(\sqrt{m}+\sqrt{n}+\sqrt{\pi})$, and choosing a large enough C subsumes $\sqrt{\pi}$. ⊛

Problem (Exercise 4.4.7)**.** Suppose that in Theorem 4.4.5 the entries A_{ij} have unit variances. Prove that for sufficiently large n and m one has

$$\mathbb{E}[\|A\|] \geq \frac{1}{4}(\sqrt{m}+\sqrt{n}).$$

Answer. Clearly, by choosing $x = e_1 \in S^{n-1}$,

$$\|A\| = \sup_{x\in S^{n-1}} \|Ax\|_2 \geq \|(A_{i1})_{1\leq i\leq m}\|_2.$$

On the other hand, by picking $x = (A_{11}/\|(A_{1j})_{1\leq j\leq n}\|_2, \dots, A_{1n}/\|(A_{1j})_{1\leq j\leq n}\|_2) \in S^{n-1}$ and $y = e_1 \in S^{m-1}$, we have

$$\|A\| = \sup_{x\in S^{n-1}, y\in S^{m-1}} \langle Ax, y\rangle \geq \sum_{j=1}^n \frac{A_{1j}}{\|(A_{1j})_{1\leq j\leq n}\|_2} A_{1j} = \|(A_{1j})_{1\leq j\leq n}\|_2.$$

Hence, $\|A\|$ is lower bounded by the norm of the first row and column, i.e.,

$$\|A\| \geq \max(\|(A_{i1})_{1\leq i\leq m}\|_2, \|(A_{1j})_{1\leq j\leq n}\|_2).$$

Exercise 3.1.4 (b), the expectation of $\|A\|$ is then greater than or equal to $\max(\sqrt{m}-o(1), \sqrt{n}-o(1))$ by Thus, $\mathbb{E}[\|A\|] \geq (\sqrt{m}+\sqrt{n}-o(1))/2$. ⊛

Remark. An easier way to deduce the second (i.e., lower bounded by the norm of the first row) is to note that $\|A^\top\| = \|A\|$ by some elementary (functional) analysis.

Week 15: Stochastic Block Model and Community Detection

4.5 Application: community detection in networks

8 Jun. 2024

Problem (Exercise 4.5.2)**.** Check that the matrix D has rank 2, and the non-zero eigenvalues λ_i and the corresponding eigenvectors u_i are

$$\lambda_1 = \left(\frac{p+q}{2}\right)n,\quad u_1 = \begin{bmatrix}\mathbb{1}_{n/2\times 1}\\ \mathbb{1}_{n/2\times 1}\end{bmatrix},\quad \lambda_2 = \left(\frac{p-q}{2}\right)n,\quad u_2 = \begin{bmatrix}\mathbb{1}_{n/2\times 1}\\ -\mathbb{1}_{n/2\times 1}\end{bmatrix}.$$

Answer. Let n be an even number. Firstly, for any $D \in \mathbb{R}^{n\times n}$, columns 1 to $n/2$ are identical, same for columns $n/2+1$ to n. Furthermore, since $p > q$, column 1 and $n/2+1$ are linear independent, so $\operatorname{rank}(D) = 2$.

Instead of solving the characteristic equation and find the eigenvalues, and find the corresponding eigenvectors later, since we know that $\operatorname{rank}(D) = 2$, it's immediate that there are only 2 non-zero eigenvalues. Hence, we directly verify that

$$\lambda_1 = \left(\frac{p+q}{2}\right)n, \quad u_1 = \mathbb{1}_{n\times 1}, \quad \lambda_2 = \left(\frac{p-q}{2}\right)n, \quad u_2 = \begin{pmatrix} \mathbb{1}_{1\times n/2} \\ -\mathbb{1}_{1\times n/2} \end{pmatrix}^\top.$$

For λ_1, indeed, since

$$Du_1 = \lambda_1 u_1 \Rightarrow \begin{pmatrix} \frac{p+q}{2}n \\ \frac{p+q}{2}n \\ \vdots \\ \frac{q+p}{2}n \\ \frac{q+p}{2}n \end{pmatrix} = \left(\frac{p+q}{2}\right)n \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}.$$

On the other hand, for λ_2, we have

$$Du_2 = \lambda_2 u_2 \Rightarrow \begin{pmatrix} \frac{p-q}{2}n \\ \frac{p-q}{2}n \\ \vdots \\ \frac{q-p}{2}n \\ \frac{q-p}{2}n \end{pmatrix} = \left(\frac{p-q}{2}\right)n \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ -1 \\ -1 \end{pmatrix},$$

which again holds. ⊛

Problem (Exercise 4.5.4)**.** Deduce Weyl's inequality from the Courant-Fisher's min-max characterization of eigenvalues.

Answer. We have that from the Courant-Fisher's min-max characterization,

$$\lambda_i(A) = \max_{\dim E = i} \min_{x\in S(E)} \langle Ax, x\rangle.$$

Now, as $\lambda_i(A) = -\lambda_{n-i+1}(-A)$, we see that

$$\lambda_i(A) = -\lambda_{n-i+1}(-A) = -\max_{\dim E = n-i+1} \min_{x\in S(E)} \langle -Ax, x\rangle = \min_{\dim E = n-i+1} \max_{x\in S(E)} \langle Ax, x\rangle.$$

We now show the Weyl's inequality.

Theorem 4.5.1 (Weyl's inequality)**.** $\lambda_{i+j-1}(A+B) \le \lambda_i(A) + \lambda_j(B) \le \lambda_{i+j-n}(A+B)$.

Proof. We first show the lower-bound. From the Courant-Fisher's min-max characterization, it suffices to show that for any E with $\dim E = i+j-1$, there exists some $x\in S(E)$ such that $\langle (A+B)x, x\rangle \le \lambda_i(A) + \lambda_j(B)$.

We first analyze $\lambda_i(A)$. We know that from the max-min characterization,

$$\lambda_i(A) = \min_{\dim E = n-i+1} \max_{x\in S(E)} \langle Ax, x\rangle,$$

i.e., there exists some E_A with $\dim E_A = n-i+1$ such that $\lambda_i(A) = \max_{x\in S(E_A)}\langle Ax, x\rangle$. Similarly, there exists some E_B with $\dim E_B = n-j+1$ satisfying the same property. Hence, it suffices to find some unit vector x in $E_A\cap E_B\cap E$. We see that

$$\dim(E_A\cap E_B) \ge \dim E_A + \dim E_B - n = n-i-j+2,$$

which implies that $E_A\cap E_B$ will have a non-trivial intersection with E since $\dim E = i+j-1$, hence we're done. For the upper-bound, taking the negative gives the result. ∎

To obtain the spectral stability, we see that from Weyl's inequality, we have

$$\begin{cases} \lambda_i(A+B) \leq \lambda_i(A) + \lambda_1(B); \\ \lambda_i(A+B) \geq \lambda_i(A) + \lambda_n(B); \end{cases} \Rightarrow \lambda_n(B) \leq \lambda_i(A+B) - \lambda_i(A) \leq \lambda_1(B).$$

Given any symmetric S, T, by setting $A \coloneqq T$ and $B \coloneqq S - T$, the upper-bound yields

$$\lambda_i(S) - \lambda_i(T) \leq \lambda_1(S-T) = \|S-T\|.$$

On the other hand, by setting $A \coloneqq S$ and $B \coloneqq T - S$, the upper-bound again yields

$$\lambda_i(T) - \lambda_i(S) \leq \lambda_1(T-S) = \|T-S\| = \|S-T\|.$$

As this holds for every i, we have

$$\max_i |\lambda_i(S) - \lambda_i(T)| \leq \|S-T\|$$

as we desired. ⊛

Week 16: Tighter Bounds on Sub-Gaussian Matrices

4.6 Two-sided bounds on sub-gaussian matrices

13 Jun. 2024

Problem (Exercise 4.6.2). Deduce from (4.22) that

$$\mathbb{E}\left[\left\|\frac{1}{m}A^\top A - I_n\right\|\right] \leq CK^2\left(\sqrt{\frac{n}{m}} + \frac{n}{m}\right).$$

Answer. We have that for any $t \geq 0$, with probability at least $1 - 2\exp(-t^2)$,

$$\left\|\frac{1}{m}A^\top A - I_n\right\| \leq K^2 \max(\delta, \delta^2), \text{ where } \delta = C\left(\sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}}\right),$$

and we want to prove

$$\mathbb{E}\left[\left\|\frac{1}{m}A^\top A - I_n\right\|\right] \leq CK^2\left(\sqrt{\frac{n}{m}} + \frac{n}{m}\right).$$

Firstly, we know that with $u \coloneqq K^2((\frac{C}{\sqrt{m}} + \frac{2C^2\sqrt{n}}{m})t + \frac{C^2}{m}t^2)$, we get exactly

$$\mathbb{P}\left(\left\|\frac{1}{m}A^\top A - I_n\right\| > K^2\left(C\sqrt{\frac{n}{m}} + C^2\frac{n}{m}\right) + u\right) \leq 2e^{-t^2}.$$

Then, from the integral identity with the substitution $v \coloneqq u + K^2(C\sqrt{\frac{n}{m}} + C^2\frac{n}{m})$,

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{m}A^\top A - I_n\right\|\right] \\
&= \left(\int_0^{K^2(C\sqrt{\frac{n}{m}}+C^2\frac{n}{m})} + \int_{K^2(C\sqrt{\frac{n}{m}}+C^2\frac{n}{m})}^\infty\right)\mathbb{P}\left(\left\|\frac{1}{m}A^\top A - I_n\right\| > v\right)\,\mathrm{d}v \\
&\leq \int_0^{K^2(C\sqrt{\frac{n}{m}}+C^2\frac{n}{m})} 1\,\mathrm{d}v + \int_{K^2(C\sqrt{\frac{n}{m}}+C^2\frac{n}{m})}^\infty \mathbb{P}\left(\left\|\frac{1}{m}A^\top A - I_n\right\| > v\right)\,\mathrm{d}v \\
&= K^2\left(C\sqrt{\frac{n}{m}} + C^2\frac{n}{m}\right) + \int_0^\infty \mathbb{P}\left(\left\|\frac{1}{m}A^\top A - I_n\right\| > K^2\left(C\sqrt{\frac{n}{m}} + C^2\frac{n}{m}\right) + u\right)\,\mathrm{d}u
\end{aligned}$$

plugging back $v = u + K^2(C\sqrt{\frac{n}{m}} + C^2\frac{n}{m})$,

$$
\begin{aligned}
&\leq K^2\left(C\sqrt{\frac{n}{m}} + C^2\frac{n}{m}\right) + \int_0^\infty 2e^{-t^2}\,\mathrm{d}u \\
&= K^2\left(C\sqrt{\frac{n}{m}} + C^2\frac{n}{m}\right) + \int_0^\infty 2e^{-t^2}K^2\left(\frac{C}{\sqrt{m}} + \frac{2C^2\sqrt{n}}{m} + \frac{2C^2}{m}t\right)\mathrm{d}t \\
&= K^2\left(C\sqrt{\frac{n}{m}} + C^2\frac{n}{m}\right) + K^2\left(\sqrt{\pi}\left(\frac{C}{\sqrt{m}} + \frac{2C^2\sqrt{n}}{m}\right) + \frac{2C^2}{m}\right),
\end{aligned}
$$

which is asymptotically $\asymp K^2(\sqrt{\frac{n}{m}} + \frac{n}{m})$. ⊛

Problem (Exercise 4.6.3)**.** Deduce from Theorem 4.6.1 the following bounds on the expectation:

$$\sqrt{m} - CK^2\sqrt{n} \leq \mathbb{E}[s_n(A)] \leq \mathbb{E}[s_1(A)] \leq \sqrt{m} + CK^2\sqrt{n}.$$

Answer. From Theorem 4.6.1, for any $t \geq 0$,

$$\sqrt{m} - CK^2(\sqrt{n} + t) \leq s_n(A) \leq s_1(A) \leq \sqrt{m} + CK^2(\sqrt{n} + t)$$

with probability at least $1 - 2\exp(-t^2)$. We want to show that

$$\sqrt{m} - CK^2\sqrt{n} \leq \mathbb{E}[s_n(A)] \leq \mathbb{E}[s_1(A)] \leq \sqrt{m} + CK^2\sqrt{n}.$$

Consider

$$\xi := \frac{\max\left(0, \sqrt{m} - CK^2\sqrt{n} - s_n(A), s_1(A) - \sqrt{m} - CK^2\sqrt{n}\right)}{CK^2} \geq 0,$$

then from the integral identity,

$$\mathbb{E}[\xi] = \int_0^\infty \mathbb{P}(\xi > t)\,\mathrm{d}t \leq \int_0^\infty 2e^{-t^2}\,\mathrm{d}t = \sqrt{\pi},$$

which proves the result. ⊛

Problem (Exercise 4.6.4)**.** Give a simpler proof of Theorem 4.6.1, using Theorem 3.1.1 to obtain a concentration bound for $\|Ax\|_2$ and Exercise 4.4.4 to reduce to a union bound over a net.

