Solutions to Selected Exercises in
High-Dimensional Probability

Yuze Han Liangyu Zhang

January 6, 2021

1 Appetizer and preliminaries on random variable

Exercise 0.0.5

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

Exercise 0.0.6 Let $k := \lceil 1/\epsilon^2 \rceil$. The number of ways to choose k elements from an N-element set with repetitions is $\binom{N+k-1}{k}$. Exercise 2.2.5 implies

$$\binom{N+k-1}{k} \leq \left(\frac{e(N+k-1)}{k}\right)^k \leq \left(e + eN\epsilon^2\right)^{\lceil 1/\epsilon^2 \rceil}.$$

Exercise 1.3.3 Let $\sigma^2 = \mathrm{Var}\,(X_1)$.

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

Note that for a sequence of r.v.s $\{X_n\}$, $X_n \xrightarrow{d} X$ does not necessarily imply $\mathbb{E}\,X_n \to \mathbb{E}\,X$. And a sufficient condition for $\mathbb{E}\,X_n \to \mathbb{E}\,X$ is uniform integrability.

2 Concentration of sums of independent random variables

Exercise 2.1.4

$$\begin{aligned}\mathbb{E}\, g^2 \mathbf{1}_{\{g>t\}} &= \int_t^\infty \frac{x^2}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_t^\infty \frac{x}{\sqrt{2\pi}} d\left(-e^{-x^2/2}\right) = t \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} + \mathbb{P}\ \{g > t\} \\ &\leq \left(t + \frac{1}{t}\right) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad \text{(by Proposition 2.1.2)}\end{aligned}$$

Exercise 2.2.7 We first prove the following lemma.

Lemma. *If $\mathbb{E}\, X = 0$ and $a \leq X \leq b$, we have $\mathbb{E}\, e^{\lambda X} \leq e^{\lambda^2 (b-a)^2/8}$.*

Proof. The convexity of e^x implies

$$\mathbb{E}\, e^{\lambda X} \leq \mathbb{E} \left(\frac{b - X}{b - a} e^{\lambda a} + \frac{x - a}{b - a} e^{\lambda b} \right) = e^{\phi(\lambda)},$$

where $\phi(x) = \ln \left[\left(\frac{b}{b-a} e^{ax} + \frac{-a}{b-a} e^{bx} \right) \right]$. One can prove that $\phi(0) = 0$, $\phi'(0) = 0$ and $\phi''(x) \leq \frac{(b-a)^2}{4}$. Taylor expansion gives the desired result. An alternative proof can be found in [1, Lemma 2.2] □

For an arbitrary $\lambda > 0$,

$$\begin{aligned}\mathbb{P}\ \left\{ \sum_{i=1}^N (X_i - \mathbb{E}\, X_i) \geq t \right\} &\leq \exp(-\lambda t) \prod_{i=1}^N \mathbb{E}\ \exp(\lambda (X_i - \mathbb{E}\, X_i)) \\ &\leq \exp(-\lambda t) \prod_{i=1}^N \exp(\lambda^2 (M_i - m_i)^2 / 8) \\ &= \exp \left(\frac{1}{8} \sum_{i=1}^N (M_i - m_i)^2 \lambda^2 - t\lambda \right)\end{aligned}$$

The minimum is attained for $\lambda = \frac{4t}{\sum_{i=1}^N (M_i - m_i)^2}$. This complete the proof of Hoeffding's inequality.

Exercise 2.2.8 Let $X_i = \mathbf{1}_{\{\text{The algorithm returns the wrong answer at the i-th time}\}}$. Then $X_i \sim B\left(1, \frac{1}{2} - \delta\right)$. As long as $N \geq \frac{1}{2\delta^2} \ln(1/\epsilon)$, Hoeffding's inequality implies

$$\mathbb{P}\ \left\{ \sum_{i=1}^N X_i > \frac{1}{2} \right\} = \mathbb{P}\ \left\{ \sum_{i=1}^N (X_i - \mathbb{E}\, X_i) > \delta N \right\} \leq \exp\left(-2N\delta^2\right) \leq \epsilon.$$

Exercise 2.2.9

(a) Let $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i$. For $N = 4\sigma^2/\epsilon^2$,

$$\mathbb{P}\ \{|\hat{\mu} - \mu| \geq \epsilon\} \leq \frac{\mathrm{Var}\,(\hat{\mu})}{\epsilon^2} = \frac{\sigma^2}{N\epsilon^2} = \frac{1}{4}$$

(b) We repeat part (a) K times and obtain K independent sapmle means $\hat{\mu}_1, \hat{\mu}_2, ..., \hat{\mu}_K$ with $K = 8\ln(1/\delta)$. Denote their median by $\tilde{\mu}$. Let $Y_i = \mathbf{1}_{|\hat{\mu}_1 - \mu| \geq \epsilon}$. Then $Y_i \sim B(1, p_i)$ with $p_i \leq 1/4$. Hoeffding's inequality implies

$$\mathbb{P}\ \{|\tilde{\mu} - \mu| \geq \epsilon\} \leq \mathbb{P}\ \left\{\sum_{i=1}^{K} Y_i \geq \frac{K}{2}\right\} \leq \mathbb{P}\ \left\{\sum_{i=1}^{K}(Y_i - \mathbb{E}\,Y_i) \geq \frac{K}{4}\right\} \leq \exp\left(-\frac{K}{8}\right) \leq \delta.$$

Exercise 2.3.2 For an arbitrary $\lambda > 0$,

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

Substituting $\lambda = \ln(\mu/t)$ gives the desired result.

Exercise 2.3.5 By Theorem 2.3.1 and Exercise 2.3.2, we have

$$\begin{aligned}\mathbb{P}\ \{|S_N - \mu| \geq \delta\mu\} &\leq e^{-\mu}\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} + e^{-\mu}\left(\frac{e}{1-\delta}\right)^{(1-\delta)\mu} \\ &= \exp\left[\delta\mu - (1+\delta)\mu\ln(1+\delta)\right] + \exp\left[-\delta\mu - (1-\delta)\mu\ln(1-\delta)\right].\end{aligned}$$

It suffices to prove

$$\delta - (1+\delta)\ln(1+\delta) \leq -c\delta^2, \quad \delta \in (-1, 1)$$

for some $c > 0$. We can choose $c \leq 2\ln 2 - 1$.

Exercise 2.3.8 The ch.f. of $\frac{X-\lambda}{\lambda}$ converges to $\exp(-t^2/2)$ as $\lambda \to \infty$. Note that $\forall\theta > 0, \lim_{n\to\infty} f(n\theta) = a$ does not imply $\lim_{x\to\infty} f(x) = a$. See [3, Example 3.4.8] for an alternative proof.

Exercise 2.4.2 Suppose $d \leq C \log n$. By Theorem 2.3.1

$$\begin{aligned}\mathbb{P}\left\{\exists i \leq n : d_i \geq KC \log n\right\} &\leq \sum_{i=1}^{n} \mathbb{P}\left\{d_i \geq KC \log n\right\} \leq n e^{-d} \left(\frac{ed}{KC \log n}\right)^{KC \log n} \\ &\leq n e^{KC \log n} K^{-KC \log n} \leq e^{(KC+1-KC\log K)\log n} \leq 0.1\end{aligned}$$

for a sufficient large constant K.

Exercise 2.4.3

$$\begin{aligned}\mathbb{P}\left\{\exists i \leq n : d_i \geq K \frac{\log n}{\log \log n}\right\} &\leq n e^{-d} \left(\frac{ed \log \log n}{K \log n}\right)^{K \log n / \log \log n} \\ &\leq e^{-(K-1)\log n + o(\log n)} \leq 0.1\end{aligned}$$

for a sufficient large constant K.

Exercise 2.4.4 Take $V' \subset V$ randomly of size $\tilde{n} = n^{1/3}$ with vertex indices $I = \{i_1, i_2, \ldots, i_{\tilde{n}}\}$. Let $\tilde{d}_j$ denote the degree of the vertex $j \in I$ of V' and $\tilde{d}$ denote the expected degree. Then $\tilde{d} = o\left(\frac{\log n}{n^{2/3}}\right)$. It follows that $\log \tilde{d} / \log n \leq -2/3 + o(1)$. We have

$$\mathbb{P}\left\{\exists i \leq \tilde{n} : \tilde{d}_i \geq 1\right\} \leq \tilde{n} e^{-\tilde{d}} e \tilde{d} \leq e^{\log \tilde{n} + \log \tilde{d} + 1} \leq 1 - \sqrt{0.9}$$

when n is sufficiently large. Denote $A = \{\text{There are no edges between vertices in } V'\}$. Then $\mathbb{P}\{A\} \geq \sqrt{0.9}$. Condition on A, $\{d_j, j \in I\}$ are independent. By Poisson approximation (for the accuracy of Poisson approximation, see [6]), we obtain that for any $j \in I$,

$$\begin{aligned}\mathbb{P}\left\{d_j = 10d \,|\, A\right\} &= \frac{1}{\sqrt{2\pi 10 d}} e^{-d} \left(\frac{ed}{10d}\right)^{10d} + O\left(\frac{1}{n}\right) \\ &= \frac{1}{\sqrt{20\pi}} \exp\left[-\left(10 \log 10 + \frac{1}{2} - 9\right) d\right] + O\left(\frac{1}{n}\right) \\ &\geq a n^{-c}\end{aligned}$$

for some constants $a > 0$ and $0 < c < 1$ and sufficiently large n. Since $d = o(\log n)$, we may assume $c < 1/3$. It follows that

$$\begin{aligned}
\mathbb{P}\left\{\exists i : d_i = 10d\right\} &\geq \mathbb{P}\left\{\exists j \in I : d_j = 10d\right\} \\
&\geq \mathbb{P}\left\{\exists j \in I : d_j = 10d \,|A\right\} \mathbb{P}\left\{A\right\} \\
&= \left[1 - (1 - \mathbb{P}\left\{d_i = 10d\right\})^{\tilde{n}}\right] \mathbb{P}\left\{A\right\} \\
&\geq \left[1 - \exp\left(-an^{1/3-c}\right)\right] \mathbb{P}\left\{A\right\} \\
&\geq 0.9
\end{aligned}$$

when n is sufficiently large.

Exercise 2.4.5 Take $V' \subset V$ randomly of size $\tilde{n} = n^{1/3}$ with vertex indices $I = \{i_1, i_2, \ldots, i_{\tilde{n}}\}$. Let $\tilde{d}_j$ denote the degree of the vertex $j \in I$ of V' and $\tilde{d}$ denote the expected degree. Then $\tilde{d} = O\left(n^{-2/3}\right)$. We have

$$\mathbb{P}\left\{\exists i \leq \tilde{n} : \tilde{d}_i \geq 1\right\} \leq \tilde{n}e^{-\tilde{d}}e\tilde{d} \leq e^{\log \tilde{n} + \log \tilde{d} + 1} \leq 1 - \sqrt{0.9}$$

when n is sufficiently large. Denote $A = \{\text{There are no edges between vertices in } V'\}$. Then $\mathbb{P}\left\{A\right\} \geq \sqrt{0.9}$. Condition on A, $\{d_j, j \in I\}$ are independent. Let $k = \frac{b \log n}{\log \log n}$ with $b < 1/4$

and $n' = n - n^{1/3} = \Theta(n)$. Assume $1 \le d \le C$. Then for any $j \in I$, we have

$$\begin{aligned}
\mathbb{P}\{d_i = k \,|A\} &= \binom{n'-1}{k}\left(\frac{d}{n-1}\right)^k\left(1-\frac{d}{n-1}\right)^{n'-k-1} \\
&\ge \frac{(n'-k)^k}{k!}\left(\frac{1}{n-1}\right)^k\left(1-\frac{C}{n-1}\right)^{n'-k-1} \\
&= \frac{1}{k!}\left(1-\frac{k-1+n^{1/3}}{n-1}\right)^k e^{-C}(1+o(1)) \\
&= \frac{e^k}{k^k\sqrt{2\pi k}}e^{-C}(1+o(1)) \\
&= \exp\left(k - k\log k - \frac{1}{2}\log(2\pi k) - C\right)(1+o(1)) \\
&= \exp(-b\log n(1+o(1))) \\
&\ge n^{-1.1b}
\end{aligned}$$

when n is sufficiently large. Note that we can also use Poisson approximation to obtain the same inequality. It follows that

$$\begin{aligned}
\mathbb{P}\{\exists i : d_i = k\} &\ge \mathbb{P}\{\exists j \in I : d_j = k\} \\
&\ge \mathbb{P}\{\exists j \in I : d_j = k \,|A\}\,\mathbb{P}\{A\} \\
&= \left[1 - (1-\mathbb{P}\{d_i = k\})^{\tilde{n}}\right]\mathbb{P}\{A\} \\
&\ge \left[1 - \exp\left(-n^{1/3-1.1b}\right)\right]\mathbb{P}\{A\} \\
&\ge 0.9
\end{aligned}$$

when n is sufficiently large.

Exercise 2.5.1

$$\begin{aligned}
\mathbb{E}\,|X|^p &= 2\int_0^\infty x^p \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \\
&= \frac{2^{p/2}}{\sqrt{\pi}}\int_0^\infty t^{\frac{p-1}{2}}e^{-t}dt \quad \text{(by change of variables } t = x^2\text{)} \\
&= 2^{p/2}\frac{\Gamma((1+p)/2)}{\Gamma(1/2)}.
\end{aligned}$$

(2.11) follows from Stirling's formula $\Gamma(x+1) \sim \sqrt{2\pi x}\left(\frac{x}{e}\right)^x$.

Exercise 2.5.4 Assume $K_4 = 1$. By Jensen's inequality,

$$\lambda \mathbb{E}\, X \leq \ln \mathbb{E}\, e^{\lambda X} \leq \lambda^2.$$

Since λ is arbitrary, $\mathbb{E}\, X = 0$.

Exercise 2.5.5

(a) For $\lambda^2 \geq 1/2$, $\mathbb{E}\,\exp(\lambda^2 X^2) = \infty$.

(b) For $M > \sqrt{K}$, we have

$$\mathbb{P}\,\left(X > M\right) \leq \exp(-\lambda^2 M^2)\,\mathbb{E}\,\exp(\lambda^2 X^2) \leq \exp((K - M^2)\lambda^2).$$

Letting $\lambda \to \infty$, $\mathbb{P}\,\left(X > M\right) = 0$.

Exercise 2.5.7 We only verify positive definiteness and the triangular inequality.

positive definiteness If $\|X\|_{\psi_2} = 0$, $\mathbb{E}\,\exp\left(X^2/t^2\right) \leq 2$ for all $t > 0$. Then for all $c > 0$, $\mathbb{P}\left(X > c\right) \leq e^{-c^2/t^2}\mathbb{E}\, e^{X^2/t^2} \leq 2e^{-c^2/t^2} \to 0$ as $t \to 0$. Thus $X = 0$.

triangular inequality Let $a = \|X\|_{\psi_2}$, $b = \|Y\|_{\psi_2}$. Since e^{x^2} is convex, we have

$$\mathbb{E}\,\exp\left(\frac{X+Y}{a+b}\right)^2 \leq \frac{a}{a+b}\mathbb{E}\,\exp\left(\frac{X}{a}\right)^2 + \frac{b}{a+b}\mathbb{E}\,\exp\left(\frac{Y}{b}\right)^2 \leq 2.$$

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

Exercise 2.5.9

Poisson If $X \sim Pois(\lambda)$, we have $\mathbb{P}\left(X \geq t\right) \geq \frac{\lambda^{\lceil t\rceil}e^{-\lambda}}{\lceil t\rceil!} = \Omega(e^{-t\ln t})$.

Exponential If $X \sim Exp(\lambda)$, we have $\mathbb{P}\left(X \geq t\right) = e^{-\lambda t}$.

Pareto If $X \sim Pa(a, \theta)$, we have $\mathbb{P}(X \geq t) = (a/t)^\theta = \Omega(e^{-\theta \ln t})$.

Cauchy The expectation of Cauchy distribution dose not exist.

Exercise 2.5.10 From the proof of Proposition 2.5.2, we can assume $c = 1$. Let $Y_i = \frac{|X_i|}{\sqrt{1+\log i}}$. For $t \geq \sqrt{2}K$,

$$\begin{aligned}
\mathbb{P}\left(\max_i Y_i > t\right) &\leq \sum_i \mathbb{P}\left(|X_i| > t\sqrt{1+\log i}\right) \\
&\leq \sum_i 2\exp\left(-\frac{t^2(1+\log i)}{K^2}\right) \\
&= \sum_i 2\exp\left(-\frac{t^2}{K^2}\right) i^{-t^2/K^2} \\
&\leq C_1 \exp\left(-\frac{t^2}{K^2}\right)
\end{aligned}$$

Then we have

$$\begin{aligned}
\mathbb{E}\max_i Y_i &= \int_0^\infty \mathbb{P}\left(\max_i Y_i > t\right) dt \\
&= \sqrt{2}K + \int_{\sqrt{2}K}^\infty \mathbb{P}\left(\max_i Y_i > t\right) dt \\
&\leq \sqrt{2}K + C_1 \int_0^\infty \exp\left(-\frac{t^2}{K^2}\right) \\
&\leq C_2 K,
\end{aligned}$$

where $C_1, C_2 > 0$ do not depend on any parameter. Finally, for every $N \geq 2$,

$$\mathbb{E}\max_{i \leq N} |X_i| \leq \sqrt{1+\log N}\,\mathbb{E}\max_i Y_i \leq C_3 K\sqrt{\log N}.$$

Exercise 2.5.11 This proof is inspired by [7]. For a tight bound, see [4].

For $N \geq 3$, let $A_N = \{\max_{i \leq N} X_i \geq C_N \sqrt{\log N}\}$ with $C_N = \sqrt{2 - \frac{\log \log N}{\log N}} > 1$. First, we have

$$\begin{aligned}
\mathbb{E} \max_{i \leq N} X_i &= \mathbb{E} \left[\max_{i \leq N} X_i \middle| A_N \right] \mathbb{P}(A_N) + \mathbb{E} \left[\max_{i \leq N} X_i \middle| A_N^c \right] \mathbb{P}(A_N^c) \\
&\geq C_N \sqrt{\log N}\, \mathbb{P}(A_N) + \mathbb{E} \left[X_1 | A_N^c \right] \mathbb{P}(A_N^c) \\
&= C_N \sqrt{\log N}\, \mathbb{P}(A_N) + \mathbb{E} \left[X_1 | X_1 < C_N \sqrt{\log N} \right] (1 - \mathbb{P}(A_N)) \\
&\geq C_N \sqrt{\log N}\, \mathbb{P}(A_N) + \mathbb{E} \left[X_1 | X_1 < 0 \right] (1 - \mathbb{P}(A_N)) \\
&\geq \left(\sqrt{\log N} - \sqrt{\frac{2}{\pi}} \right) \mathbb{P}(A_N) - \sqrt{\frac{2}{\pi}}.
\end{aligned}$$

Now we bound $\mathbb{P}(A_N)$.

$$\begin{aligned}
\mathbb{P}(A_N) &= 1 - \left[1 - \mathbb{P}(X_1 \geq C_N \sqrt{\log N}) \right]^N \\
&\geq 1 - \left[1 - \frac{1}{C_N \sqrt{2\pi \log N}} \left(1 - \frac{1}{C_N^2 \log N} \right) N^{-C_N^2/2} \right]^N \\
&\geq 1 - \exp \left[- \frac{1}{C_N \sqrt{2\pi \log N}} \left(1 - \frac{1}{C_N^2 \log N} \right) N^{1 - C_N^2/2} \right] \\
&= 1 - \exp \left[- \frac{1}{C_N \sqrt{2\pi}} \left(1 - \frac{1}{C_N^2 \log N} \right) \right] \\
&= 1 - \exp \left[\frac{1}{C_N \sqrt{2\pi}} \left(\frac{1}{2 \log N - \log \log N} - 1 \right) \right] \\
&\geq 1 - \exp \left[\frac{1}{\sqrt{2\pi}} \left(\frac{1}{\log N} - 1 \right) \right].
\end{aligned}$$

For $N \geq 8$, $\mathbb{P}(A_N) \geq 1 - \exp(-1/2\sqrt{2\pi}) \triangleq C$. When $N \geq \exp\left(\frac{8(C+1)^2}{\pi C^2} \right)$, we obtain

$$\mathbb{E} \max_{i \leq N} X_i \geq C \sqrt{\log N} - \sqrt{\frac{2}{\pi}} (C+1) \geq \frac{C}{2} \sqrt{\log N}.$$

Since $\mathbb{E}\, X_1 = 0$ and $\mathbb{E} \max_{i \leq N} X_i > 0$ for $N \geq 2$, there exists $c > 0$ such that $\mathbb{E} \max_{i \leq N} X_i \geq c\sqrt{\log N}$ for all $N \geq 1$.

Exercise 2.6.5 The lower bound follows from

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

The upper bound follows from

$$\left\|\sum_{i=1}^N a_i X_i\right\|_{L^p} \leq C_1\sqrt{p}\left\|\sum_{i=1}^N a_i X_i\right\|_{\psi_2} \leq C_2\sqrt{p}\sqrt{\sum_{i=1}^N \|a_i X_i\|_{\psi_2}^2} \leq CK\sqrt{p}\left(\sum_{i=1}^N a_i^2\right)^{1/2}.$$

Exercise 2.6.7 The upper bound follows from

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

To prove the lower bound, we first note that.

$$\|Z\|_{L^2}^2 = \mathbb{E}\,|Z|^{p/2}|Z|^{2-p/2} \leq \||Z|^{p/2}\|_{L^2}\||Z|^{2-p/2}\|_{L^2} = \|Z\|_{L^p}^{p/2}\|Z\|_{L^{4-p}}^{2-p/2}.$$

It follows that

$$\left\|\sum_{i=1}^N a_i X_i\right\|_{L^p} \geq \frac{\left\|\sum_{i=1}^N a_i X_i\right\|_{L^2}^{4/p}}{\left\|\sum_{i=1}^N a_i X_i\right\|_{L^{4-p}}^{4/p-1}} \geq \frac{\left(\sum_{i=1}^N a_i^2\right)^{2/p}}{\left[CK\sqrt{4-p}\left(\sum_{i=1}^N a_i^2\right)^{1/2}\right]^{4/p-1}} = c(K)\left(\sum_{i=1}^N a_i^2\right)^{1/2}$$

with $c(K) = \left(CK\sqrt{4-p}\right)^{1-4/p}$.

Exercise 2.6.9 Let X be $\sqrt{\log 2}$ with probability $1/4$ and $-\sqrt{\log 2}$ with probability $3/4$. We have $\|X\|_{\psi_2} = 1$ but $\mathbb{E}\exp(X - \mathbb{E}\,X)^2 > 2$.