Answer. From the proof of Theorem 4.6.1, we know that S^{n-1} admits a $1/4$-net $\mathcal{N}$ such that $|\mathcal{N}| \leq 9^n$. Furthermore, for any $x \in \mathcal{N}$, we have

- $\mathbb{E}[\langle A_i, x\rangle] = \langle \mathbb{E}[A_i], x\rangle = \langle 0, x\rangle = 0$;
- $\mathbb{E}[\langle A_i, x\rangle^2] = x^\top \mathbb{E}[A_i^\top A_i]x = x^\top I_n x = 1$ ($x \in S^{n-1}$ too);
- $\|\langle A_i, x\rangle\|_{\psi_2} \leq \|A_i\|_{\psi_2} \leq K$ for all i,

by Theorem 3.1.1, we have $\|\|Ax\|_2 - \sqrt{m}\|_{\psi_2} \leq CK^2$. From Proposition 2.5.2 (i), for any $t > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(|\|Ax\|_2 - \sqrt{m}| > CK(\sqrt{n\log 9} + t)\right) \\
&\leq 2\exp\left(-(\sqrt{n\log 9} + t)^2\right) \leq 2\exp\left(-(n\log 9 + t^2)\right) = 2\cdot 9^{-n}\cdot e^{-t^2}.
\end{aligned}
$$

Finally, from Exercise 4.4.4, with a union bound over $\mathcal{N}$, we have

$$\mathbb{P}\left(\neg\left\{\sqrt{m} - 2CK^2(\sqrt{n\log 9} + t) \leq s_n(A) \leq s_1(A) \leq \sqrt{m} + 2CK^2(\sqrt{n\log 9} + t)\right\}\right)$$

by the definition of $s_n(A)$ and $s_1(A)$, we have

$$
\begin{aligned}
&\leq \mathbb{P}\left(\max_{x\in S^{n-1}} \left|\|Ax\|_2 - \sqrt{m}\right| > 2CK^2(\sqrt{n\log 9} + t)\right) \\
&\leq \mathbb{P}\left(2\max_{x\in\mathcal{N}} \left|\|Ax\|_2 - \sqrt{m}\right| > 2CK^2(\sqrt{n\log 9} + t)\right) \\
&\leq \sum_{x\in\mathcal{N}} \mathbb{P}\left(\left|\|Ax\|_2 - \sqrt{m}\right| > CK^2(\sqrt{n\log 9} + t)\right) \\
&\leq 9^n \cdot 2 \cdot 9^{-n} \cdot e^{-t^2} = 2e^{-t^2}.
\end{aligned}
$$

Scaling C accommodates the additional $\log 9$ factor finishes the proof. ⊛

4.7 Application: covariance estimation and clustering

Problem (Exercise 4.7.3). Our argument also implies the following high-probability guarantee. Check that for any $u \geq 0$, we have

$$
\|\Sigma_m - \Sigma\| \leq CK^2\left(\sqrt{\frac{n+u}{m}} + \frac{n+u}{m}\right)\|\Sigma\|
$$

with probability at least $1 - 2e^{-u}$.

Answer. Omit ⊛

Problem (Exercise 4.7.6). Prove Theorem 4.7.5 for the spectral clustering algorithm applied for the Gaussian mixture model. Proceed as follows.

(a) Compute the covariance matrix Σ of X; note that the eigenvector corresponding to the largest eigenvalue is parallel to μ.

(b) Use results about covariance estimation to show that the sample covariance matrix Σ_m is close to Σ, if the sample size m is relatively large.

(c) Use the Davis-Kahan Theorem 4.5.5 to deduce that the first eigenvector $v = v_1(\Sigma_m)$ is close to the direction of μ.

(d) Conclude that the signs of $\langle \mu, X_i \rangle$ predict well which community X_i belongs to.

(e) Since $v \approx \mu$, conclude the same for v.

Answer. Omit ⊛

Chapter 5

Concentration without independence

Week 17: Concentration of Lipschitz Functions on Spheres

5.1 Concentration of Lipschitz functions on the sphere

22 Jun. 2024

Problem (Exercise 5.1.2). Prove the following statements.

(a) Every Lipschitz function is uniformly continuous.

(b) Every differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$ is Lipschitz, and

$$\|f\|_{\mathrm{Lip}} \leq \sup_{x \in \mathbb{R}^n} \|\nabla f(x)\|_2.$$

(c) Give an example of a non-Lipschitz but uniformly continuous function $f\colon [-1, 1] \to \mathbb{R}$.

(d) Give an example of a non-differentiable but Lipschitz function $f\colon [-1, 1] \to \mathbb{R}$.

Answer. Omit. ⊛

Problem (Exercise 5.1.3). Prove the following statements.

(a) For a fixed $\theta \in \mathbb{R}^n$, the linear functional

$$f(x) = \langle x, \theta \rangle$$

is a Lipschitz function on $\mathbb{R}^n$, and $\|f\|_{\mathrm{Lip}} = \|\theta\|_2$.

(b) More generally, an $m \times n$ matrix A acting as a linear operator

$$A\colon (\mathbb{R}^n, \|\cdot\|_2) \to (\mathbb{R}^m, \|\cdot\|_2)$$

is Lipschitz, and $\|A\|_{\mathrm{Lip}} = \|A\|$.

(c) Any norm $f(x) = \|x\|$ on $(\mathbb{R}^n, \|\cdot\|_2)$ is a Lipschitz function. The Lipschitz norm of f is the smallest L that satisfies

$$\|x\| \leq L\|x\|_2 \text{ for all } x \in \mathbb{R}^n.$$

Answer. Omit. ⊛

Problem (Exercise 5.1.8). Prove inclusion (5.2), i.e., $H_t \supseteq \{x \in \sqrt{n}S^{n-1} \colon x_1 \leq t/\sqrt{2}\}$.

Answer. Omit. ⊛

Problem (Exercise 5.1.9). Let A be the subset of the sphere $\sqrt{n}S^{n-1}$ such that

$$\sigma(A) > 2\exp\left(-cs^2\right) \text{ for some } s > 0.$$

(a) Prove that $\sigma(A_s) > 1/2$.

(b) Deduce from this that for any $t \geq s$,

$$\sigma(A_{2t}) \geq 1 - 2\exp\left(-ct^2\right).$$

Here $c > 0$ is the absolute constant from Lemma 5.1.7.

Answer. Omit. ⊛

Problem (Exercise 5.1.11). We proved Theorem 5.1.4 for functions f that are Lipschitz with respect to the Euclidean metric $\|x - y\|_2$ on the sphere. Argue that the same result holds for the geodesic metric, which is the length of the shortest arc connecting x and y.

Answer. Omit. ⊛

Problem (Exercise 5.1.12). We stated Theorem 5.1.4 for the scaled sphere $\sqrt{n}S^{n-1}$. Deduce that a Lipschitz function f on the *unit* sphere S^{n-1} satisfies

$$\|f(X) - \mathbb{E}[f(X)]\|_{\psi_2} \leq \frac{C\|f\|_{\text{Lip}}}{\sqrt{n}},$$

where $X \sim \mathcal{U}(S^{n-1})$. Equivalently, for every $t \geq 0$, we have

$$\mathbb{P}\left(|f(X) - \mathbb{E}[f(X)]| \geq t\right) \leq 2\exp\left(-\frac{cnt^2}{\|f\|_{\text{Lip}}^2}\right).$$

Answer. Omit. ⊛

Problem (Exercise 5.1.13). Consider a random variable Z with median M. Show that

$$c\|Z - \mathbb{E}[Z]\|_{\psi_2} \leq \|Z - M\|_{\psi_2} \leq C\|Z - \mathbb{E}[Z]\|_{\psi_2},$$

where $c, C > 0$ are some absolute constants.

Answer. Omit. ⊛

Problem (Exercise 5.1.14). Consider a random vector X taking values in some metric space (T, d). Assume that there exists $K > 0$ such that

$$\|f(X) - \mathbb{E}[f(X)]\|_{\psi_2} \leq K\|f\|_{\text{Lip}}$$

for every Lipschitz function $f\colon T \to \mathbb{R}$. For a subset $A \subseteq T$, define $\sigma(A) := \mathbb{P}(X \in A)$. (Then σ is a probability measure on T.) Show that if $\sigma(A) \geq 1/2$ then, for every $t \geq 0$,

$$\sigma(A_t) \geq 1 - 2\exp\left(-ct^2/K^2\right)$$

where $c > 0$ is an absolute constant.

Answer. Omit. ⊛

Problem (Exercise 5.1.15). From linear algebra, we know that any set of orthonormal vectors in $\mathbb{R}^n$

must contain at most n vectors. However, if we allow the vectors to be almost orthogonal, there can be *exponentially many* of them! Prove this counterintuitive fact as follows. Fix $\epsilon \in (0,1)$. Show that there exists a set $\{x_1, \dots, x_N\}$ of unit vectors in $\mathbb{R}^n$ which are mutually almost orthogonal:

$$|\langle x_i, x_j\rangle| \leq \epsilon \text{ for all } i \neq j,$$

and the set is *exponentially large* in n:

$$N \geq \exp(c(\epsilon)n).$$

Answer. Omit. ⊛

5.2 Concentration on other metric measure spaces

Problem (Exercise 5.2.3)**.** Deduce Gaussian concentration inequality (Theorem 5.2.2) from Gaussian isoperimetric inequality (Theorem 5.2.1).

Answer. Omit. ⊛

Problem (Exercise 5.2.4)**.** Prove that in the concentration results for sphere and Gauss space (Theorem 5.1.4 and 5.2.2), the expectation $\mathbb{E}[f(X)]$ can be replaced by the L^p norm $(\mathbb{E}[f(X)^p])^{1/p}$ for any $p \geq 1$ and for any non-negative function f. The constants may depend on p.