Exercise 2.7.2

a $\Rightarrow$ b Assume $K_1 = 1$. We have

$$\mathbb{E}\,|X|^p = \int_0^\infty p t^{p-1}\mathbb{P}\left(|X| \geq t\right)dt = 2p\int_0^\infty t^{p-1}e^{-t}dt = 2p\Gamma(p) \leq 2p^p.$$

b $\Rightarrow$ c Assume $K_2 = 1$. For $0 < \lambda < 1/2e$,

$$\mathbb{E}\,\exp(\lambda|X|) = \sum_{p=0}^{\infty} \frac{\lambda^p \mathbb{E}\,|X|^p}{p!} \leq \sum_{p=0}^{\infty} \frac{\lambda^p \mathbb{E}\,|X|^p e^p}{p^p} \leq \sum_{p=0}^{\infty} \lambda^p e^p = \frac{1}{1-\lambda e} \leq \exp(2e\lambda).$$

c $\Rightarrow$ d is trivial.

d $\Rightarrow$ a Assume $K_4 = 1$. $\mathbb{P}\,(|X| \geq t) \leq e^{-t}\mathbb{E}\,e^{|X|} \leq 2e^{-t}$.

Exercise 2.7.3 We first state the proposition.

Proposition. *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

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

(b) The moments of X satisfy

$$\|X\|_{L^p} = (\mathbb{E}\,|X|^p)^{1/p} \leq K_2 \left(\frac{p}{\alpha}\right)^{1/\alpha} \quad \text{for all } p \geq 1.$$

(c) The MGF of $|X|^\alpha$ satisfies

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

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

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

Moreover, if $\mathbb{E}\,X = 0$ and $\alpha \geq 2$, then properties a-d are also equivalent to the following one.

(e) The MGF of X satisfies

$$\mathbb{E}\,\exp(\lambda X) \leq \exp\left(K_5^{\alpha/(\alpha-1)}|\lambda|^{\alpha/(\alpha-1)}\right) \quad \text{for all } \lambda \in \mathbb{R}.$$

If $1 < \alpha < 2$, then properties a-d are also equivalent to the following one.

(f) The MGF of X satisfies

$$\mathbb{E}\,\exp(\lambda X) \leq \exp\left(K_6^{\alpha/(\alpha-1)}|\lambda|^{\alpha/(\alpha-1)}\right) \quad \textit{for all } |\lambda| \geq 1/2e.$$

Proof.

a $\Rightarrow$ b. Assume $K_1 = 1$. Note that p is not necessarily an integer. For all $p \geq 1$,

$$\mathbb{E}\,|X|^p = \int_0^\infty pt^{p-1}\mathbb{P}\left(|X| \geq t\right)dt = 2p\int_0^\infty t^{p-1}e^{-t^\alpha}dt = \frac{2p}{\alpha}\int_0^\infty u^{p/\alpha-1}e^{-u}du = 2\Gamma\left(\frac{p}{\alpha}+1\right).$$

There exists $K_2 > 0$ such that $\mathbb{E}\,|X|^p = 2\Gamma\left(\frac{p}{\alpha}+1\right) \leq K_2\left(\frac{p}{\alpha}\right)^{p/\alpha}$.

b $\Rightarrow$ c. Assume $K_2 = 1$. For $0 < \lambda < 1/2e^{1/\alpha}$,

$$\mathbb{E}\,\exp(\lambda^\alpha|X|^\alpha) = \sum_{p=0}^\infty \frac{\lambda^{\alpha p}\mathbb{E}\,|X|^{\alpha p}}{p!} \leq \sum_{p=0}^\infty \frac{\lambda^{\alpha p}p^p e^p}{p^p} \leq \sum_{p=0}^\infty \lambda^{\alpha p}e^p = \frac{1}{1-2e\lambda^\alpha} \leq \exp(2e\lambda^\alpha).$$

c $\Rightarrow$ d is trivial.

d $\Rightarrow$ a. Assume $K_4 = 1$. $\mathbb{P}\left(|X| \geq t\right) \leq e^{-t^\alpha}\,\mathbb{E}\,e^{|X|^\alpha} \leq 2e^{-t^\alpha}$.

b c $\Rightarrow$ e. Assume $K_2, K_3 \leq 1$. Let $\beta = \alpha/(\alpha-1)$. We have $1 < \beta < \alpha$. We have the following inequality

$$\exp(x) \leq x + \exp\left(|x|^\beta\right) \quad \text{for all } x \in [-1, 1].$$

It follows that

$$
\begin{aligned}
\mathbb{E}\,\exp(\lambda X) &\leq \mathbb{E}\,\exp\left(|\lambda|^\beta |x|^\beta\right) \\
&= \sum_{p=0}^{\infty} \frac{|\lambda|^{\beta p}\mathbb{E}\,|X|^{\beta p}}{p!} \\
&\leq \sum_{p=0}^{\infty} |\lambda|^{\beta p} \left(\frac{\beta p}{\alpha}\right)^{\beta p/\alpha} \frac{e^p}{p^p} \\
&= \sum_{p=0}^{\infty} |\lambda|^{\beta p} p^{p(\beta/\alpha - 1)} (\alpha-1)^{-p/(\alpha-1)} e^p \\
&\leq \sum_{p=0}^{\infty} |\lambda|^{\beta p} (\alpha-1)^{-p/(\alpha-1)} e^p \\
&= \frac{1}{1-|\lambda|^\beta(\alpha-1)^{-1/(\alpha-1)}e} \\
&\leq \exp\left(2(\alpha-1)^{-1/(\alpha-1)} e|\lambda|^\beta\right) \\
&\leq \exp\left(2e|\lambda|^\beta\right) \\
&\leq \exp\left((2e)^\beta|\lambda|^\beta\right)
\end{aligned}
$$

for all λ satisfying $|\lambda|^\beta \leq (\alpha-1)^{1/(\alpha-1)}/2e$. For λ with large absolute value, by Young's inequality,

$$\lambda x \leq |\lambda||x| \leq \frac{|\lambda|^\beta}{\beta} + \frac{|x|^\alpha}{\alpha}.$$

It follows that

$$\mathbb{E}\,e^{\lambda X} \leq e^{|\lambda|^\beta/\beta}\mathbb{E}\,e^{|X|^\alpha/\alpha} \leq e^{|\lambda|^\beta/\beta}e^{1/\alpha} \leq e^{C|\lambda|^\beta}$$

for $|\lambda|^\beta \geq (\alpha-1)^{1/(\alpha-1)}/2e$ and $C = \frac{\alpha-1}{\alpha} + \frac{2e(\alpha-1)^{-1/(\alpha-1)}}{\alpha} \leq 2e+1$. Thus $\mathbb{E}\,e^{\lambda X} \leq \exp\left[(2e+1)^\beta|\lambda|^\beta\right]$.

e $\Rightarrow$ a. Assume $K_5 = 1$. Let $\lambda > 0$.

$$\mathbb{P}\left(X \geq t\right) = \mathbb{P}\left(e^{\lambda X} \geq e^{\lambda t}\right) \leq e^{-\lambda t}\mathbb{E}\,e^{\lambda X} \leq \exp\left(-\lambda t + \lambda^{\alpha/(\alpha-1)}\right).$$

Substituting $\lambda = t^{\alpha-1}(\alpha-1)^{\alpha-1}/\alpha^{\alpha-1}$ gives

$$\mathbb{P}\left(X \geq t\right) \leq \exp\left(-\frac{(\alpha-1)^{\alpha-1}}{\alpha^\alpha}t^\alpha\right) \leq \exp\left(-t^\alpha/2^\alpha\right).$$

Repeating this argument for $-X$, we obtain the same bound for $\mathbb{P}(X \leq -t)$. We conclude that

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

c $\Rightarrow$ f. Assume $K_3 = 1$. Let $\beta = \alpha/(\alpha-1)$. By Young's inequality,

$$\mathbb{E}\, e^{\lambda X} \leq e^{|\lambda|^\beta/\beta}\mathbb{E}\, e^{|X|^\alpha/\alpha} \leq e^{|\lambda|^\beta/\beta} e^{1/\alpha} \leq \exp\left(C|\lambda|^\beta\right)$$

for $|\lambda| \geq 1/2e$ and $C = \frac{1}{\beta} + \frac{(2e)^\beta}{\alpha} \leq (2e+1)^\beta$.

f $\Rightarrow$ a. Assume $K_6 = 1$. If $t \leq 2(\log 2)^{1/\alpha}$, we have $2\exp(-t^\alpha/2^\alpha) \geq 1$. Otherwise,

$$\lambda := t^{\alpha-1}(\alpha-1)^{\alpha-1}/\alpha^{\alpha-1} \geq 2^{\alpha-1}(\log 2)^{(\alpha-1)/\alpha}\frac{(\alpha-1)^{\alpha-1}}{\alpha^{\alpha-1}} \geq 1/2e.$$

From the proof of e $\Rightarrow$ a, we have

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

$\square$

Exercise 2.7.4 Let $X \sim Exp(1)$. For $|\lambda| < 1$, we have $\mathbb{E}\exp(\lambda X) = \frac{1}{1-\lambda}$. Suppose $f(x) := e^{kx} - \frac{1}{1-x} \geq 0$ for $x \in [-\delta, \delta]$ with $\delta > 0$. $f(\delta) \geq 0$ gives $k \geq -\frac{\log(1-\delta)}{\delta} > 1$. Then we have $f'(0) > 0$. Note that $f(0) = 0$. This induces a contradiction.

Exercise 2.7.11 The proof is similar to Exercise 2.5.7.

Exercise 2.8.5 For $|z| < 3$, we have

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

Then

$$\mathbb{E}\exp(\lambda X) \leq 1 + \mathbb{E}\,\frac{\lambda^2 X^2/2}{1-\lambda|X|/3} \leq \exp(\mathbb{E}\, X^2 g(\lambda)).$$

Exercise 2.8.6

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

Substituting $\lambda = \frac{t}{\sigma^2+Kt/3}$ gives

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

Repeating the argument for $-X_i$, we obtain the same bound for $\mathbb{P}\left\{\sum_{i=1}^{N} \leq -t\right\}$. A combination of the two bounds completes the proof.

3 Random vectors in high dimensions

Exercise 3.1.4

(a) By Theorem 3.1.1 we have $\|\|X\|_2 - \sqrt{n}\|_{\psi_2} \leq C_0 K^2$. Recall that for sub-gaussian r.v. X, $\|X\|_{L^p} \leq C_1 \|X\|_{\psi_2}\sqrt{p}$, for all $p \geq 1$. Hence we have:

$$|\mathbb{E}\,\|X\|_2 - \sqrt{n}| \leq \mathbb{E}\,|\|X\|_2 - \sqrt{n}| \leq C_1 \|X\|_{\psi_2} \leq C_0 C_1 K^2$$

The proof is completed.

(b) By the same argument used in **(a)** we can get $\mathbb{E}\,(\|X\|_2 - \sqrt{n})^2 \leq (CK^2\sqrt{2})^2$, which means $2\sqrt{n}\mathbb{E}\,\|X\|_2 \geq n + \mathbb{E}\,\|X\|_2^2 - 2C^2K^4$. Note that $\mathbb{E}\,\|X\|_2^2 = n$, thus $\mathbb{E}\,\|X\|_2 \geq \sqrt{n} - 2C^2K^4/\sqrt{n} = \sqrt{n} + o(1)$. And it is trivial that $E\|X\|_2 \leq (E\|X\|_2^2)^{1/2} = \sqrt{n}$. So the answer is yes.

Exercise 3.1.5 Use Exercise 3.1.4 we have $\mathbb{E}\,\|X\|_2 \geq \sqrt{n} - 2C^2K^4/\sqrt{n}$. Therefore,

$$\mathrm{Var}\,(\|X\|_2) = \mathbb{E}\,\|X\|_2^2 - (\mathbb{E}\,\|X\|_2)^2 \leq n - \left(\sqrt{n} - \frac{2C^2K^4}{\sqrt{n}}\right)^2 = 4C^2K^4 - \frac{4C^4K^8}{n}$$

The proof is completed.

Exercise 3.1.6 We have

$$\mathbb{E}\,(\|X\|_2-\sqrt{n})^2 \le \frac{\mathbb{E}\,(\|X\|-\sqrt{n})^2(\|X\|+\sqrt{n})^2}{n} = \frac{\mathbb{E}\,(\|X\|_2^2-n)^2}{n}.$$

Note that

$$\begin{aligned}\mathbb{E}\,(\|X\|_2^2-n)^2 &= \mathbb{E}\,\|X\|_2^4+n^2-2n\mathbb{E}\,\|X\|_2^2\\ &= \sum_{i\neq j}\mathbb{E}\,X_i^2X_j^2+\sum_i \mathbb{E}\,X_i^4 = nK^4+n(n-1)+n^2-2n^2\\ &\le nK^4.\end{aligned}$$

Therefore we can get $\mathrm{Var}\,(\|X\|_2)\le \mathbb{E}\,(\|X\|_2-\sqrt{n})^2\le \mathbb{E}\,(\|X\|_2^2-n)^2/n = K^4$.

Exercise 3.1.7 For any $\lambda>0$, we have

$$\begin{aligned}\mathbb{P}\,(\|X\|_2\le\epsilon\sqrt{n}) &= \mathbb{P}\,(-\|X\|_2^2\ge-\epsilon^2 n)\\ &= \mathbb{P}\,(e^{-\lambda\|X\|_2^2}\ge e^{-\lambda\epsilon^2 n})\\ &\le \frac{\mathbb{E}\,e^{-\lambda\|X\|_2^2}}{e^{-\lambda\epsilon^2 n}} \qquad\qquad ,\\ &= \left(\frac{\mathbb{E}\,e^{-\lambda X_1^2}}{e^{-\lambda\epsilon^2}}\right)^n\end{aligned}$$

and

$$\mathbb{E}\,e^{-\lambda x_1^2} = \int e^{-\lambda t^2}f(t)dt \le \int e^{-\lambda t^2}dt = \sqrt{\frac{\pi}{\lambda}}.$$

Setting $\lambda=\epsilon^{-2}$ yields

$$\begin{aligned}\mathbb{P}\,(\|X\|_2\le\epsilon\sqrt{n}) &\le \left(\frac{\mathbb{E}\,e^{-\lambda X_1^2}}{e^{-\lambda\epsilon^2}}\right)^n\\ &\le (e^{\lambda\epsilon^2}\sqrt{\pi/\lambda})^n\\ &= (e\sqrt{\pi}\epsilon)^n.\end{aligned}$$

Exercise 3.3.1 By rotational invariance property of X, it suffices to prove $\mathbb{E}\,X_1^2=1$ and $\mathbb{E}\,X_1X_2=0$. Since we always have $X_1^2+\cdots+X_n^2=n$, $\mathbb{E}\,X_1^2=1$ holds by the rotational invariance, too. Note that $X_1X_2\overset{d}{=}-X_1X_2$, so we have $\mathbb{E}\,X_1X_2=0$. It is obvious the coordinates of

X are not independent, as $X_1^2 + \cdots + X_n^2 = n$ always holds.

Exercise 3.3.7 Note that $dg = r^{n-1} dr d\sigma(\theta)$, where $\sigma(\theta)$ denotes the area element of $\mathbb{S}^{n-1}$. Thus we have $\mathrm{e}^{-\|g\|_2^2/2}\mathrm{d}g = C\mathrm{e}^{-r^2/2} r^{n-1}\mathrm{d}r \cdot \mathrm{d}\sigma(\theta)$, which completes the proof.

Exercise 3.4.3

1. By triangular inequality we have

$$\begin{aligned}
\|\langle X, x\rangle\|_{\psi_2} &= \left\|\sum_{i=1}^n X_i x_i\right\|_{\psi_2} \\
&\leq \sum_{i=1}^n \|X_i x_i\|_{\psi_2} \\
&= \sum_{i=1}^n |x_i| \|X_i\|_{\psi_2} \\
&\leq \left(\sum_{i=1}^n \|X_i\|_{\psi_2}^2\right)^{1/2}, \quad \forall x \in \mathbb{S}^{n-1}.
\end{aligned}$$

The last inequality holds by Cauchy-Schwartz's inequality.

2. Let $X_1 \sim N(0,1)$, and we simply define $X = (X_1, \ldots, X_1)$. Then

$$\begin{aligned}
\|X\|_{\psi_2} &\geq \|\langle X, 1_n/\sqrt{n}\rangle\|_{\psi_2} \\
&= \sqrt{n}\|X_1\|_{\psi_2} \\
&\gg \|X_1\|_{\psi_2}.
\end{aligned}$$

Exercise 3.4.4 First we'll verify $\|X_i\|_{\psi_2} = \sqrt{n/\log(n+1)}$, $\forall i \leq n$. Recall that:

$$\|X_i\|_{\psi_2} = \inf\{K > 0 : \mathbb{E}\exp\{X_i^2/K^2\} \leq 2\}.$$

Let $K = \sqrt{n/\log(n+1)}$ and plug it into $\mathbb{E}\exp\{X^2/K^2\}$ yields:

$$\begin{aligned}\mathbb{E}\exp\{X_i^2/K^2\} &= \frac{1}{n}\exp\{n/K^2\} + \frac{n-1}{n} \\ &= \frac{n+1}{n} + \frac{n-1}{n} \\ &= 2.\end{aligned}$$

So $\|X_i\|_{\psi_2} = \sqrt{n/\log(n+1)},\ i \le n$, and $\|X\|_{\psi_2} \ge \|\langle X, e_i\rangle\|_{\psi_2} = \|X_i\|_{\psi_2} = \sqrt{n/\log(n+1)}$. In addition, $\forall x \in \mathbb{S}^{n-1}$, we have

$$\begin{aligned}\mathbb{E}\exp\{\langle X, x\rangle^2/K^2\} &= \frac{1}{n}\sum_{i=1}^{n}\exp\{nx_i^2/K^2\} \\ &\le \frac{1}{n}\exp\{n/K^2\} + \frac{n-1}{n} \quad \text{(by Karamata's inequality)} \\ &= \mathbb{E}\exp\{X_i^2/K^2\}\end{aligned}$$

Therefore, $\|X\|_{\psi_2} \le \|X_i\|_{\psi_2} = \sqrt{n/\log(n+1)}$. And we may conclude that $\|X\|_{\psi_2} = \sqrt{n/\log(n+1)}$.

Exercise 3.4.5 Recall that:

$$\|X\|_{\psi_2} = \sup_{\|\mathbf{y}\|=1} \inf\{K > 0 : \mathbb{E}\exp\{(X^\top \mathbf{y})^2/K^2\} \le 2\}.$$

By the assumption $\|X\|_{\psi_2} \le C$,

$$\exists C_0 > 0, \text{ s.t. } \mathbb{E}\exp\{(X^\top \mathbf{y})^2/C_0^2\} \le 3,\ \forall\, \mathbf{y} \in \mathbb{S}^{d-1}.$$

Let $T := \{\mathbf{x}_1, \dots,\ \mathbf{x}_{|T|}\}, P(X = \mathbf{x}_i) := p_i$.

$$p_j \exp\{(\mathbf{x}_j^\top \mathbf{y})^2/C_0^2\} \le \sum_{i=1}^{|T|} p_i \exp\{(\mathbf{x}_i^\top \mathbf{y})^2/C_0^2\} = \mathbb{E}\exp\{(X^\top \mathbf{y})^2/C_0^2\} \le 3$$

Thus

$$\left(\mathbf{x}_j^\top \mathbf{y}\right)^2 \le C_0^2 \ln\frac{3}{p_j},\ \forall\, \mathbf{y} \in \mathbb{S}^{d-1}.$$

Therefore

$$\|\mathbf{x}_j\|^2 \le C_0^2 \ln \frac{3}{p_j}, \ \forall\, \mathbf{y} \in \mathbb{S}^{d-1}.$$

Notice that $\mathbb{E} XX^\top = I_n$ (X is isotropic),

$$n = \operatorname{tr}\left(\mathbb{E} XX^\top\right) = \sum_{i=1}^{|T|} p_i \|\mathbf{x}_i\|_2^2 \le \sum_{i=1}^{|T|} C_0^2 p_i \ln \frac{3}{p_i} \le t_0^2 \ln(3|T|).$$

The last inequality follows from $x \ln x$ is convex and Jensen inequality. Hence

$$|T| \ge \frac{1}{3} \exp\left\{ \frac{n}{C_0^2} \right\}.$$

Exercise 3.4.7 Since X enjoys the rotational invariance property, it suffices to prove $\|X_1\|_{\psi_2} \le C$. Define $Y = \sqrt{n}\frac{X}{\|X\|_2}$, it follows that $Y \sim \text{Unif}(\sqrt{n}\mathbb{S}^{n-1})$. From Theorem 3.4.6 we know $\|Y_1\|_{\psi_2} \le C$, and $|Y_1| \ge |X_1|$ always holds. So we can get $\|X_1\|_{\psi_2} \le C$.

Exercise 3.4.9

(a) By rotational invariance we have $\mathbb{E}\, X_i X_j = 0,\ i \ne j$ and $\mathbb{E}\, X_1^2 = \cdots \mathbb{E}\, X_n^2$. Then it is clear that one can make the distribution isotropic by scaling the ball. We denote a n-dimensional l_1-ball of radius r by $B_n(r)$. Suppose $X \sim \text{Unif}(B_n(1))$, then

$$\begin{aligned}
\mathbb{P}\left(|X_1| \le t\right) &= \int_{-t}^{t} \int_{-1+|x_1|}^{1-|x_1|} \cdots \int_{-1+|x_1|+\cdots+|x_{n-1}|}^{1-|x_1|-\cdots-|x_{n-1}|} \frac{1}{\text{Vol}(B_n(1))} dx_n \ldots dx_2 dx_1 \\
&= \int_{-t}^{t} \frac{\text{Vol}(B_{n-1}(1-|x_1|))}{\text{Vol}(B_n(1))} dx_1 \\
&= \int_{-t}^{t} (1-|x_1|)^{n-1} dx_1 \frac{\text{Vol}(B_{n-1}(1))}{\text{Vol}(B_n(1))}
\end{aligned}$$

Setting $t = 1$ yields $\mathbb{P}\left(|X_1| \le 1\right) = \frac{2}{n}\frac{\text{Vol}(B_{n-1}(1))}{\text{Vol}(B_n(1))} = 1$, so

$$\mathbb{P}\left(|X_1| > t\right) = 1 - \frac{n}{2} \int_{-t}^{t} (1-|x_1|)^{n-1} dx_1 = (1-t)^n.$$

Now using results from Exercise 1.2.3 we have:

$$\mathbb{E}\, X_1^2 = \int_0^1 2t\mathbb{P}\left(|X_1| > t\right)dt = \int_0^1 2t(1-t)^n dt = \frac{2}{(n+1)(n+2)}.$$

Therefore the scale factor should be $r = \sqrt{(n+1)(n+2)/2}$.

(b) Using similar argument in part (a), we may get for $X \sim \mathrm{Unif}(B_n(r))$,

$$\mathbb{P}\left(|X_1| > t\right) = \left(1 - \frac{t}{r}\right)^n \to e^{-t},\ n \to \infty,$$

since $r \asymp n$. So obviously $\|X_1\|_{\psi_2}$ is not bounded by any an absolute constant as n grows, and neither is $\|X\|_{\psi_2}$.