Answer. Omit. ⊛

Problem (Exercise 5.2.11)**.** Let $\Phi(x)$ denote the cumulative distribution function of the standard normal distribution $\mathcal{N}(0,1)$. Consider a random vector $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n)$. Check that

$$\phi(Z) \coloneqq (\Phi(Z_1), \dots, \Phi(Z_n)) \sim \mathcal{U}([0,1]^n).$$

Answer. Omit. ⊛

Problem (Exercise 5.2.12)**.** Expressing $X = \phi(Z)$ by the previous exercise, use Gaussian concentration to control the deviation of $f(\phi(Z))$ in terms of $\|f \circ \phi\|_{\mathrm{Lip}} \leq \|f\|_{\mathrm{Lip}}\|\phi\|_{\mathrm{Lip}}$. Show that $\|\phi\|_{\mathrm{Lip}}$ is bounded by an absolute constant and complete the proof of Theorem 5.2.10.

Answer. Omit. ⊛

Problem (Exercise 5.2.14)**.** Use a similar method to prove Theorem 5.2.13. Define a function $\phi\colon \mathbb{R}^n \to \sqrt{n}B_2^n$ that pushes forward the Gaussian measure on $\mathbb{R}^n$ into the uniform measure on $\sqrt{n}B_2^n$, and check that ϕ has bounded Lipschitz norm.

Answer. Omit. ⊛

5.3 Application: Johnson-Lindenstrauss Lemma

Problem (Exercise 5.3.3)**.** Let A be an $m \times n$ random matrix whose rows are independent, mean zero, sub-gaussian isotropic random vectors in $\mathbb{R}^n$. Show that the conclusion of Johnson-Lindenstrauss lemma holds for $Q = (1/\sqrt{m})A$.

Answer. Omit. ⊛

Problem (Exercise 5.3.4)**.** Give an example of a set $\mathcal{X}$ of N points for which no scaled projection onto a subspace of dimension $m \ll \log N$ is an approximate isometry.

Answer. Omit. ⊛

Week 18: Tighter Bounds on Sub-Gaussian Matrices

5.4 Matrix Bernstein's inequality

29 Jun. 2024

Problem (Exercise 5.4.3). (a) Consider a polynomial

$$f(x) = a_0 + a_1 x + \cdots + a_p x^p.$$

Check that for a matrix X, we have

$$f(X) = a_0 I + a_1 X + \cdots + a_p X^p.$$

In the right side, we use the standard rules for matrix addition and multiplication, so in particular, $X^p = X \cdots X$ (p times) there.

(b) Consider a convergent power series expansion of f about x_0:

$$f(x) = \sum_{k=1}^{\infty} a_k (x - x_0)^k.$$

Check that the series of matrix terms converges, and

$$f(X) = \sum_{k=1}^{\infty} a_k (X - x_0 I)^k.$$

Answer. Let $X =: U\Lambda U^\top$ be the symmetric eigendecomposition of X.

(a) Since $X^k = U\Lambda U^\top \cdots U\Lambda U^\top = U\Lambda I \cdots I\Lambda U^\top = U\Lambda^k U^\top$ for all $k \geq 0$, then

$$f(X) = U f(\Lambda) U^\top = U \left(\sum_{k=0}^{p} a_k \Lambda^k \right) U^\top = \sum_{k=0}^{p} a_k U \Lambda^k U^\top = \sum_{k=0}^{p} a_k X^k.$$

(b) Since $X - x_0 I = U(\Lambda - x_0 I)U^\top$, then by (a),

$$f(X) = U \left(\sum_{k=1}^{\infty} a_k (\Lambda - x_0 I)^k \right) U^\top = \sum_{k=1}^{\infty} a_k U(\Lambda - x_0 I)^k U^\top = \sum_{k=0}^{\infty} a_k (X - x_0 I)^k.$$

⊛

Problem (Exercise 5.4.5). Prove the following properties.

(a) $\|X\| \leq t$ if and only if $-tI \preceq X \preceq tI$.

(b) Let $f, g \colon \mathbb{R} \to \mathbb{R}$ be two functions. If $f(x) \leq g(x)$ for all $x \in \mathbb{R}$ satisfying $|x| \leq K$, then $f(X) \preceq g(X)$ for all X satisfying $\|X\| \leq K$.

(c) Let $f \colon \mathbb{R} \to \mathbb{R}$ be an increasing function and X, Y are commuting matrices. Then $X \preceq Y$ implies $f(X) \preceq f(Y)$.

(d) Given an example showing that property (c) may fail for non-commuting matrices.

(e) In the following parts of the exercise, we develop weaker versions of property (c) that hold for arbitrary, nor necessarily commuting, matrices. First, show that $X \preceq Y$ always implies $\operatorname{tr} f(X) \leq \operatorname{tr} f(Y)$ for any increasing function $f : \mathbb{R} \to \mathbb{R}$.

(f) Show that $0 \preceq X \preceq Y$ implies $X^{-1} \preceq Y^{-1}$ if X is invertible.

(g) Show that $0 \preceq X \preceq Y$ implies $\log X \preceq \log Y$.

Answer. Let $X =: U\Lambda U^\top$ and $Y =: VMV^\top$ denote the symmetric eigendecompositions of X and Y, respectively. Additionally, let $\lambda := \operatorname{diag}(\Lambda)$ and $\mu := \operatorname{diag}(M)$ in $\mathbb{R}^n$.

(a) By the Courant-Fisher min-max theorem w.r.t. λ_1 and λ_n,

$$\|X\| \leq t \Leftrightarrow -t\mathbb{1} \leq \lambda \leq t\mathbb{1} \Leftrightarrow t\mathbb{1} \pm \lambda \geq 0 \Leftrightarrow tI \pm X \succeq 0 \Leftrightarrow -tI \preceq X \preceq tI.$$

(b) Since $|\lambda| \leq K\mathbb{1}$, then $g(\lambda) - f(\lambda) \geq 0$. This implies that $g(X) - f(X) = U(g(\Lambda) - f(\Lambda))U^\top$ has non-negative eigenvalues. Therefore, $g(X) \succeq f(X)$.

(c) Since X and Y are symmetric and commute, then Y admits an eigendecomposition with $V = U$. This implies $\lambda \leq \mu$. It follows that $f(\mu) - f(\lambda) \geq 0$, so $f(Y) - f(X) = U(f(M) - f(\Lambda))U^\top$ has non-negative eigenvalues. Therefore, $f(X) \preceq f(Y)$.

(d) We see that

$$\lambda\left(\begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} - \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\right) = \{5, 0\},$$

while

$$\lambda\left(\begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}^3 - \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}^3\right) = \left\{\frac{\sqrt{43993} + 197}{2}, -\frac{\sqrt{43993} - 197}{2}\right\}.$$

(e) Since $X - Y \preceq 0$, then by the Courant-Fisher min-max theorem, for any $i = 1, \ldots, n$,

$$\begin{aligned}
\lambda_i - \mu_i &= \max_{\dim E = i} \min_{v \in S(E)} v^\top X v - \max_{\dim E = i} \min_{v \in S(E)} v^\top Y v \\
&\leq \max_{\dim E = i} \left(\min_{v \in S(E)} v^\top X v - \min_{v \in S(E)} v^\top Y v \right) \\
&\leq \max_{\dim E = i} \max_{v \in S(E)} \left(v^\top X v - v^\top Y v \right) = \max_{\dim E = i} \max_{v \in S(E)} v^\top (X - Y) v \leq 0
\end{aligned}$$

Since f is increasing, then $f(\lambda_i) \leq f(\mu_i)$ for all i. It follows that

$$\operatorname{tr} f(X) = \sum_{i=1}^n f(\lambda_i) \leq \sum_{i=1}^n f(\mu_i) = \operatorname{tr} f(Y).$$

(f) Since $X \preceq Y$, then $I = X^{-1/2} X X^{-1/2} \preceq X^{-1/2} Y X^{-1/2}$. This implies $\lambda(X^{-1/2} Y X^{-1/2}) \geq 1$. Thus, $\lambda(X^{1/2} Y^{-1} X^{1/2}) = \lambda^{-1}(X^{-1/2} Y X^{-1/2}) \leq 1$, so $X^{1/2} Y^{-1} X^{1/2} \preceq I$. It follows that

$$Y^{-1} = X^{-1/2}(X^{1/2} Y^{-1} X^{1/2}) X^{-1/2} \preceq X^{-1/2} I X^{-1/2} = X^{-1}.$$

(g) By (f), $(X + tI)^{-1} \succeq (Y + tI)^{-1}$ for $t \geq 0$. Since $\log z = \log \frac{1+t}{z+t}\Big|_{t=0}^{t=\infty} = \int_0^\infty \frac{1}{1+t} - \frac{1}{z+t} \,\mathrm{d}t$, then

$$\log X = \int_0^\infty ((1+t)^{-1} I - (X + tI)^{-1}) \,\mathrm{d}t \preceq \int_0^\infty ((1+t)^{-1} I - (Y + tI)^{-1}) \,\mathrm{d}t = \log Y.$$

⊛

Problem (Exercise 5.4.6). Let X and Y be $n \times n$ symmetric matrices.

(a) Show that if the matrices commute, i.e., $XY = YX$, then

$$e^{X+Y} = e^X e^Y.$$

(b) Find and example of matrices X and Y such that

$$e^{X+Y} \neq e^X e^Y.$$

Answer. (a) Since X and Y commute, by the binomial theorem and the substitution $i := k - j$,

$$e^{X+Y} = \sum_{k=0}^{\infty} \frac{(X+Y)^k}{k!} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{j=0}^{k} \frac{k!}{(k-j)!j!} X^{k-j} Y^j = \sum_{i=0}^{\infty} \frac{X^i}{i!} \sum_{j=0}^{\infty} \frac{Y^j}{j!} = e^X e^Y.$$

(b) For $X := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $Y := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$,

$$e^{X+Y} = \begin{pmatrix} \cosh\sqrt{2} + \frac{\sinh\sqrt{2}}{\sqrt{2}} & \frac{\sinh\sqrt{2}}{\sqrt{2}} \\ \frac{\sinh\sqrt{2}}{\sqrt{2}} & \cosh\sqrt{2} - \frac{\sinh\sqrt{2}}{\sqrt{2}} \end{pmatrix}, \quad e^X e^Y = \frac{1}{2} \begin{pmatrix} e^2+1 & e^2-1 \\ 1-e^{-2} & 1+e^{-2} \end{pmatrix}.$$

⊛

Problem (Exercise 5.4.11). Let $X_1, \ldots, X_N$ be independent, mean zero, $n \times n$ symmetric random matrices, such that $\|X_i\| \leq K$ almost surely for all i. Deduce from Bernstein's inequality that

$$\mathbb{E}\left[\left\|\sum_{i=1}^{N} X_i\right\|\right] \lesssim \left\|\sum_{i=1}^{N} \mathbb{E}[X_i^2]\right\|^{1/2} \sqrt{1+\log n} + K(1+\log n).$$

Answer. Let $\sigma^2 := \|\sum_{i=1}^N \mathbb{E}[X_i^2]\|$. By the matrix Berstein's inequality, for every $u > 0$, with the substitution $t := c^{-1/2}\sigma\sqrt{u+\log n} + c^{-1}K(u+\log n)$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^{N} X_i\right\| \geq t\right) \leq 2ne^{-c\min\left(\frac{t^2}{\sigma^2}, \frac{t}{K}\right)} \leq 2ne^{-(u+\log n)} = 2e^{-u}.$$