Exercise 3.4.10 (This example is copied from *https://mathoverflow.net/questions/326183/*) Define $\mu_X = \frac{1}{2}\mu_{aZ} + \frac{1}{2}\mu_{bZ}$, where μ_U denotes the probability distribution of a random vector U, $Z \sim N(0, I_n)$, and $a,\ b$ are constants such that $0 < a < 1 < b$, $a^2 + b^2 = 2$. Then one can verify that X is isotropic. For any unit vector u and a real number $s > 0$,

$$\mathbb{E}\, \exp\left\{\langle X, u\rangle^2 / s^2\right\} = \frac{1}{2\sqrt{1 - 2a^2/s^2}} + \frac{1}{2\sqrt{1 - 2b^2/s^2}} < 2,$$

if s is large enough (depending only on a, b), so that by the definition of sub-gaussian norm we have $\|X\|_{\psi_2} \le s$.

On the other hand, let $t := (b-1)\sqrt{n}/2$, we have

$$\begin{aligned}
2\mathbb{E}\, e^{(\|X\| - \sqrt{n})^2/t^2} &> \mathbb{E}\, e^{(\|bZ\| - \sqrt{n})^2/t^2} \\
&> \mathbb{E}\, e^{(\|bZ\| - \sqrt{n})^2/t^2} 1_{\|Z\|^2 > n} \\
&> e^{(b\sqrt{n} - \sqrt{n})^2/t^2}\, \mathbb{P}\left(\|Z\|^2 > n\right) \\
&= e^4 \mathbb{P}\left(\|Z\|^2 > n\right) \to e^4/2 \\
&> 4,
\end{aligned}$$

since $\mathbb{P}\left(\|Z\|^2 > n\right) \to 1/2$ by CLT. Therefore, $\|\|X\| - \sqrt{n}\|_{\psi_2} \ge t = (b-1)\sqrt{n}/2 \to \infty$, as desired.

Exercise 3.5.3 First we'll prove the following lemma:

Lemma. *Suppose A is either positive-semidefinite or has zero diagonal, we have*

$$\max_{x\in\{-1,1\}^n} |\langle Ax, x\rangle| = \max_{x\in\{-1,0,1\}^n} |\langle Ax, x\rangle|$$

Proof. Obviously we have $\max_{x\in\{-1,1\}^n} |\langle Ax, x\rangle| \le \max_{x\in\{-1,0,1\}^n} |\langle Ax, x\rangle|$. And next we will prove $\max_{x\in\{-1,1\}^n} |\langle Ax, x\rangle| \ge \max_{x\in\{-1,0,1\}^n} |\langle Ax, x\rangle|$.

When A is positive-semidefinite, we always have $|\langle Ax, x\rangle| = \langle Ax, x\rangle$, and $\langle Ax, x\rangle$ is convex in x. Also note that $[-1,1]^n = \mathrm{conv}\{u : u \in \{-1,1\}^n\}$ is a polyhedron with vertices $\{u : u \in \{-1,1\}^n\}$, so $\langle Ax, x\rangle$ must attain its maximum at one of the vertices. Then we may get

$$\max_{x\in\{-1,0,1\}^n} \langle Ax, x\rangle \le \max_{x\in[-1,1]^n} \langle Ax, x\rangle = \max_{x\in\{-1,1\}^n} \langle Ax, x\rangle$$

When A has zero diagonal, suppose

$$\max_{x\in\{-1,1\}^n} |\langle Ax, x\rangle| < \max_{x\in\{-1,0,1\}^n} |\langle Ax, x\rangle|,$$

define $x_0 = \mathrm{argmax}_{x\in\{-1,0,1\}^n} \langle Ax, x\rangle$, then x_0 has at least one zero coordinate, thus we can assume $x_0(i_0) = 0$. Consider $x \in \{-1,0,1\}^n$ such that $x(j) = x_0(j)$, $\forall j \ne i_0$, we have

$$\langle Ax, x\rangle - \langle Ax_0, x_0\rangle = 2\sum_{j\ne i_0} a_{i_0 j} x(i_0)x(j) = 2x(i_0)\sum_{j\ne i_0} a_{i_0 j} x(j).$$

Then setting $x_1 = x_0 + \mathrm{sgn}(\langle Ax_0, x_0\rangle) \cdot \mathrm{sgn}(\sum_{j\ne i_0} a_{i_0 j} x(j)) \cdot e_{i_0}$ yields $|\langle Ax_1, x_1\rangle| \ge |\langle Ax_0, x_0\rangle|$. Repeat this procedure for at most n times we will get a $x_* \in \{-1,1\}^n$ such that $|\langle Ax_*, x_*\rangle| \ge |\langle Ax_0, x_0\rangle|$, which is a contradiction. $\square$

Suppose $x, y \in \{-1,1\}^n$, by polarization identity we'll get $\langle Ax, y\rangle = \langle Au, u\rangle - \langle Av, v\rangle$, where $u = (x+y)/2 \in \{-1,0,1\}$ and $v = (x-y)/2 \in \{-1,0,1\}$. Since A is positive-semidefinite or has zero diagonal and $|\langle Ax, x\rangle| \le 1$, $\forall x \in \{-1,1\}^n$, from the lemma we know $|\langle Au, u\rangle| \le 1$, $|\langle Av, v\rangle| \le 1$. Thus $|\langle Ax, y\rangle| \le |\langle Au, u\rangle| + |\langle Av, v\rangle| = 2$, $\forall x, y \in \{-1,1\}$. Then the conclusion holds via Grothendieck's inequality.

Exercise 3.6.4 The algorithm is simple. We may just repeat the procedure in Proposition 3.6.3, and terminate as long as the number of cuts exceeds $(0.5-\epsilon)|E|$. Define $p_0 = \mathbb{P}\,(\text{a single try fails})$, then we may find

$$\begin{aligned} p_0 &= \mathbb{P}\left(\frac{1}{4}\sum_{i,j=1}^{n} A_{ij}(1-x_i x_j) \le \left(\frac{1}{2}-\epsilon\right)|E|\right) \\ &= \mathbb{P}\left(\frac{1}{4}\sum_{i,j=1}^{n} A_{ij}x_i x_j \ge \epsilon|E|\right) \\ &= \mathbb{P}\left(\frac{1}{4}\sum_{i,j=1}^{n} A_{ij}\,\mathrm{sign}(x_i)\,\mathrm{sign}(x_j) \ge \epsilon|E|\right) \end{aligned}$$

Let $f(x) = \frac{1}{4}\sum_{i,j=1}^{n} A_{ij}\,\mathrm{sign}(x_i)\,\mathrm{sign}(x_j)$. By Theorem 2.9.1, we have

$$p_0 \le \exp\left(-\frac{8\epsilon^2|E|^2}{\sum_{i=1}^{n} d_i^2}\right),$$

where d_i is the degree of the i-th vertex. It remains to bound $\sum_{i=1}^{n} d_i^2$. D.de Caen has proved [2]

$$\sum_{i=1}^{n} d_i^2 \le |E|\left(\frac{2|E|}{n-1}+n-2\right) \quad \text{when } n \ge 2.$$

Then we obtain

$$p_0 \le \exp\left(-\frac{8\epsilon^2|E|}{2|E|/(n-1)+n-2}\right) \quad \text{when } n \ge 2.$$

Futhermore, the number of runs needed N obeys the geometric distribution with parameter $1-p_0$. And $\mathbb{P}\,(N<\infty)=1$, $\mathbb{E}\,N = 1/(1-p_0)$.

Exercise 3.6.7 First we set $(X_1, X_2) := (\langle g,u\rangle, \langle g,v\rangle) \sim N(0, \left[\begin{smallmatrix} 1 & \cos\alpha \\ \cos\alpha & 1\end{smallmatrix}\right])$, then we have

$$\begin{aligned} &\mathbb{E}\,\mathrm{sign}\langle g,u\rangle \mathrm{sign}\langle g,v\rangle \\ &= \mathbb{P}\,(X_1>0, X_2>0) + \mathbb{P}\,(X_1\le 0, X_2\le 0) - \mathbb{P}\,(X_1\le 0, X_2>0) - \mathbb{P}\,(X_1>0, X_2\le 0) \\ &= 2\mathbb{P}\,(X_1>0,X_2>0) - (1-2\mathbb{P}\,(X_1>0,X_2>0)) \qquad \text{(by symmetry)} \\ &= 4\mathbb{P}\,(X_1>0,X_2>0) - 1 \end{aligned}$$

Note that $(X_1, X_2) \stackrel{d}{=} (Z_1, \sin\alpha Z_1 + \cos\alpha Z_2)$, where $(Z_1, Z_2) \sim N(0, I_2)$. So

$$\mathbb{P}\left(X_1 > 0, X_2 > 0\right) = \mathbb{P}\left(Z_1 > 0, \sin\alpha Z_1 + \cos\alpha Z_2 > 0\right) = \frac{\pi - \alpha}{2\pi}.$$

Thus $\mathbb{E}\,\mathrm{sign}\langle g, u\rangle \mathrm{sign}\langle g, v\rangle = 4\mathbb{P}\left(X_1 > 0, X_2 > 0\right) - 1 = (\pi - 2\alpha)/\pi = 2\arcsin\langle u, v\rangle/\pi$.

4 Random matrices

Exercise 4.1.6 Choose $x \in S^{n-1}$ such that $\|A^\top A - I_n\| = \langle \big(A^\top A - I_n\big)x, x\rangle$. Note that (4.7) implies $|\|Ax\|_2 - 1| \le \delta$. Then we have

$$\|A^\top A - I_n\| = |\|Ax\|_2^2 - 1| = |\|Ax\|_2 - 1|\left(\|Ax\|_2 + 1\right) \le \delta(2+\delta) \le 3\,\max\big(\delta, \delta^2\big).$$

Exercise 4.2.5 (b) Let $T = \{0,1\}^3$ with d being the Hamming distance. Consider $K = \{(1,0,0),(1,1,0),(1,1,1)\}$ and $\epsilon = 1.2$. Then $\mathcal{P}(K, d, 1.1) = 2$. However, there are 3 disjoint balls with centers in K and radii $\epsilon/2$.

Exercise 4.2.9 The first inequality is trivial. To prove the second inequality, let $\mathcal{N} = \{x_1, x_2, \ldots, x_{|\mathcal{N}|}\}$ be an $(\epsilon/2)$-net of K without requiring $\mathcal{N} \subset K$. For each $x_i \in \mathcal{N}$, there exists $\tilde{x}_i \in K$ such that $\|x_i - \tilde{x}_i\| \le \epsilon/2$ and $B(x_i, \epsilon/2) \subset B(\tilde{x}_i, \epsilon)$. It follows that $K \subset \cup_{i=1}^{|\mathcal{N}|} B(x_i, \epsilon/2) \subset \cup_{i=1}^{|\mathcal{N}|} B(\tilde{x}_i, \epsilon)$, which implies $\mathcal{N}(K, d, \epsilon) \le \mathcal{N}^{\mathrm{ext}}(K, d, \epsilon/2)$.

Exercise 4.2.10 Let $T = \mathbb{R}^n$ with d being the Euclidean distance. Consider $K = B_2^n$, $L = B_2^n \backslash \{0\}$ and $\epsilon = 1$. We have $\mathcal{N}(K, d, \epsilon) = 1$ but $\mathcal{N}(L, d, \epsilon) > 1$. The second inequality follows from Exercise 4.2.9 and the monotonicity of $\mathcal{N}^{\mathrm{ext}}(K, d, \epsilon)$.

Exercise 4.2.16 Let $B_m = \{x \in K : d_H(x, e_0) \le m\} \subset K$ where $e_0 = (0, 0, \ldots, 0)$. By symmetry, $\mathcal{N}(K, d_H, m) \ge \frac{|K|}{|B_m|}$ and $\mathcal{P}(K, d_H, m) \le \frac{|K|}{|B_{m/2}|}$. Since $|K| = 2^n$, $|B_m| = \sum_{k=0}^{m} \binom{n}{k}$ and $|B_{m/2}| = \sum_{k=0}^{\lfloor m/2 \rfloor} \binom{n}{k}$, we get the desired inequality.

Exercise 4.3.7

(a) If $\mathcal{P}(\{0,1\}^n, d_H, 2r) \le 2^k - 1$, for any encoding may E, there exist $x, y \in \{0,1\}^k$ such that $d_H(E(x), E(y)) \le 2r$ and we can find $z \in \{0,1\}^n$ such that $d_H(E(x), z) \le r$ and $d_H(E(y), z) \le r$. If we receive z, we can not determine whether the original letter is x or y.

(b) Note that $f(\delta) \ge h(\delta)$ where $h(\delta) = -x\log_2(x) - (1-x)\log_2(1-x)$. The upper bound $R \le 1 - f(\delta)$ is tighter than the Hamming bound in Section 4.8. We prove a weaker result $R \le 1 - \delta\log_2\left(\frac{1}{\delta}\right)$. From part (a), we have $\mathcal{P}(\{0,1\}^n, d_H, 2r) \ge 2^k$. Combining Exercise 0.0.5 and 4.2.16, we obtain

$$2^k \le \frac{2^n}{\sum_{k=0}^{r}\binom{n}{k}} \le \frac{2^n}{(n/r)^r}.$$

It follows that $R \le 1 - \delta\log_2\left(\frac{1}{\delta}\right)$.

Exercise 4.4.4 If $\mu \le 0$, Lemma 4.4.1 implies

$$\sup_{x\in S^{n-1}} |\|Ax\| - \mu| = \sup_{x\in S^{n-1}} \|Ax\| - \mu \le \frac{1}{1-\epsilon}\left(\sup_{x\in\mathcal{N}} \|Ax\| - \mu\right) = \frac{1}{1-\epsilon}\sup_{x\in\mathcal{N}} |\|Ax\| - \mu|.$$

If $\mu > 0$, without loss of generality we assume $\mu = 1$.
When $\|A\| > 2$, we have $\sup_{x\in S^{n-1}} |\|Ax\| - 1| = \|A\| - 1$ and $\sup_{x\in\mathcal{N}} |\|Ax\| - 1| \ge (1-\epsilon)\|A\| - 1$. It follows that

$$\sup_{x\in S^{n-1}} |\|Ax\| - 1| = \frac{(1-\epsilon)\|A\| - 1 - \epsilon(\|A\| - 2)}{1-2\epsilon} \le \frac{1}{1-2\epsilon}\sup_{x\in\mathcal{N}} |\|Ax\| - 1|.$$

When $\|A\| \le 2$, by Exercise 4.4.3, we have

$$\sup_{x\in S^{n-1}} |\|Ax\| - 1| \le \sup_{x\in S^{n-1}} |\|Ax\|^2 - 1| \le \frac{1}{1-2\epsilon}\sup_{x\in\mathcal{N}} |\|Ax\|^2 - 1| \le \frac{3}{1-2\epsilon}\sup_{x\in\mathcal{N}} |\|Ax\| - 1|$$

Thus we can choose $C = 3$.

Exercise 4.4.6 By Lemma 1.2.1 and change of variables,

$$\begin{aligned}
\mathbb{E}\|A\| &= \int_0^\infty \mathbb{P}\{\|A\| \geq u\} du \\
&\leq CK\big(\sqrt{m}+\sqrt{n}\big) + \int_{CK(\sqrt{m}+\sqrt{n})}^\infty \mathbb{P}\{\|A\| \geq u\} du \\
&= CK\big(\sqrt{m}+\sqrt{n}\big) + CK\int_0^\infty \mathbb{P}\{\|A\| \geq CK\big(\sqrt{m}+\sqrt{n}+t\big)\} dt \\
&= CK\big(\sqrt{m}+\sqrt{n}\big) + 2CK\int_0^\infty \exp(-t^2) dt \\
&\leq \tilde{C}K\big(\sqrt{m}+\sqrt{n}\big)
\end{aligned}$$

for $m, n \geq 1$.

Exercise 4.4.7 Let $A_{1\cdot}^\top$ denote the first row of A and $A_{\cdot 1}$ denote the first column of A. Note that $Ae_1 = A_{\cdot 1}$ and $A^\top \tilde{e}_1 = A_{1\cdot}$ where $e_1 = (1,0,0,\ldots,0)^\top \in \mathbb{R}^n$ and $\tilde{e}_1 = (1,0,0,\ldots,0)^\top \in \mathbb{R}^m$. Then we have

$$\|A\| = \frac{\|A\| + \|A^\top\|}{2} \geq \frac{\|A_{\cdot 1}\| + \|A_{1\cdot}\|}{2}.$$

By Exercise 3.1.4, $\mathbb{E}\|A\| \geq C(\sqrt{m}+\sqrt{n})$ for some $C > 0$.

Exercise 4.5.4 For any i, we have

$$\begin{aligned}
\lambda_i(S) &= \max_{\dim E = i} \min_{x \in S(E)} \langle Sx, x\rangle \\
&\leq \max_{\dim E = i} \min_{x \in S(E)} \left(\langle Tx, x\rangle + \langle (S-T)x, x\rangle\right) \\
&\leq \max_{\dim E = i} \min_{x \in S(E)} \langle Tx, x\rangle + \|S-T\| \\
&= \lambda_i(T) + \|S-T\|
\end{aligned}$$

By symmetry, $\max_i |\lambda_i(S) - \lambda_i(T)| \leq \|S-T\|$.

Exercise 4.6.2 Denote $M = \big\|\frac{1}{m}A^\top A - I_n\big\|$. Consider $u = K^2\max\big(\delta, \delta^2\big)$ with $\delta = C\Big(\sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}}\Big)$. When $t \leq \sqrt{m}/C - \sqrt{n}$, $u = K^2C\Big(\sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}}\Big) \leq K^2$. Otherwise, $u = K^2C^2\Big(\sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}}\Big)^2 \geq K^2$.

By Lemma 1.2.1 and change of variables,

$$
\begin{aligned}
\mathbb{E}\, M &= \int_0^{K^2} \mathbb{P}\{M \geq u\} du + \int_{K^2}^{\infty} \mathbb{P}\{M \geq u\} du \\
&= K^2 \wedge \left(K^2 C \sqrt{\frac{n}{m}}\right) + \int_{K^2 \wedge \left(K^2 C \sqrt{\frac{n}{m}}\right)}^{K^2} \mathbb{P}\{M \geq u\} du \\
&\quad + \left(K^2 C^2 \frac{n}{m}\right) \vee K^2 - K^2 + \int_{(K^2C^2)\frac{n}{m} \vee K^2}^{\infty} \mathbb{P}\{M \geq u\} du \\
&\leq K^2 C \sqrt{\frac{n}{m}} + \frac{2K^2 C}{\sqrt{m}} \int_0^{\infty} \exp(-t^2) dt + \frac{K^2 C^2 n}{m} + \frac{4K^2C^2}{m} \int_0^{\infty} \left(\sqrt{n} + t\right) \exp(-t^2) dt \\
&\leq \tilde{C} K^2 \left(\sqrt{\frac{n}{m}} + \frac{n}{m}\right)
\end{aligned}
$$

for some $\tilde{C} \geq 0$ and $m, n \geq 1$.

Exercise 4.6.4 Note that for any $x \in S^{n-1}$, $Ax \in \mathbb{R}^m$ satisfy the assumptions of Theorem 3.1.1. Let $\mathcal{N}$ be an 1/4-net of S^{n-1}. By Exercise 4.4.4, we have

$$
\mathbb{P}\left\{\sup_{x \in S^{n-1}} \left|\|Ax\| - \sqrt{m}\right| \geq u\right\} \leq \mathbb{P}\left\{\sup_{x \in \mathcal{N}} \left|\|Ax\| - \sqrt{m}\right| \geq \frac{u}{6}\right\} \leq 9^n \cdot 2\exp\left(-\frac{Cu^2}{K^4}\right)
$$

for some $C > 0$. Taking $u = \tilde{C}K^2(\sqrt{n} + t)$ for a sufficiently large $\tilde{C}$, we obtain the desired result.

Exercise 4.7.3 Recall Theorem 4.6.1 and note that with $\delta = C\left(\sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}}\right)$ and $u = t^2$, we have

$$
\max(\delta, \delta^2) \leq \delta + \delta^2 \leq C\left(\sqrt{\frac{n}{m}} + \sqrt{\frac{u}{m}}\right) + 2K^2C^2\left(\frac{n}{m} + \frac{u}{m}\right) \leq \tilde{C}\left(\sqrt{\frac{n+u}{m}} + \frac{n+u}{m}\right),
$$

where $\tilde{C} = 2\max(C, C^2)$.

5 Concentration without independence

Exercise 5.1.8 For any $x \in \sqrt{n}S^{n-1}$ such that $0 \leq x_1 \leq t/\sqrt{2}$, let

$$
\tilde{x} = (0, \operatorname{sign}(x_2)\sqrt{x_1^2 + x_2^2}, x_3, x_4, \cdots, x_n)
$$

We have $\tilde{x} \in H$ and

$$\|x - \tilde{x}\|^2 = x_1^2 + \left(x_2 - \operatorname{sign}(x_2)\sqrt{x_1^2 + x_2^2}\right)^2 = 2x_1^2 + 2x_2^2 - 2|x_2|\sqrt{x_1^2 + x_2^2} \le t^2.$$

Exercise 5.1.11 Let $d(x, y)$ denote the length of the shortest arc connecting x and y. Without loss of generality, we assume $x = Re_1$ and $y = R(e_1 \cos\theta + e_2 \sin\theta)$ with $R > 0$ and $0 \le \theta \le \pi$. Then we have $\|x - y\| = 2R\sin\frac{\theta}{2}$ and $d(x, y) = R\theta$. It follows that $\|x - y\| \le d(x, y) \le \pi/2\|x - y\|$.

Exercise 5.1.13 It suffices to prove the upper bound. Without loss of generality, we assume that $\mathbb{E}\, Z = 0$ and $K = \|Z\|_{\psi_2}$. For any $t \ge 0$, we have $\mathbb{P}\,(Z \ge t) \le 2\exp(-c_1 t^2/K^2)$ for some $c_1 > 0$. Setting $t = K\sqrt{\log 4/c_1}$ yields $\mathbb{P}\,(Z \ge K\sqrt{\log 4/c_1}) \le 1/2$, which implies $M \le K\sqrt{\log 4/c_1}$. Similarly, we can prove $M \ge -K\sqrt{\log 4/c_1}$. Thus $\|Z - M\|_{\psi_2} \le \|Z\|_{\psi_2} + \|M\|_{\psi_2} \le CK$ for some $C > 0$.

Exercise 5.1.15 Let $X \sim \mathrm{Unif}(S^{n-1})$. For any fixed $x_1, x_2, \ldots, x_k \in S^{n-1}$, Exercise 5.1.12 implies

$$\mathbb{P}\,(\exists i : |\langle X, x_i\rangle| \ge \epsilon) \le 2k\exp(-cn\epsilon^2)$$

for some $c > 0$. There exists $x_{k+1} \in S^{n-1}$ such that $|\langle x_{k+1}, x_i\rangle| \le \epsilon$ for all $1 \le i \le k$ as long as $k < \exp(cn\epsilon^2)/2$. By induction, we can construct $\{x_1, x_2, \cdots, x_N\}$ such that $|\langle x_i, x_j\rangle| \le \epsilon$ for all $i \ne j$ with $N = \lceil \exp(cn\epsilon^2)/2 - 1\rceil \ge \exp(cn\epsilon^2/2)$ for sufficiently large n.