Then by Lemma 1.2.1,

$$\begin{aligned}
&\mathbb{E}\left[\left\|\sum_{i=1}^{N} X_i\right\|\right] \\
&= \left(\int_0^{c^{-1/2}\sigma\sqrt{1+\log n}+c^{-1}K(1+\log n)} + \int_{c^{-1/2}\sigma\sqrt{1+\log n}+c^{-1}K(1+\log n)}^{\infty}\right) \mathbb{P}\left(\left\|\sum_{i=1}^{N} X_i\right\| \geq t\right) \,\mathrm{d}t \\
&\leq \int_0^{c^{-1/2}\sigma\sqrt{1+\log n}+c^{-1}K(1+\log n)} 1 \,\mathrm{d}t + \int_{c^{-1/2}\sigma\sqrt{1+\log n}+c^{-1}K(1+\log n)}^{\infty} 2e^{-u} \,\mathrm{d}t \\
&= c^{-1/2}\sigma\sqrt{1+\log n} + c^{-1}K(1+\log n) + \int_1^{\infty} 2e^{-u} \left(\frac{2^{-1}c^{-1/2}\sigma}{\sqrt{u+\log n}} + c^{-1}K\right) \mathrm{d}u \\
&\leq c^{-1/2}\sigma\sqrt{1+\log n} + c^{-1}K(1+\log n) + \int_1^{\infty} 2e^{-u} \left(\frac{2^{-1}c^{-1/2}\sigma}{\sqrt{1+\log n}} + c^{-1}K\right) \mathrm{d}u \\
&= c^{-1/2}\sigma\sqrt{1+\log n} + c^{-1}K(1+\log n) + 2e^{-1} \left(\frac{2^{-1}c^{-1/2}\sigma}{\sqrt{1+\log n}} + c^{-1}K\right) \\
&\lesssim \sigma\sqrt{1+\log n} + K(1+\log n),
\end{aligned}$$

which is exactly what we want to show. ⊛

Problem (Exercise 5.4.12). Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent symmetric Bernoulli random variables and let $A_1, \ldots, A_N$ be symmetric $n \times n$ matrices (deterministic). Prove that, for any $t \geq 0$, we have

$$\mathbb{P}\left(\left\|\sum_{i=1}^N \varepsilon_i A_i\right\| \geq t\right) \leq 2n\exp\left(-t^2/2\sigma^2\right),$$

where $\sigma^2 = \|\sum_{i=1}^N A_i^2\|$.

Answer. Let $\sigma^2 := \|\sum_{i=1}^N A_i^2\|$ and $\lambda := t/\sigma^2 \geq 0$. By Exercise 2.2.3,

$$\operatorname{tr} e^{\sum_{i=1}^N \log \mathbb{E}[e^{\lambda\varepsilon_i A_i}]} = \operatorname{tr} e^{\sum_{i=1}^N \log\cosh(\lambda A_i)} \leq \operatorname{tr} e^{\sum_{i=1}^N \frac{\lambda^2}{2} A_i^2} \leq n e^{\frac{\lambda^2}{2}\lambda_{\max}(\sum_{i=1}^N A_i^2)} = n e^{\frac{\lambda^2\sigma^2}{2}}$$

Then by the Chernoff bound and Lieb's inequality,

$$\begin{aligned}\mathbb{P}\left(\lambda_{\max}\left(\sum_{i=1}^N \varepsilon_i A_i\right) \geq t\right) &\leq e^{-\lambda t}\mathbb{E}[e^{\lambda\cdot\lambda_{\max}(\sum_{i=1}^N \varepsilon_i A_i)}] \\ &= e^{-\lambda t}\operatorname{tr} e^{\sum_{i=1}^N \log\mathbb{E}[e^{\lambda\varepsilon_i A_i}]} \leq e^{-\lambda t} n e^{\frac{\lambda^2\sigma^2}{2}} = n e^{-\frac{t^2}{2\sigma^2}}.\end{aligned}$$

Similarly, $\mathbb{P}(\lambda_{\min}(\sum_{i=1}^N \varepsilon_i A_i) \leq -t) \leq n e^{-\frac{t^2}{2\sigma^2}}$. ⊛

Problem (Exercise 5.4.13). Let $\varepsilon_1, \ldots, \varepsilon_N$ be independent symmetric Bernoulli random variables and let $A_1, \ldots, A_N$ be symmetric $n \times n$ matrices (deterministic).

1. Prove that

$$\mathbb{E}\left[\left\|\sum_{i=1}^N \varepsilon_i A_i\right\|\right] \leq C\sqrt{1+\log n}\left\|\sum_{i=1}^N A_i^2\right\|^{1/2}.$$

2. More generally, prove that for every $p \in [1, \infty)$, we have

$$\left(\mathbb{E}\left[\left\|\sum_{i=1}^N \varepsilon_i A_i\right\|^p\right]\right)^{1/p} \leq C\sqrt{p+\log n}\left\|\sum_{i=1}^N A_i^2\right\|^{1/2}.$$

Answer. Since (a) follows from (b) with $p = 1$, we will only prove (b) here. As the inequality trivially holds for $n = 1$ with $C = 1$, let's assume $n \geq 2$ from now on.

Note that if $1 \leq p \leq 2$, then by Stirling's approximation $\Gamma(z) \leq \sqrt{\frac{2\pi}{z}}\left(\frac{z}{e}\right)^z e^{\frac{1}{12z}}$,

$$\left(\int_0^\infty e^{-s}(\log(2n)+s)^{\frac{p}{2}-1}\,\mathrm{d}s\right)^{1/p} \leq \left(\int_0^\infty e^{-s}(0+s)^{\frac{p}{2}-1}\,\mathrm{d}s\right)^{1/p} = \Gamma\left(\frac{p}{2}\right)^{1/p} \leq \frac{\pi^{\frac{1}{2p}} p^{\frac{p-1}{2p}}}{2^{\frac{1}{2}-\frac{1}{p}} e^{\frac{1}{2}-\frac{1}{6p^2}}},$$

and that if $p > 2$, then by Minkowski's inequality (and the same Stirling's approximation),

$$\left(\int_0^\infty e^{-s}(\log(2n)+s)^{\frac{p}{2}-1}\,\mathrm{d}s\right)^{1/p} = \left(\int_0^\infty (e^{-\frac{s}{p}}(\log(2n)+s)^{\frac{1}{2}-\frac{1}{p}})^p\,\mathrm{d}s\right)^{1/p}$$

since $p > 2$, and as $x, y > 0$, we have $(x+y)^{\frac{1}{2}-\frac{1}{p}} \leq x^{\frac{1}{2}-\frac{1}{p}} + y^{\frac{1}{2}-\frac{1}{p}}$,

$$\leq \left(\int_0^\infty (e^{-\frac{s}{p}}((\log(2n))^{\frac{1}{2}-\frac{1}{p}} + s^{\frac{1}{2}-\frac{1}{p}}))^p\,\mathrm{d}s\right)^{1/p}$$

then by Minkowski's inequality (i.e., $\|f+g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}$),

$$\leq \left(\int_0^\infty (e^{-\frac{s}{p}}(\log(2n))^{\frac{1}{2}-\frac{1}{p}})^p\,\mathrm{d}s\right)^{1/p} + \left(\int_0^\infty (e^{-\frac{s}{p}} s^{\frac{1}{2}-\frac{1}{p}})^p\,\mathrm{d}s\right)^{1/p}$$

then by some direct calculations,

$$
\begin{aligned}
&= (\log(2n))^{\frac{1}{2}-\frac{1}{p}} \left(\int_0^\infty e^{-s}\,\mathrm{d}s\right)^{1/p} + \left(\int_0^\infty e^{-s} s^{\frac{p}{2}-1}\,\mathrm{d}s\right)^{1/p} \\
&= (\log(2n))^{\frac{1}{2}-\frac{1}{p}} + \Gamma\left(\frac{p}{2}\right)^{1/p} \\
&\leq (\log(2n))^{\frac{1}{2}-\frac{1}{p}} + \frac{\pi^{\frac{1}{2p}} p^{\frac{p-1}{2p}}}{2^{\frac{1}{2}-\frac{1}{p}} e^{\frac{1}{2}-\frac{1}{6p^2}}}.
\end{aligned}
$$

Let $\sigma^2 \coloneqq \lVert \sum_{i=1}^N A_i^2 \rVert$. By Exercise 5.4.12, for any $t \geq 0$,

$$
\mathbb{P}\left(\left\lVert \sum_{i=1}^N \varepsilon_i A_i \right\rVert^p \geq t\right) = \mathbb{P}\left(\left\lVert \sum_{i=1}^N \varepsilon_i A_i \right\rVert \geq t^{1/p}\right) \leq 2n e^{-\frac{t^{2/p}}{2\sigma^2}}.
$$

Then with the substitution $t =: (\sigma\sqrt{2(\log(2n)+s)})^p$, by Lemma 1.2.1 and Minkowski's inequality,

$$
\begin{aligned}
\left(\mathbb{E}\left[\left\lVert \sum_{i=1}^N \varepsilon_i A_i \right\rVert^p\right]\right)^{1/p} &= \left(\left(\int_0^{(\sigma\sqrt{2\log(2n)})^p} + \int_{(\sigma\sqrt{2\log(2n)})^p}^\infty\right) \mathbb{P}\left(\left\lVert \sum_{i=1}^N \varepsilon_i A_i \right\rVert^p \geq t\right)\mathrm{d}t\right)^{1/p} \\
&\leq \left(\int_0^{(\sigma\sqrt{2\log(2n)})^p} 1\,\mathrm{d}t + \int_{(\sigma\sqrt{2\log(2n)})^p}^\infty 2n e^{-\frac{t^{2/p}}{2\sigma^2}}\,\mathrm{d}t\right)^{1/p} \\
&= \left((\sigma\sqrt{2\log(2n)})^p + \int_0^\infty e^{-s}\frac{(\sqrt{2}\sigma)^p p}{2}(\log(2n)+s)^{\frac{p}{2}-1}\,\mathrm{d}s\right)^{1/p} \\
&= \sqrt{2}\sigma\left(\sqrt{\log(2n)}^p + \frac{p}{2}\int_0^\infty e^{-s}(\log(2n)+s)^{\frac{p}{2}-1}\,\mathrm{d}s\right)^{1/p} \\
&\leq \sqrt{2}\sigma\left(\sqrt{\log(2n)} + \left(\frac{p}{2}\right)^{1/p}\left(\int_0^\infty e^{-s}(\log(2n)+s)^{\frac{p}{2}-1}\,\mathrm{d}s\right)^{1/p}\right)
\end{aligned}
$$

plugging in the bound we have established in the beginning,

$$
\begin{aligned}
&\leq \sqrt{2}\sigma\left(\sqrt{\log(2n)} + \left(\frac{p}{2}\right)^{1/p}\left((\log(2n))^{\frac{1}{2}-\frac{1}{p}}\mathbb{1}_{p>2} + \frac{\pi^{\frac{1}{2p}} p^{\frac{p-1}{2p}}}{2^{\frac{1}{2}-\frac{1}{p}} e^{\frac{1}{2}-\frac{1}{6p^2}}}\right)\right) \\
&= \sqrt{2}\sigma\left(\left(1 + \left(\frac{p}{2}\right)^{1/p}(\log(2n))^{-\frac{1}{p}}\mathbb{1}_{p>2}\right)\sqrt{\log(2n)} + \frac{\pi^{\frac{1}{2p}} p^{\frac{p+1}{2p}}}{\sqrt{2}e^{\frac{1}{2}-\frac{1}{6p^2}}}\right) \\
&\leq \sqrt{2}\sigma\left((1 + e^{\frac{1}{e\log(16)}})\sqrt{\log(2n)} + \frac{\sqrt{\pi}}{\sqrt{2}e^{\frac{1}{3}}}\sqrt{p}\right) \\
&\asymp \sigma\sqrt{p + \log n},
\end{aligned}
$$

which is exactly what we want to show. ⊛

Problem (Exercise 5.4.14). Let X be an $n \times n$ random matrix that takes values $e_k e_k^\top$, $k = 1, \ldots, n$, with probability $1/n$ each. (Here (e_k) denotes the standard basis in $\mathbb{R}^n$.) Let $X_1, \ldots, X_N$ be independent copies of X. Consider the sum $S = \sum_{i=1}^N X_i$, which is a diagonal matrix.