Exercise 5.2.3 We only prove the blow-up inequality. Let σ denote the Gaussian measure on $\mathbb{R}^n$. Define $H = \{x \in \mathbb{R}^n : x_1 \le 0\}$. If $A \subset \mathbb{R}^n$ such that $\sigma(A) \ge 1/2$, Theorem 5.2.1 implies that $\sigma(A_t) \ge \sigma(H_t)$ for any $\epsilon \ge 0$. Note that $H_t = \{x \in \mathbb{R}^n : x_1 \le t\}$. Then $\sigma(H_t) = 1 - \mathbb{P}\,(g > t)$ where $g \sim N(0, 1)$. By Proposition 2.1.2, we have

$$\mathbb{P}\,(g > t) \le \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \le e^{-t^2/2}, & \text{for } t \ge 1, \\ \frac{1}{2} \le e^{-t^2/2}, & \text{for } 0 < t < 1. \end{cases}$$

It follows that $\sigma(A_t) \ge 1 - \exp(-t^2/2)$.

Exercise 5.2.4 Denote $K = \|f(X) - \mathbb{E}\, f(X)\|_{\psi_2}$. Since f is non-negative, we have

$$
\begin{aligned}
\|f(X) - \|f(X)\|_p\|_{\psi_2} &\le \|f(X) - \mathbb{E}\, f(X)\|_{\psi_2} + \|\|\mathbb{E}\, f(X)\|_p - \|f(X)\|_p\|_{\psi_2} \\
&\le K + \|\|f(X) - \mathbb{E}\, f(x)\|_p\|_{\psi_2} \\
&\le K + C_1\sqrt{p}\, K \\
&\le C_2(1+\sqrt{p})\|f\|_{\mathrm{Lip}}
\end{aligned}
$$

for some $C_1, C_2 > 0$.

Exercise 5.2.14 Let $X \sim N(0, I_n)$. By rotation variance, it suffices to define $\phi(X) = \frac{X}{\|X\|} g(\|X\|)$ where $g : \mathbb{R}_+ \to \mathbb{R}_+$. Since $\|X\|^2 \sim \chi^2_n$, we take $g(r) = \sqrt{n}\big[F_n\big(r^2\big)\big]^{1/n}$ where F_n is the c.d.f of χ^2_n. To check the Lipschitz continuity, we first prove the following lemma.

Lemma. *Suppose $x, y \in \mathbb{R}^n$ and $\|x\| \ge \|y\|$. Then we have $\|x - \tilde{x}\| + \|\tilde{x} - y\| \le \sqrt{2}\|x - y\|$percolation where $\tilde{x} = \|y\| x / \|x\|$.*

Proof. Let $\pi(x)$ denote the projection onto $\|y\|B^n_2$. Then $\tilde{x} = \pi(x)$ and $y \in \|y\|B^n_2$. Since $\|y\|B^n_2$ is a closed convex set, we have that for any $x \in \mathbb{R}^n$ and $z \in \|y\|B^n_2$, $\langle x - \pi(x), \pi(x) - z\rangle \ge 0$. It follows that

$$
\begin{aligned}
\|x - y\|^2 &= \|x - \pi(x)\|^2 + \|\pi(x) - y\|^2 + 2\langle x - \pi(x), \pi(x) - y\rangle \\
&\ge \|x - \pi(x)\|^2 + \|\pi(x) - y\|^2 \\
&\ge \frac{(\|x - \pi(x)\| + \|\pi(x) - y\|)^2}{2}
\end{aligned}
$$

$\square$

Note that $g(r)$ is an increasing function and $g(0) = 0$. Then we have

$$
\begin{aligned}
\|\phi\|_{\mathrm{Lip}} &\le \sqrt{2}\max\left\{\sup_{r\in\mathbb{R}_+} g'(r), \sup_{\|x\|=\|y\|, \|x-y\|>0} \frac{\|\phi(x) - \phi(y)\|}{\|x - y\|}\right\} \\
&\le \sqrt{2}\max\left\{\sup_{r\in\mathbb{R}_+} g'(r), \sup_{r\in\mathbb{R}_+} \frac{g(r)}{r}\right\} \\
&\le \sqrt{2}\sup_{r\in\mathbb{R}_+} g'(r).
\end{aligned}
$$

Let f_n denote the p.d.f. of χ_n^2. We have $g'(r) = 2n^{-1/2} r f_n(r^2) \big[F_n(r^2)\big]^{1/n-1}$. Note that $F_n(x) = \gamma(n/2, x/2)/\Gamma(n/2)$ where the lower incomplete gamma function $\gamma(s,x)$ is defined as $\gamma(s,x) = \int_0^s t^{s-1} e^{-t} dt$. Then

$$g'(r) = \frac{n^{-1/2} r^{n-1} e^{-r^2/2}}{[\Gamma(n/2)]^{1/n} 2^{n/2-1}} \big[\gamma(n/2, r^2/2)\big]^{1/n-1}$$

Edward Neuman has proved [5]

$$\gamma(a,x) \geq \frac{x^a}{a} \exp\biggl(\frac{-ax}{a+1}\biggr).$$

It follows that

$$\begin{aligned}
g'(r) &\leq \frac{n^{-1/2} r^{n-1} e^{-r^2/2}}{[\Gamma(n/2)]^{1/n} 2^{n/2-1}} \left[\frac{r^n}{2^{n/2-1} n} \exp\biggl(\frac{-nr^2}{2(n+2)}\biggr)\right]^{1/n-1} \\
&= \frac{\sqrt{n} 2^{1/n-1/2}}{n^{1/n} [\Gamma(n/2)]^{1/n}} \exp\biggl(-\frac{3r^2}{2(n+2)}\biggr) \\
&\leq \frac{\sqrt{2n}}{[\Gamma(n/2)]^{1/n}}.
\end{aligned}$$

Stirling's approximation implies $\sup_{r \in \mathbb{R}_+} g'(r) \leq C$ for some constant $C > 0$ independent of n. Thus ϕ is Lipschitz continuous.

Exercise 5.3.3 Assume that $A = (a_1, a_2, \ldots, a_m)^\top$, $\max_{1 \leq i \leq m} \|a_i\|_{\psi_2} \leq K$ and $K \geq 1$. It suffices to prove that for any $\|x\| = 1$, and $t > 0$

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\|Ax\| - 1\right| \geq t\right) \leq 2\exp\biggl(-\frac{cmt^2}{K^2}\biggr)$$

for some $c > 0$. Note that $\mathbb{E}\,(a_i^\top x)^2 = n$. Denote $s = \max(t, t^2)$. Recalling (3.2) and Corollary 2.8.3, we have

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\|Ax\| - 1\right| \geq t\right) &\leq \mathbb{P}\left(\left|\frac{1}{n}\|Ax\|^2 - 1\right| \geq s\right) \\
&\leq 2\exp\biggl(-c\min\biggl(\frac{s^2}{K^4}, \frac{s}{K^2}\biggr)m\biggr) \\
&\leq 2\exp\biggl(-\frac{cmt^2}{K^4}\biggr).
\end{aligned}$$

Exercise 5.3.4 Let $x_1, x_2, \ldots, x_n$ be an orthonormal basis of $\mathbb{R}^n$ and P be the orthogonal projection onto an m-dimensional subspace of $\mathbb{R}^n$. Denote $Q = \frac{n}{m}P$. Suppose that $\|Qx_i - Qx_j\| \geq (1-\epsilon)\|x_i - x_j\| = \sqrt{2}(1-\epsilon)$ for any $i \neq j$. For any $\epsilon \leq 1 - 1/\sqrt{2}$, $\{Qx_i\}_{i=1}^n$ is 1-separated. By Proposition 4.2.12 and Corollary 4.2.13, we have $n \leq 3^m$. Since $N = n$, $m \ll \log N$ can not hold.

Exercise 5.4.11 Denote $\sigma^2 = \left\|\sum_{i=1}^N \mathbb{E}\, X_i^2\right\|$ and $Y = \left\|\sum_{i=1}^N X_i\right\|$. Setting $u = -\log n + \frac{t^2/2}{\sigma^2 + Kt/3}$ yields

$$t = \frac{K(\log n + u)}{3} + \sqrt{\frac{K^2(\log n + u)^2}{9} + 2\sigma^2(\log n + u)} \leq \frac{2K(\log n + u)}{3} + \sqrt{2}\sigma\sqrt{\log n + u}.$$

Then we have

$$\mathbb{P}\left(Y \geq C\sigma\sqrt{\log n + u} + CK(\log n + u)\right) \leq 2e^{-u}$$

with $C = \sqrt{2}$. By Lemma 1.2.1 and change of variables,

$$\begin{aligned}
\mathbb{E}\, Y &\leq C\sigma\sqrt{\log n} + CK\log n + \int_{C\sigma\sqrt{\log n} + CK\log n}^{\infty} \mathbb{P}\,(Y \geq v) dv \\
&= C\sigma\sqrt{\log n} + CK\log n + 2C\int_0^\infty \left(K + \frac{\sigma}{2\sqrt{\log n + u}}\right)e^{-u}du \\
&= C\sigma\sqrt{\log n} + CK\log n + 2C\int_0^\infty \left(K + \frac{\sigma}{2\sqrt{u}}\right)e^{-u}du \\
&\leq \tilde{C}\left(\sigma\sqrt{1 + \log n} + K(1 + \log n)\right)
\end{aligned}$$

for some constant $\tilde{C} > 0$.

Exercise 5.4.13 Denote $\sigma^2 = \left\|\sum_{i=1}^N A_i^2\right\|$ and $Y = \left\|\sum_{i=1}^N \epsilon_i A_i\right\|$. Setting $u = -\log n + t^2/2\sigma^2$ yields $t = \sqrt{2}\sigma\sqrt{\log n + u}$. Then we have

$$\mathbb{P}\left(Y \geq \sqrt{2}\sigma\sqrt{\log n + u}\right) \leq 2e^{-u}.$$

By Lemma 1.2.1 and change of variables,

$$
\begin{aligned}
\mathbb{E}\, Y^p &\le \left(\sqrt{2}\sigma\sqrt{\log n}\right)^p + \int_{\sqrt{2}\sigma\sqrt{\log n}}^{\infty} pt^{p-1}\mathbb{P}\,(Y \ge t)dt \\
&\le \left(\sqrt{2}\sigma\sqrt{\log n}\right)^p + 2p\int_0^{\infty}\left(\sqrt{2}\sigma\sqrt{\log n + u}\right)^{p-1}\frac{\sqrt{2}\sigma}{2\sqrt{\log n + u}}e^{-u}du \\
&= \left(\sqrt{2}\sigma\sqrt{\log n}\right)^p + p\left(\sqrt{2}\sigma\right)^p\int_0^{\infty}(\log n + u)^{p/2-1}e^{-u}du \\
&\le \left(\sqrt{2}\sigma\sqrt{\log n}\right)^p + p\left(\sqrt{2}\sigma\right)^p\int_0^{\infty}\max\{1, 2^{p2-2}\}\left((\log n)^{p/2-1} + u^{p/2-1}\right)e^{-u}du \\
&\le C^p\sigma^p(p + \log n)^{p/2}
\end{aligned}
$$

for some constant $C > 0$.