(a) Show that the entry S_{ii} has the same distribution as the number of balls in i-th bin when N balls are thrown into n bins independently.

(b) Relating this to the classical coupon collector's problem, show that if $N \asymp n$, then

$$
\mathbb{E}\lVert S \rVert \asymp \frac{\log n}{\log\log n}.
$$

Deduce that the bound in Exercise 5.4.11 would fail if the logarithmic factors were removed from it.

Answer. (a) We see that $X = e_k e_k^\top$ with k being chosen uniformly randomly among $[n]$, where $e_k e_k^\top$ is a matrix with all 0's except the k^{th} diagonal element being 1. Hence, by interpreting each X_i as "throwing a ball into n bins," S_{kk} records the number of balls in the k^{th} bin when N balls are thrown into n bins independently.

(b) We first observe that since S is diagonal, $\|S\| = \lambda_1(S) = \max_k S_{kk}$ as all the diagonal elements are eigenvalues of S. We first answer the question of how this related to the coupon collector's problem. Firstly, let's introduce the problem formally:

Problem 5.4.1 (Coupon collector's problem). Say we have n different types of coupons to collect, and we buy N boxes, where each box contains a (uniformly) random type of coupon. The classical *coupon collector's problem* asks for the expected number of boxes (i.e., N) we need in order to collect all coupons.

Intuition. From (a), we can view S_{kk} as the number of coupons we have collected for the k^{th} type of the coupon, where N is the number of boxes we have bought.

Hence, the coupon collector's problem asks for the expected N we need for $\lambda_n(S) = \min_k S_{kk} > 0$, while (b) is asking for the expected number of *the most frequent* coupons (i.e., $\max_k S_{kk}$) we will see when buying only $N \asymp n$ boxes.

Next, let's prove the upper bound and the lower bound separately. Let $0 < c < C$ to be some constants satisfying $N \leq Cn$ and $n \leq cN$.

Claim (Upper bound). $\mathbb{E}[\|S\|] \lesssim \log n / \log\log n$.

Proof. We first note that $S_{kk} \sim \operatorname{Binomial}(N, 1/n)$ for all k, so by Exercise 2.4.3, for any $m > N/n$, we have

Fix

$$\mathbb{P}(\|S\| \geq m) = \mathbb{P}(\exists k \colon S_{kk} \geq m) \leq \sum_{k=1}^{n} \mathbb{P}(S_{kk} \geq m) \leq 3^{\frac{N}{n} + 1 - \frac{m \log\log n}{\log n}}.$$

Let $L \coloneqq \left\lfloor \frac{(C+1)\log n}{\log\log n} \right\rfloor + 1 > C + 1 > N/n$, then

$$\begin{aligned}
\mathbb{E}[\|S\|] &= \left(\sum_{m=1}^{L-1} + \sum_{m=L}^{\infty} \right) \mathbb{P}(\|S\| \geq m) \\
&\leq \sum_{m=1}^{L-1} 1 + \sum_{m=L}^{\infty} 3^{\frac{N}{n} + 1 - \frac{m \log\log n}{\log n}} \\
&= L - 1 + \frac{3^{\frac{N}{n} + 1 - \frac{L \log\log n}{\log n}}}{1 - 3^{-\frac{\log\log n}{\log n}}} \leq \frac{(C+1)\log n}{\log\log n} + \frac{3^{C+1-(C+1)}}{\frac{2}{3} \cdot \frac{\log\log n}{\log n}} = \frac{(C + \frac{5}{2})\log n}{\log\log n},
\end{aligned}$$

establishing the desired upper bound. ⊛

The hard part lies in the lower bound. We will need the following fact.

Lemma 5.4.1 (Maximum of Poisson [Kim83; BSP09]). Given $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \operatorname{Pois}(1)$,

$$\mathbb{E}\left[\max_{1 \leq k \leq n} Y_k\right] \asymp \frac{\log n}{\log\log n}.$$

Such a concentration is *very* tight.

Claim (Lower bound). $\mathbb{E}[\|S\|] \gtrsim \log n / \log\log n$

Proof. Let $M_T := \mathbb{E}[\max_k\{Y_k\} \mid \sum_{k=1}^n Y_k = T]$ with $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \mathrm{Pois}(1)$. As $(Y_1, \ldots, Y_n) \mid \sum_{k=1}^n Y_k = T \sim \mathrm{Multinomial}(T; 1/n, \ldots, 1/n)$, we know that M_T is non-decreasing w.r.t. T. Moreover, as $\sum_{k=1}^n Y_k \sim \mathrm{Pois}(n)$, by the law of total expectation and maximum of Poisson lemma,

$$
\begin{aligned}
\frac{\log n}{\log\log n} \asymp \mathbb{E}\left[\max_{1\leq k\leq n} Y_k\right] &= \left(\sum_{T=0}^{\lfloor ne^{2+\frac{1}{2e}}\rfloor} + \sum_{T=\lfloor ne^{2+\frac{1}{2e}}\rfloor+1}^{\infty}\right)\frac{e^{-n}n^T}{T!}M_T \\
&\leq \sum_{T=0}^{\lfloor ne^{2+\frac{1}{2e}}\rfloor}\frac{e^{-n}n^T}{T!}M_{\lfloor ne^{2+\frac{1}{2e}}\rfloor} + \sum_{T=\lfloor ne^{2+\frac{1}{2e}}\rfloor+1}^{\infty}\frac{e^{-n}n^T}{T!}T \\
&\leq M_{\lfloor ne^{2+\frac{1}{2e}}\rfloor}\cdot 1 + \sum_{T=\lfloor ne^{2+\frac{1}{2e}}\rfloor+1}^{\infty}\frac{e^{-n}n^T}{\Gamma(T)}
\end{aligned}
$$

From Stirling's approximation, $\Gamma(z) \geq \sqrt{2\pi}z^{z-1/2}e^{-z}$ for $z > 0$,

$$
\begin{aligned}
&\leq M_{\lfloor ne^{2+\frac{1}{2e}}\rfloor} + \sum_{T=\lfloor ne^{2+\frac{1}{2e}}\rfloor+1}^{\infty}\frac{e^{-n}n^T}{\sqrt{2\pi}T^{T-1/2}e^{-T}} \\
&= M_{\lfloor ne^{2+\frac{1}{2e}}\rfloor} + \frac{e^{-n}}{\sqrt{2\pi}}\sum_{T=\lfloor ne^{2+\frac{1}{2e}}\rfloor+1}^{\infty}\left(\frac{neT^{\frac{1}{2T}}}{T}\right)^T
\end{aligned}
$$

since for all $x > 0$, $x^{1/2x} \leq e^{1/2e}$, for $x = T$, we have

$$
\begin{aligned}
&\leq M_{\lfloor ne^{2+\frac{1}{2e}}\rfloor} + \frac{e^{-n}}{\sqrt{2\pi}}\sum_{T=\lfloor ne^{2+\frac{1}{2e}}\rfloor+1}^{\infty}\left(\frac{ne^{1+\frac{1}{2e}}}{T}\right)^T \\
&\leq M_{\lfloor ne^{2+\frac{1}{2e}}\rfloor} + \frac{e^{-n}}{\sqrt{2\pi}}\sum_{T=\lfloor ne^{2+\frac{1}{2e}}\rfloor+1}^{\infty} e^{-T} \\
&= M_{\lfloor ne^{2+\frac{1}{2e}}\rfloor} + \frac{e^{-n-\lfloor ne^{2+\frac{1}{2e}}\rfloor-1}}{\sqrt{2\pi}(1-e^{-1})},
\end{aligned}
$$

leading to

$$
M_{\lfloor ne^{2+\frac{1}{2e}}\rfloor} \gtrsim \frac{\log n}{\log\log n}
$$

as the trailing term is decreasing exponentially fast. Finally, we have

$$
M_{\lfloor ne^{2+\frac{1}{2e}}\rfloor} \leq M_{\left\lceil\frac{\lfloor ne^{2+\frac{1}{2e}}\rfloor}{N}\right\rceil N} \leq \left\lceil\frac{\lfloor ne^{2+\frac{1}{2e}}\rfloor}{N}\right\rceil M_N \leq \left\lceil\frac{ne^{2+\frac{1}{2e}}}{N}\right\rceil M_N \leq \lceil ce^{2+\frac{1}{2e}}\rceil M_N,
$$

where the second inequality follows from the triangle inequality of max. This leads to

$$
\mathbb{E}[\|S\|] = M_N \geq \frac{1}{\lceil ce^{2+\frac{1}{2e}}\rceil}M_{\lfloor ne^{2+\frac{1}{2e}}\rfloor} \gtrsim \frac{\log n}{\log\log n}
$$

as desired. ⊛

Finally, the bound in Exercise 5.4.11 will fail if the logarithmic factors were removed becomes obvious after a direct substitution. Indeed, since $\|X_i\| = 1 =: K$, Exercise 5.4.11 states that

$$
\mathbb{E}\left[\left\|\sum_{i=1}^N X_i\right\|\right] \lesssim \left\|\sum_{i=1}^N \mathbb{E}[X_i^2]\right\|^{1/2},
$$

where the logarithmic factors were removed along with $K = 1$. Now, using the bound for $S := \sum_{i=1}^{N} X_i$ we have, with the observation that $X_i^2 = X_i$ and $\mathbb{E}[X_i^2] = \mathbb{E}[X_i] = \operatorname{diag}(1/n, \dots, 1/n)$, we see that the bound becomes $\sqrt{N/n} = \Theta(1)$, while the left-hand side grows as $\log n / \log\log n \to \infty$, which is clear not valid.

⊛

Remark (Alternative examples). We give another example to demonstrate the sharpness of the matrix Bernstein's inequality. Consider the following random $n \times n$ matrix (slightly different from S)

$$T := \sum_{i=1}^{N} \sum_{k=1}^{n} b_{ik}^{(N)} e_k e_k^\top,$$

where $b_{ik}^{(N)} \overset{\text{i.i.d.}}{\sim} \operatorname{Ber}(1/N)$. Here, we view $X_i := \sum_{k=1}^{n} b_{ik}^{(N)} e_k e_k^\top$

Intuition. In expectation, T and S should behave the same. However, this is easier to work with from independence.

Claim. As $N \to \infty$, with $Y_k \sim \operatorname{Pois}(1)$, we have

$$\mathbb{E}[\lambda_1(T)] = \mathbb{E}\left[\max_{1 \le k \le n} T_{kk}\right] \to \mathbb{E}\left[\max_{1 \le k \le n} Y_k\right] = \Theta\left(\frac{\log n}{\log\log n}\right).$$

Noticeably, the above claim doesn't require n to vary with N.

Proof. For every $k \in [n]$, we apply the Poisson limit theorem since as $N \to \infty$, $p_{N,ik} = 1/N \to 0$ and $\mathbb{E}[S_N^k] = \mathbb{E}[\sum_{i=1}^{N} b_{ik}^{(N)}] = 1 =: \lambda$ as $N \to \infty$. So as $N \to \infty$, $S_N^k \overset{D}{\to} \operatorname{Pois}(1)$.

With a similar interpretation as in (a), we can interpret $S_N^k = \sum_{i=1}^{N} b_{ik}^{(N)}$ as the value of the k^{th} diagonal element of T, i.e., T_{kk}. Hence, as $N \to \infty$, for all k, $T_{kk} \overset{D}{\to} Y_k$ where $Y_k \overset{\text{i.i.d.}}{\sim} \operatorname{Pois}(1)$. Since T_{kk}'s are independent, we have $T \overset{D}{\to} \operatorname{diag}(Z_1, \dots, Z_n)$, therefore

$$\mathbb{E}[\lambda_1(T)] = \mathbb{E}\left[\max_{1 \le k \le n} T_{kk}\right] \to \mathbb{E}\left[\max_{1 \le k \le n} Y_k\right] \asymp \Theta\left(\frac{\log n}{\log\log n}\right)$$

from the maximum of Poisson lemma. ⊛

A simple calculation of $\|\sum_{i=1}^{N} \mathbb{E}[X_i^2]\|^{1/2}$ reveals that the logarithmic factors can't be removed.

Problem (Exercise 5.4.15). Let $X_1, \dots, X_N$ be independent, mean zero, $m \times n$ random matrices, such that $\|X_i\| \le K$ almost surely for all i. Prove that for $t \ge q0$, we have

$$\mathbb{P}\left(\left\|\sum_{i=1}^{N} X_i\right\| \ge qt\right) \le 2(m+n)\exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right),$$

where

$$\sigma^2 = \max\left(\left\|\sum_{i=1}^{N} \mathbb{E}[X_i^\top X_i]\right\|, \left\|\sum_{i=1}^{N} \mathbb{E}[X_i X_i^\top]\right\|\right).$$

Answer. Consider the following N independent $(m+n) \times (m+n)$ symmetric, mean 0 matrices

$$X_i' := \begin{pmatrix} 0_{n \times n} & X_i^\top \\ X_i & 0_{m \times m} \end{pmatrix}.$$

To apply the matrix Bernstein's inequality (Theorem 5.4.1), we need to show that $\|X_i'\| \le K'$ for some K', where we know that $\|X_i\| \leq K$. However, it's easy to see that since $\|X_i\| = \|X_i^\top\|$, we have $\|X_i'\| \leq K$ as well since the characteristic equation for X_i' is

$$\det(X_i' - \lambda I) = \det\left(\begin{pmatrix} -\lambda I & X_i^\top \\ X_i & -\lambda I \end{pmatrix}\right) = \det\left(\lambda^2 I - X_i^\top X_i\right) = 0,$$

so $\|X_i'\| = \sqrt{\|X_i^\top X_i\|} \leq K$.

Claim. Actually, we have $\|X_i'\| = \|X_i\|$, hence $\|X_i'\| \leq K$.

Proof. Observe that for any matrix $A \in \mathbb{R}^{m\times n}$, as $\|A\| = \sqrt{\lambda_1(AA^\top)} = \sqrt{\lambda_1(A^\top A)}$, we have

$$\|A\| = \sqrt{\lambda_1\left(\begin{pmatrix} A^\top A & 0 \\ 0 & AA^\top \end{pmatrix}\right)} = \sqrt{\lambda_1\left(\begin{pmatrix} 0 & A^\top \\ A & 0 \end{pmatrix}^2\right)} = \left\|\begin{pmatrix} 0 & A^\top \\ A & 0 \end{pmatrix}\right\|.$$

Plugging in $X_i =: A$, we're done. ⊛

Hence, from matrix Bernstein's inequality, for every $t \geq q0$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^N X_i\right\| \geq qt\right) \leq 2(m+n)\exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right),$$

where $\sigma^2 = \|\sum_{i=1}^N \mathbb{E}[(X_i')^2]\|$. A quick calculation reveals that

$$(X_i')^2 = \begin{pmatrix} 0 & X_i^\top \\ X_i & 0 \end{pmatrix}\begin{pmatrix} 0 & X_i^\top \\ X_i & 0 \end{pmatrix} = \begin{pmatrix} X_i^\top X_i & 0 \\ 0 & X_i X_i^\top \end{pmatrix},$$

hence we have

$$\sigma^2 = \max\left(\left\|\sum_{i=1}^N \mathbb{E}[X_i^\top X_i]\right\|, \left\|\sum_{i=1}^N \mathbb{E}[X_i X_i^\top]\right\|\right),$$

which completes the proof. ⊛

5.5 Application: community detection in sparse networks

5.6 Application: covariance estimation for general distributions

Chapter 6

Quadratic forms, symmetrization and contraction

Week 19: Decoupling and Hanson-Wright Inequality

6.1 Decoupling

6 Jul. 2024

Problem (Exercise 6.1.4). Prove the following generalization of Theorem 6.1.1. Let $A = (a_{ij})$ be an $n \times n$ matrix. Let $X_1, \ldots, X_n$ be independent, mean zero random vectors in some Hilbert space. Show that for every convex function $F \colon \mathbb{R} \to \mathbb{R}$, one has

$$\mathbb{E}\left[F\left(\sum_{i,j\colon i\neq j} a_{ij}\langle X_i, X_j\rangle\right)\right] \leq \mathbb{E}\left[F\left(4\sum_{i,j} a_{ij}\langle X_i, X_j'\rangle\right)\right],$$

where (X_i') is an independent copy of (X_i).

Answer. Omit. ⊛

Problem (Exercise 6.1.5). Prove the following alternative generalization of Theorem 6.1.1. Let $(u_{ij})_{i,j=1}^n$ be fixed vectors in some normed space. Let $X_1, \ldots, X_n$ be independent, mean zero random variables. Show that, for every convex and increasing function F, one has

$$\mathbb{E}\left[F\left(\left\|\sum_{i,j\colon i\neq j} X_i X_j u_{ij}\right\|\right)\right] \leq \mathbb{E}\left[F\left(4\left\|\sum_{i,j} X_i X_j' u_{ij}\right\|\right)\right],$$

where (X_i') is an independent copy of (X_i).

Answer. Omit. ⊛

6.2 Hanson-Wright Inequality

Problem (Exercise 6.2.4). Complete the proof of Lemma 6.2.3. Replace X' by g'; write all details carefully.

Answer. Omit. ⊛

Problem (Exercise 6.2.5). Give an alternative proof of Hanson-Write inequality for normal distributions, without separating the diagonal part or decoupling.

Answer. Omit. ⊛

Problem (Exercise 6.2.6). Consider a mean zero, sub-gaussian random vector X in $\mathbb{R}^n$ with $\|X\|_{\psi_2} \leq K$. Let B be an $m \times n$ matrix. Show that

$$\mathbb{E}\left[\exp\left(\lambda^2 \|BX\|_2^2\right)\right] \leq \exp\left(CK^2\lambda^2\|B\|_F^2\right) \text{ provided } |\lambda| \leq \frac{c}{K\|B\|}.$$

To prove this bound, replace X with a Gaussian random vector $g \sim \mathcal{N}(0, I_m)$ along the following lines:

(a) Prove the comparison inequality

$$\mathbb{E}[\exp(\lambda^2\|BX\|_2^2)] \leq \mathbb{E}[\exp(CK^2\lambda^2\|B^\top g\|_2^2)]$$

for every $\lambda \in \mathbb{R}$.

(b) Check that

$$\mathbb{E}[\exp(\lambda^2\|B^\top g\|_2^2)] \leq \exp(C\lambda^2\|B\|_F^2)$$

provided that $|\lambda| \leq c/\|B\|$.

Answer. Omit. ⊛

Problem (Exercise 6.2.7). Let $X_1, \ldots, X_n$ be independent, mean zero, sub-gaussian random vectors in $\mathbb{R}^d$. Let $A = (a_{ij})$ be an $n \times n$ matrix. prove that for every $t \geq 0$, we have

$$\mathbb{P}\left(\left|\sum_{i,j\colon i\neq j}^{n} a_{ij}\langle X_i, X_j\rangle\right| \geq t\right) \leq 2\exp\left(-c\min\left(\frac{t^2}{K^4 d\|A\|_F^2}, \frac{t}{K^2\|A\|}\right)\right)$$

where $K = \max_i \|X_i\|_{\psi_2}$.

Answer. Omit. ⊛

6.3 Concentration of anisotropic random vectors

Problem (Exercise 6.3.1). Let B be an $m \times n$ matrix and X be an isotropic random vector in $\mathbb{R}^n$. Check that

$$\mathbb{E}[\|BX\|_2^2] = \|B\|_F^2.$$

Answer. Omit. ⊛

Problem (Exercise 6.3.3). Let D be a $k \times m$ matrix and B be an $m \times n$ matrix. Prove that

$$\|DB\|_F \leq \|D\|\|B\|_F.$$

Answer. Let $B = (b_1, \ldots, b_n)$, then

$$\|DB\|_F^2 = \sum_{i=1}^{n}\|Db_i\|_2^2 \leq \sum_{i=1}^{n}\|D\|^2\|b_i\|_2^2 = \|D\|^2\|B\|_F^2,$$

where we use the fact that $\|A\| = \sqrt{\sum_{i,j} a_{ij}^2}$ for any matrix A. ⊛

Problem (Exercise 6.3.4). Let E be a subspace of $\mathbb{R}^n$ of dimension d. Consider a random vector

$X = (X_1, \dots, X_n) \in \mathbb{R}^n$ with independent, mean zero, unit variance, sub-gaussian coordinates.

(a) Check that
$$(\mathbb{E}[\operatorname{dist}(X, E)^2])^{1/2} = \sqrt{n-d}.$$

(b) Prove that for any $t \geq 0$, the distance nicely concentrates:
$$\mathbb{P}\left(\left|\operatorname{dist}(X, E) - \sqrt{n-d}\right| > t\right) \leq 2\exp(-ct^2/K^4)$$
where $K = \max_i \|X_i\|_{\psi_2}$.

Answer. Omit. ⊛

Problem (Exercise 6.3.5). Let B be an $m \times n$ matrix, and let X be a mean zero, sub-gaussian random vector in $\mathbb{R}^n$ with $\|X\|_{\psi_2} \leq K$. Prove that for any $t \geq 0$, we have
$$\mathbb{P}(\|BX\|_2 \geq CK\|B\|_F + t) \leq \exp\left(-\frac{ct^2}{K^2\|B\|^2}\right).$$

Answer. Omit. ⊛

Problem (Exercise 6.3.6). Show that there exists a mean zero, isotropic, and sub-gaussian random vector X in $\mathbb{R}^n$ such that
$$\mathbb{P}(\|X\|_2 = 0) = \mathbb{P}(\|X\|_2 \geq 1.4\sqrt{n}) = \frac{1}{2}.$$
In other words, $\|X\|_2$ does not concentrate near $\sqrt{n}$.

Answer. Omit. ⊛

Week 20: The Symmetrization Trick

6.4 Symmetrization

13 Jul. 2024

Problem (Exercise 6.4.1). Let X be a random variable and ξ be an independent symmetric Bernoulli random variable.

(a) Check that ξX and $\xi|X|$ are symmetric random variables, and they have the same distribution.

(b) If X is symmetric, show that the distribution of ξX and $\xi|X|$ is the same as of x.

(c) Let X' be an independent copy of X. Check that $X - X'$ is symmetric.

Answer. (a) For any random variable X and a symmetric Bernoulli random variable ξ, we first prove that $\xi X \overset{D}{=} -\xi X$, i.e., $\mathbb{P}(\xi X \geq t) = \mathbb{P}(-\xi X \geq t)$ for any $t \in \mathbb{R}$. Indeed, since
$$\mathbb{P}(\xi X \geq t) = \frac{\mathbb{P}(\xi X \geq t \mid \xi = 1) + \mathbb{P}(\xi X \geq t \mid \xi = -1)}{2} = \frac{\mathbb{P}(X \geq t) + \mathbb{P}(-X \geq t)}{2}$$
while
$$\mathbb{P}(-\xi X \geq t) = \frac{\mathbb{P}(-\xi X \geq t \mid \xi = 1) + \mathbb{P}(-\xi X \geq t \mid \xi = -1)}{2} = \frac{\mathbb{P}(-X \geq t) + \mathbb{P}(X \geq t)}{2}.$$
This proves that both ξX and $\xi|X|$ are symmetric (by substituting X as $|X|$). Secondly, we

show that $\xi X \overset{D}{=} \xi|X|$, i.e., $\mathbb{P}(\xi X \geq t) = \mathbb{P}(\xi|X| \geq t)$ for any $t \in \mathbb{R}$. Again, we have

$$\begin{aligned}
\mathbb{P}(\xi|X| \geq t) &= \frac{\mathbb{P}(\xi|X| \geq t \mid \xi = 1) + \mathbb{P}(\xi|X| \geq t \mid \xi = -1)}{2} \\
&= \frac{\mathbb{P}(|X| \geq t) + \mathbb{P}(-|X| \geq t)}{2} \\
&= \frac{\mathbb{P}(|X| \geq t \mid X \geq 0)\mathbb{P}(X \geq 0) + \mathbb{P}(|X| \geq t \mid X < 0)\mathbb{P}(X < 0)}{2} \\
&\quad + \frac{\mathbb{P}(-|X| \geq t \mid X \geq 0)\mathbb{P}(X \geq 0) + \mathbb{P}(-|X| \geq t \mid X < 0)\mathbb{P}(X < 0)}{2} \\
&= \frac{(\mathbb{P}(X \geq t) + \mathbb{P}(-X \geq t))\mathbb{P}(X \geq 0) + (\mathbb{P}(-X \geq t) + \mathbb{P}(X \geq t))\mathbb{P}(X < 0)}{2} \\
&= \frac{(\mathbb{P}(X \geq t) + \mathbb{P}(-X \geq t))(\mathbb{P}(X \geq 0) + \mathbb{P}(X < 0))}{2} \\
&= \frac{\mathbb{P}(X \geq t) + \mathbb{P}(-X \geq t)}{2},
\end{aligned}$$

which is just $\mathbb{P}(\xi X \geq t)$, as we desired.

(b) Moreover, if X is symmetric, we want to show that $\xi X \overset{D}{=} \xi|X| \overset{D}{=} X$. The first equation is from (a); as for the second, we see that for any $t \geq 0$,

$$\mathbb{P}(X \geq t) = \mathbb{P}(-X \geq t) = \frac{\mathbb{P}(X \geq t) + \mathbb{P}(-X \geq t)}{2} = \mathbb{P}(\xi X \geq t)$$

from the proof of (a).

(c) It suffices to show that $X - X' \overset{D}{=} X' - X$, but this is trivial since $(X, X') \overset{D}{=} (X', X)$.

⊛

Problem (Exercise 6.4.3). Where in this argument did we use the independence of the random variables X_i? Is mean zero assumption needed for both upper and lower bounds?

Answer. If X_i's are not independent, then $\{\varepsilon_i(X_i - X_i')\}_{i=1}^N$ might not have the same joint distribution as $\{(X_i - X_i')\}_{i=1}^N$. For the mean zero assumption, see Exercise 6.4.4. ⊛

Problem (Exercise 6.4.4). (a) Prove the following generalization of Symmetrization Lemma 6.4.2 for random vectors X_i that do not necessarily have zero means:

$$\mathbb{E}\left[\left\lVert \sum_{i=1}^N X_i - \sum_{i=1}^N \mathbb{E}[X_i] \right\rVert\right] \leq 2\mathbb{E}\left[\left\lVert \sum_{i=1}^N \varepsilon_i X_i \right\rVert\right].$$

(b) Argue that there can not be any non-trivial reverse inequality.

Answer. (a) We see that using Lemma 6.1.2 again, we have

$$\begin{aligned}
\mathbb{E}\left[\left\lVert \sum_{i=1}^N X_i - \sum_{i=1}^N \mathbb{E}[X_i] \right\rVert\right] &= \mathbb{E}\left[\left\lVert \sum_{i=1}^N (X_i - \mathbb{E}[X_i]) \right\rVert\right] \\
&\leq \mathbb{E}\left[\left\lVert \sum_{i=1}^N ((X_i - \mathbb{E}[X_i]) - (X_i' - \mathbb{E}[X_i'])) \right\rVert\right]
\end{aligned}$$

as $\mathbb{E}[X_i] = \mathbb{E}[X_i']$, and using Exercise 6.4.1, we have

$$\begin{aligned} &= \mathbb{E}\left[\left\lVert\sum_{i=1}^{N}(X_i - X_i')\right\rVert\right] \\ &= \mathbb{E}\left[\left\lVert\sum_{i=1}^{N}\varepsilon_i(X_i - X_i')\right\rVert\right] \\ &\leq \mathbb{E}\left[\left\lVert\sum_{i=1}^{N}\varepsilon_i X_i\right\rVert\right] + \mathbb{E}\left[\left\lVert\sum_{i=1}^{N}\varepsilon_i X_i'\right\rVert\right] = 2\mathbb{E}\left[\left\lVert\sum_{i=1}^{N}\varepsilon_i X_i\right\rVert\right]. \end{aligned}$$

(b) Let $N = 1$ and $X_1 = \lambda\mathbb{1}$ for some $\lambda > 0$. Then,

$$\mathbb{E}[\lVert X_1 - \mathbb{E}[X_1]\rVert_2] = 0,$$

while

$$\mathbb{E}[\lVert\varepsilon_1 X_1\rVert_2] = \lambda\lVert\mathbb{1}\rVert_2$$

can be arbitrarily large as $\lambda \to \infty$.

⊛

Problem (Exercise 6.4.5). Prove the following generalization of Symmetrization Lemma 6.4.2. Let $F\colon \mathbb{R}_+ \to \mathbb{R}$ be an increasing, convex function. Show that the same inequalities in Lemma 6.4.2 hold if the norm $\lVert\cdot\rVert$ is replaced with $F(\lVert\cdot\rVert)$, namely

$$\mathbb{E}\left[F\left(\frac{1}{2}\left\lVert\sum_{i=1}^{N}\varepsilon_i X_i\right\rVert\right)\right] \leq \mathbb{E}\left[F\left(\left\lVert\sum_{i=1}^{N}X_i\right\rVert\right)\right] \leq \mathbb{E}\left[F\left(2\left\lVert\sum_{i=1}^{N}\varepsilon_i X_i\right\rVert\right)\right].$$

Answer. We see that for the lower bound, we have

$$\begin{aligned} \mathbb{E}\left[F\left(\frac{1}{2}\left\lVert\sum_{i=1}^{n}\varepsilon_i X_i\right\rVert\right)\right] &= \mathbb{E}\left[F\left(\frac{1}{2}\left\lVert\mathbb{E}_{X'}\left[\sum_{i=1}^{n}\varepsilon_i(X_i - X_i')\right]\right\rVert\right)\right] && (\mathbb{E}_{X_i'}[\varepsilon_i X_i'] = 0) \\ &\leq \mathbb{E}\left[F\left(\mathbb{E}_{X'}\left[\frac{1}{2}\left\lVert\sum_{i=1}^{n}\varepsilon_i(X_i - X_i')\right\rVert\right]\right)\right] \\ & && \text{(Jensen's inequality, } F \text{ increasing)} \\ &\leq \mathbb{E}\left[F\left(\frac{1}{2}\left\lVert\sum_{i=1}^{n}\varepsilon_i(X_i - X_i')\right\rVert\right)\right] && \text{(Jensen's inequality)} \\ &= \mathbb{E}\left[F\left(\frac{1}{2}\left\lVert\sum_{i=1}^{n}(X_i - X_i')\right\rVert\right)\right] && \text{(Exercise 6.4.1 (b) and (c))} \\ &\leq \mathbb{E}\left[F\left(\frac{1}{2}\left\lVert\sum_{i=1}^{n}X_i\right\rVert + \frac{1}{2}\left\lVert\sum_{i=1}^{n}X_i'\right\rVert\right)\right] && (F \text{ increasing}) \\ &\leq \mathbb{E}\left[\frac{1}{2}F\left(\left\lVert\sum_{i=1}^{n}X_i\right\rVert\right) + \frac{1}{2}F\left(\left\lVert\sum_{i=1}^{n}X_i'\right\rVert\right)\right] && (F \text{ convex}) \\ &= \mathbb{E}\left[F\left(\left\lVert\sum_{i=1}^{n}X_i\right\rVert\right)\right]. \end{aligned}$$

On the other hand, for the upper bound, we also have

$$
\begin{aligned}
\mathbb{E}\left[F\left(\left\|\sum_{i=1}^n X_i\right\|\right)\right] &= \mathbb{E}\left[F\left(\left\|\mathbb{E}_{X'}\left[\sum_{i=1}^n (X_i - X_i')\right]\right\|\right)\right] && (\mathbb{E}_{X_i'}[X_i'] = 0)\\
&\leq \mathbb{E}\left[F\left(\mathbb{E}_{X'}\left[\left\|\sum_{i=1}^n (X_i - X_i')\right\|\right]\right)\right] && (\text{Jensen's inequality, } F \text{ increasing})\\
&\leq \mathbb{E}\left[F\left(\left\|\sum_{i=1}^n (X_i - X_i')\right\|\right)\right] && (\text{Jensen's inequality})\\
&= \mathbb{E}\left[F\left(\left\|\sum_{i=1}^n \varepsilon_i (X_i - X_i')\right\|\right)\right] && (\text{Exercise 6.4.1 (b) and (c)})\\
&\leq \mathbb{E}\left[F\left(\left\|\sum_{i=1}^n \varepsilon_i X_i\right\| + \left\|\sum_{i=1}^n \varepsilon_i X_i'\right\|\right)\right] && (F \text{ increasing})\\
&\leq \mathbb{E}\left[\frac{1}{2}F\left(2\left\|\sum_{i=1}^n \varepsilon_i X_i\right\|\right) + \frac{1}{2}F\left(2\left\|\sum_{i=1}^n \varepsilon_i X_i'\right\|\right)\right] && (F \text{ convex})\\
&= \mathbb{E}\left[F\left(2\left\|\sum_{i=1}^n \varepsilon_i X_i\right\|\right)\right].
\end{aligned}
$$

⊛

Problem (Exercise 6.4.6). Let $X_1, \dots, X_N$ be independent, mean zero random variables. Show that their sum $\sum_i X_i$ is sub-gaussian if and only if $\sum_i \varepsilon_i X_i$ is sub-gaussian, and

$$
c\left\|\sum_{i=1}^N \varepsilon_i X_i\right\|_{\psi_2} \leq \left\|\sum_{i=1}^N X_i\right\|_{\psi_2} \leq C\left\|\sum_{i=1}^N \varepsilon X_i\right\|_{\psi_2}.
$$

Answer. Consider $F_K(x) \coloneqq \exp(x^2/K^2) - 1$ for some $K \geq 0$, which is clearly convex. Hence, by Exercise 6.4.5, if $\|\sum_{i=1}^n \varepsilon_i X_i\|_{\psi_2} \leq K$, then

$$
\mathbb{E}\left[F_{2K}\left(\left|\sum_{i=1}^n X_i\right|\right)\right] \leq \mathbb{E}\left[F_{2K}\left(2\left|\sum_{i=1}^n \varepsilon_i X_i\right|\right)\right] = \mathbb{E}\left[F_K\left(\left|\sum_{i=1}^n \varepsilon_i X_i\right|\right)\right] \leq 1,
$$

implying $\|\sum_{i=1}^n X_i\|_{\psi_2} \leq 2K$. Conversely, if $\|\sum_{i=1}^n X_i\|_{\psi_2} \leq K$, then

$$
\mathbb{E}\left[F_{2K}\left(\left|\sum_{i=1}^n \varepsilon_i X_i\right|\right)\right] = \mathbb{E}\left[F_K\left(\frac{1}{2}\left|\sum_{i=1}^n \varepsilon_i X_i\right|\right)\right] \leq \mathbb{E}\left[F_K\left(\left|\sum_{i=1}^n X_i\right|\right)\right] \leq 1,
$$

thus $\|\sum_{i=1}^n \varepsilon_i X_i\|_{\psi_2} \leq 2K$. ⊛

Week 21: Random Matrices with Non-I.I.D. Entries

6.5 Random matrices with non-i.i.d. entries

20 Jul. 2024

6.6 Application: matrix completion

Week 22: Contraction Trick

6.7 Contraction Principle

25 Jul. 2024

Problem (Exercise 6.7.2). Check that the function f defined in (6/16) is convex. For reference, $f\colon \mathbb{R}^N \to \mathbb{R}$ is defined as

$$f(a) \coloneqq \mathbb{E}\left[\left\|\sum_{i=1}^N a_i\varepsilon_i x_i\right\|\right].$$

Answer. To prove that for $f\colon \mathbb{R}^N \to \mathbb{R}$ where

$$f(a) = \mathbb{E}\left[\left\|\sum_{i=1}^N a_i\varepsilon_i x_i\right\|\right]$$

is convex, consider $a, b \in \mathbb{R}^N$ and some $\lambda \in (0,1)$, we have

$$\begin{aligned} f(\lambda a + (1-\lambda)b) &= \mathbb{E}\left[\left\|\sum_{i=1}^N \left[\lambda a_i + (1-\lambda)b_i\right]\varepsilon_i x_i\right\|\right] \\ &\leq \mathbb{E}\left[\lambda\left\|\sum_{i=1}^N a_i\varepsilon_i x_i\right\| + (1-\lambda)\left\|\sum_{i=1}^N b_i\varepsilon_i x_i\right\|\right] = \lambda f(a) + (1-\lambda) f(b), \end{aligned}$$

implying that f is convex. ⊛

Problem (Exercise 6.7.3). Prove the following generalization of Theorem 6.7.1. Let $X_1, \ldots, X_N$ be independent, mean zero random vectors in a normed space, and let $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. Then

$$\mathbb{E}\left[\left\|\sum_{i=1}^N a_i X_i\right\|\right] \leq 4\|a\|_\infty \cdot \mathbb{E}\left[\left\|\sum_{i=1}^N X_i\right\|\right].$$

Answer. Let ε_i's be independent Bernoulli's random variables, then from the symmetrization and Theorem 6.7.1 with conditioning on X_i's, we have

$$\mathbb{E}\left[\left\|\sum_{i=1}^N a_i X_i\right\|\right] \leq 2\mathbb{E}\left[\left\|\sum_{i=1}^N a_i\varepsilon_i X_i\right\|\right] \leq 2\|a\|_\infty \cdot \mathbb{E}\left[\left\|\sum_{i=1}^N \varepsilon_i X_i\right\|\right] \leq 4\|a\|_\infty \cdot \mathbb{E}\left[\left\|\sum_{i=1}^N X_i\right\|\right],$$

where the last inequality follows again from the symmetrization. ⊛

Problem (Exercise 6.7.5). Show that the factor $\sqrt{\log N}$ in Lemma 6.7.4 is needed in general, and is optimal. Thus, symmetrization with Gaussian random variables is generally weaker than symmetrization with symmetric Bernoullis.

Answer. Consider e_i's being i^{th} standard basis in $\mathbb{R}^N$, and consider $X_i \coloneqq \varepsilon_i e_i$ for all $i \geq 1$. We have

$$\mathbb{E}\left[\left\|\sum_{i=1}^N X_i\right\|_\infty\right] = \mathbb{E}\left[\left\|\sum_{i=1}^N \varepsilon_i e_i\right\|_\infty\right] = \mathbb{E}[\|(\varepsilon_1, \ldots, \varepsilon_N)\|_\infty] = 1,$$

while given $g_i \sim \mathcal{N}(0,1)$, we have

$$\mathbb{E}\left[\left\|\sum_{i=1}^N g_i X_i\right\|_\infty\right] = \mathbb{E}\left[\left\|\sum_{i=1}^N g_i\varepsilon_i e_i\right\|_\infty\right] = \mathbb{E}[\|(g_1, \ldots, g_N)\|_\infty] \asymp \sqrt{\log N}$$

due to symmetry of g_i's and Exercise 2.5.10 and 2.5.11. ⊛

Problem (Exercise 6.7.6). Let $F\colon \mathbb{R}_+ \to \mathbb{R}$ be a convex increasing function. Generalize the symmetrization and contraction results of this and previous section by replacing the norm $\|\cdot\|$ with $F(\|\cdot\|)$ throughout.

Answer. Omit. ⊛

Problem (Exercise 6.7.7). Consider a bounded subset $T \subseteq \mathbb{R}^n$, and let $\varepsilon_1, \ldots, \varepsilon_n$ be independent symmetric Bernoulli random variables. Let $\phi_i \colon \mathbb{R} \to \mathbb{R}$ be contractions, i.e., Lipschitz functions with $\|\phi_i\|_{\mathrm{Lip}} \leq 1$. Then

$$\mathbb{E}\left[\sup_{t \in T} \sum_{i=1}^{n} \varepsilon_i \phi_i(t_i)\right] \leq \mathbb{E}\left[\sup_{t \in T} \sum_{i=1}^{n} \varepsilon_i t_i\right].$$

To prove this result, do the following steps:

(a) First let $n = 2$. Consider a subset $T \subseteq \mathbb{R}^2$ and contraction $\phi \colon \mathbb{R} \to \mathbb{R}$, and check that

$$\sup_{t \in T}(t_1 + \phi(t_2)) + \sup_{t \in T}(t_1 - \phi(t_2)) \leq \sup_{t \in T}(t_1 + t_2) + \sup_{t \in T}(t_1 - t_2).$$

(b) Use induction on n complete proof.

Answer. (a) Writing t by t' in the second term on both sides, which gives

$$\begin{aligned}
\sup_{t \in T}(t_1 + \phi(t_2)) + \sup_{t' \in T}(t_1' - \phi(t_2')) &= \sup_{t, t' \in T}\left(t_1 + \phi(t_2) + t_1' - \phi(t_2')\right) \\
&\leq \sup_{t, t' \in T}\left(t_1 + t_1' + |t_2 - t_2'|\right) \\
&= \sup_{t, t' \in T}\left(t_1 + t_1' + t_2 - t_2'\right) = \sup_{t \in T}(t_1 + t_2) + \sup_{t' \in T}(t_1' - t_2'),
\end{aligned}$$

where we use symmetry strategically.

(b) Firstly, we observe that conditioning on $\varepsilon_1, \ldots, \varepsilon_{n-1}$ gives

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t \in T} \sum_{i=1}^{n-1} \varepsilon_i \phi_i(t_i) + \varepsilon_n \phi_n(t_n)\right] \\
&= \frac{1}{2}\left(\sup_{t \in T} \sum_{i=1}^{n-1} \varepsilon_i \phi_i(t_i) + \phi_n(t_n) + \sup_{t \in T} \sum_{i=1}^{n-1} \varepsilon_i \phi_i(t_i) - \phi_n(t_n)\right) \\
&\leq \frac{1}{2}\left(\sup_{t \in T} \sum_{i=1}^{n-1} \varepsilon_i \phi_i(t_i) + t_n + \sup_{t \in T} \sum_{i=1}^{n-1} \varepsilon_i \phi_i(t_i) - t_n\right) = \mathbb{E}\left[\sup_{t \in T} \sum_{i=1}^{n-1} \varepsilon_i \phi_i(t_i) + \varepsilon_n t_n\right],
\end{aligned}$$

where the inequality comes from (a) by considering the supremum over

$$T^{(n)} := \left\{(x, y) \in \mathbb{R}^2 \colon x = \sum_{i=1}^{n-1} \varepsilon_i \phi_i(t_i), y = t_n, (t_1, \ldots, t_{n-1}, t_n) \in T\right\}.$$

Explicitly, we get

$$\mathbb{E}\left[\mathbb{E}\left[\sup_{t \in T} \sum_{i=1}^{n-1} \varepsilon_i \phi_i(t_i) + \varepsilon_n \phi_n(t_n)\right] \mid \varepsilon_{1\colon n-1}\right] \leq \mathbb{E}\left[\mathbb{E}\left[\sup_{t \in T} \sum_{i=1}^{n-1} \varepsilon_i \phi_i(t_i) + \varepsilon_n t_n\right] \mid \varepsilon_{1\colon n-1}\right].$$

By iterating this with conditioning on $\varepsilon_{1\colon k}$ for every k and apply (a) on

$$T^{(k)} := \left\{(x, y) \in \mathbb{R}^2 \colon x = \sum_{i=1}^{k-1} \varepsilon_i \phi_i(t_i) + \sum_{i=k+1}^{n} \varepsilon_i t_i, y = t_k, (t_1, \ldots, t_{n-1}, t_n) \in T\right\},$$

we get the desired result.

⊛

Problem (Exercise 6.7.8)**.** Generalize Talagrand's contraction principle for arbitrary Lipschitz functions $\phi_i \colon \mathbb{R} \to \mathbb{R}$ without restriction on their Lipschitz norms.

Answer. Look into the proof of Exercise 6.7.7, we see that for general Lipschitz functions ϕ_i's,

$$\mathbb{E}\left[\sup_{t\in T}\sum_{i=1}^{n}\varepsilon_i\phi_i(t_i)\right] \leq \mathbb{E}\left[\sup_{t\in T}\sum_{i=1}^{n}\varepsilon_i\|\phi_i\|_{\mathrm{Lip}}t_i\right] \leq \max_{1\leq i\leq n}\|\phi_i\|_{\mathrm{Lip}}\mathbb{E}\left[\sup_{t\in T}\sum_{i=1}^{n}\varepsilon_i t_i\right],$$

where the last inequality follows from Theorem 6.7.1, by noting that $\sup_{t\in T}$ satisfies all the conditions we need in Theorem 6.7.1. ⊛

Bibliography

[BSP09] K. M. Briggs, L. Song, and T. Prellberg. *A note on the distribution of the maximum of a set of Poisson random variables*. 2009. arXiv: 0903.4373 [math.PR]. URL: https://arxiv.org/abs/0903.4373.

[Kim83] AC Kimber. "A note on Poisson maxima". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 63.4 (1983), pp. 551–552.

[Ver24] Roman Vershynin. *High-Dimensional Probability*. Vol. 47. Cambridge University Press, 2024. URL: https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.html.

[Vla+20] Mariia Vladimirova et al. "Sub-Weibull distributions: Generalizing sub-Gaussian and sub-Exponential properties to heavier tailed distributions". In: *Stat* 9.1 (Jan. 2020). ISSN: 2049-1573. DOI: 10.1002/sta4.318. URL: http://dx.doi.org/10.1002/sta4.318.