Exercise 5.4.14 (b) This proof is inspired by Exercise 3.18 of [1] and [8]. Note that S is positive definite and $S_{ii} \sim B(N, 1/n)$. Thus $\|S\| = \max_{1\le i\le n} S_{ii}$. Suppose that $an \le N \le bn$ for some $0 < a < b$. We first give an upper bound of $\mathbb{E}\max_{1\le i\le n} S_{ii}$. By Jensen's inequality,

$$
\exp\left(\lambda\mathbb{E}\max_{1\le i\le n} S_{ii}\right) \le \mathbb{E}\,\exp\left(\lambda\max_{1\le i\le n} S_{ii}\right) \le \mathbb{E}\max_{1\le i\le n}\exp(\lambda S_{ii}) \le \sum_{i=1}^{n}\mathbb{E}\,\exp(\lambda S_{ii})
$$

for any $\lambda > 0$. Since $\mathbb{E}\,\exp(\lambda S_{ii}) = \left(e^{\lambda}/n + 1 - 1/n\right)^N \le \exp\left(N(e^{\lambda} - 1)/n\right) \le \exp\left(b(e^{\lambda} - 1)\right)$, it follows that

$$
\mathbb{E}\,\exp(\lambda S_{ii}) \le \frac{\log n + b(e^{\lambda} - 1)}{\lambda}.
$$

Let $W(x)$ denote the solution to $W(x)e^{W(x)} = x$ for $x > 0$. The upper bound is minimized for $\lambda = 1 + W\left(\frac{\log n - b}{eb}\right)$, which yields

$$
\mathbb{E}\,\|S\| \le \frac{\log n - b}{\lambda - 1} = \frac{\log n - b}{W\left(\frac{\log n - b}{eb}\right)}.
$$

Note that for $x \ge e$, $W(x) \ge \log x - \log\log x$. When n is sufficiently large, we have

$$
\mathbb{E}\,\|S\| \le \frac{\log n - b}{\log\left(\frac{\log n - b}{eb}\right) - \log\log\left(\frac{\log n - b}{eb}\right)} \le \frac{2\log n}{\log\log n}.
$$

The proof of the lower bound is based on the following lemma.

Lemma. *Let $X_1, X_2, \ldots, X_n$ be Bernoulli random variables and $X = \sum_{i=1}^n X_i$. If (a) $\mathbb{E}\, X_i = \mathbb{E}\, X_1$ for any $1 \le i \le n$, (b) $\max_{1\le i<j\le n} \mathbb{E}\, X_i X_j \le (1+o(1))(\mathbb{E}\, X_1)^2$ and (c) $n\mathbb{E}\, X_1 \to \infty$, then $\mathbb{P}\,(X=0) = o(1)$.*

Proof. By Chebyshev's inequality,

$$\mathbb{P}\,(X=0) \le \mathbb{P}\,(|X - \mathbb{E}\, X| \ge \mathbb{E}\, X) \le \frac{\mathrm{Var}\,(X)}{(\mathbb{E}\, X)^2} = \frac{\mathbb{E}\, X^2}{(\mathbb{E}\, X)^2} - 1.$$

Since

$$\mathbb{E}\, X^2 = \sum_{i=1}^n \mathbb{E}\, X_i^2 + 2\sum_{i=1}^n \sum_{j=i+1}^n \mathbb{E}\, X_i X_j = n\mathbb{E}\, X_1 + n(n-1)(\mathbb{E}\, X_1)^2 \le \mathbb{E}\, X + (\mathbb{E}\, X)^2,$$

it follows that

$$\frac{\mathbb{E}\, X^2}{(\mathbb{E}\, X)^2} \le \frac{1}{\mathbb{E}\, X} + 1 = 1 + o(1).$$

Thus $\mathbb{P}\,(X=0) = o(1)$. $\square$

Now we establish the lower bound. Let $k = \frac{\eta \log n}{\log\log n}$ for some $0 < \eta < 1$ and $X_i = \mathbf{1}_{\{S_{ii} \ge k\}}$. Clearly. $\mathbb{E}\, X_i = \mathbb{E}\, X_1$. Moreover,

$$\begin{aligned}
n\mathbb{E}\, X_1 &\ge n\mathbb{P}\,(S_{11} = k) \\
&= n\binom{N}{k}\left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{N-k} \\
&\ge \frac{n}{k!}\frac{(an-k)^k}{n^k}\left(1 - \frac{1}{n}\right)^{bn-k} \\
&= \frac{ne^k}{k^k\sqrt{2\pi k}} a^k e^{-b}(1+o(1)) \\
&= \exp\left(\log n + k - k\log k + k\log a - \frac{1}{2}\log(2\pi k) - b\right)(1+o(1)) \\
&= \exp((1-\eta)\log n(1+o(1))).
\end{aligned}$$

Then condition (c) is satisfied. Finally, we verify the condition (b). For any $1 \le i < j \le n$, we have

$$
\begin{aligned}
\mathbb{E}\, X_i X_j &= \mathbb{P}\left(S_{ii} \ge k, S_{jj} \ge k\right) \\
&= \sum_{l=k}^{N-k} \sum_{m=k}^{N-l} \binom{N}{l} \binom{N-l}{m} \left(\frac{1}{n}\right)^{l+m} \left(1-\frac{2}{n}\right)^{N-l-m} \\
&\le \sum_{l=k}^{N} \sum_{m=k}^{N} \binom{N}{l} \binom{N}{m} \left(\frac{1}{n}\right)^{l+m} \left(1-\frac{1}{n}\right)^{2N-2l-2m} \\
&= \left[\sum_{l=k}^{N} \binom{N}{l} \left(\frac{1}{n}\right)^{l} \left(1-\frac{1}{n}\right)^{N-2l}\right]^2
\end{aligned}
$$

Note that for $k \le l < N$,

$$
\frac{\binom{N}{l+1}\left(\frac{1}{n}\right)^{l+1}\left(1-\frac{1}{n}\right)^{N-2l-2}}{\binom{N}{l}\left(\frac{1}{n}\right)^{l}\left(1-\frac{1}{n}\right)^{N-2l}} = \frac{(N-l)n}{(l+1)(n-1)^2} \le \frac{(bn-k)n}{(k+1)(n-1)^2}.
$$

Denote $\lambda \triangleq \frac{(bn-k)n}{(k+1)(n-1)^2}$. Since $\lambda = o(1)$, it follows that

$$
\begin{aligned}
\mathbb{E}\, X_i X_j &\le \left[\binom{N}{k}\left(\frac{1}{n}\right)^{k}\left(1-\frac{1}{n}\right)^{N-2k} \frac{1}{1-\lambda}\right]^2 \\
&\le (\mathbb{E}\, X_1)^2 \left(1-\frac{1}{n}\right)^{-2k} \frac{1}{(1-\lambda)^2} \\
&= (1+o(1))(\mathbb{E}\, X_1)^2
\end{aligned}
$$

The lemma implies

$$
\mathbb{E}\, \|S\| = \mathbb{E} \max_{1\le i\le n} S_{ii} \ge k\, \mathbb{P}\left(\max_{1\le i\le n} S_{ii} \ge k\right) = k\, \mathbb{P}\left(\sum_{i=1}^{n} X_i > 0\right) \ge \frac{\eta \log n}{2\log\log n}
$$

when n is sufficiently large. Therefore, we have $\mathbb{E}\, \|S\| \asymp \frac{\log n}{\log\log n}$.

Exercise 5.6.5 This argument is inspired by Remark 5.42 of [10]. Suppose Y is an isotropic random vector in $\mathbb{R}^n$ with an arbitrary distribution and ξ is a $\{0,1\}$-valued random variable independent of Y with $\mathbb{E}\,\xi = \delta$. Let X_i be independent copies of $\delta^{-1/2}\xi Y$. Then $\Sigma = \mathbb{E}\, X_i X_i^\top = I_n$, but $\Sigma_m = 0$ with probability at least $(1-\delta)^m$. For any m, we can find δ such that $(1-\delta)^m \ge 0.9$.

Exercise 5.6.7 Let $X \sim \mathrm{Unif}(\sqrt{n}e_i : i = 1, 2, \dots, n)$ where $\{e_i\}_{i=1}^n$ is the canonical basis of $\mathbb{R}^n$. Then $\Sigma = \mathbb{E}\, XX^\top = I_n$. Suppose that we sequentially throw m balls into n bins by placing each ball into a bin chosen independently and uniformly at random. $\|\Sigma_m\|/n$ is the maximum number of balls in any bin. If $m \ll n \log n$, Theorem 1 of [8] implies that

$$\mathbb{E}\,\|\Sigma_m - \Sigma\| \geq \mathbb{E}\,\|\Sigma_m\| - 1 \geq \frac{n \log n}{2m \log \frac{n \log n}{m}} - 1$$

for sufficiently large n. Thus $\|\Sigma_m - \Sigma\| \leq \epsilon \|\Sigma\|$ can not hold.

Exercise 5.6.8 We prove that for every $t \geq t_0$, (5.22) holds with probability at least $1 - 2n^{-ct^2}$, where t_0 and c are absolute constants. Without loss of generality, we assume $K \geq 1$. Let $X_i = \frac{1}{m}\big(A_i A_i^\top - I_n\big)$. Just like in the proof of Theorem 5.6.1, we can obtain $\|X_i\| \leq \frac{2K^2 n}{m}$ and $\|\mathbb{E} \sum_{i=1}^m X_i^2\| \leq \frac{K^2 n}{m}$. Denote $\delta = Kt\sqrt{\frac{n \log n}{m}}$ and $\epsilon = \max(\delta, \delta^2)$. By matrix Bernstein's inequality, we have

$$\mathbb{P}\left(\left\|\sum_{i=1}^m X_i\right\| \geq \epsilon\right) \leq 2n \exp\left[-\frac{c'm}{K^2 n} \min(\epsilon, \epsilon^2)\right] \leq 2n \exp\left(-\frac{c'm}{K^2 n}\delta^2\right) = 2n^{1-c't^2}$$

for some $c' > 0$. Since $\sum_{i=1}^m X_i = \frac{1}{m}A^\top A - I_n$, Lemma 4.1.5 implies that (5.22) holds with probability at least $1 - 2n^{1-c't^2}$. Then we can find t_0 and c such that $1 - c't^2 \leq ct^2$ for every $t \geq t_0$.

6 Quadratic forms, symmertrization, and contraction

Exercise 6.1.5

Lemma. *Let Y and Z be independent random vectors s.t. $\mathbb{E}\, Z = 0$. Then for every convex and increasing function F, one has*

$$\mathbb{E}\, F(\|Y\|) \leq \mathbb{E}\, F(\|Y + Z\|)$$

Proof. We have the fact that for every convex and increasing function F, $F(\|\cdot\|)$ is also convex.

Then it is straightforward to see:

$$
\begin{aligned}
\mathbb{E}_Y F(\|Y\|) &= \mathbb{E}_Y F(\|\mathbb{E}_Z(Y+Z)\|) \\
&\leq \mathbb{E}_Y \mathbb{E}_Z F(\|Y+Z\|) \qquad \text{(Jensen's inequality)} \\
&= \mathbb{E}_{Y,Z} F(\|Y+Z\|) \qquad \text{(independence)}
\end{aligned}
$$

□

With the lemma above, we may then complete the proof via similar arguments as in the proof of Theorem 6.1.1. Specifically, we have

$$
\begin{aligned}
&\mathbb{E}_X F\left(\left\|\sum_{i\neq j} X_i X_j u_{ij}\right\|\right) \\
&= \mathbb{E}_X F\left(\left\|\mathbb{E}_\delta \sum_{i\neq j} 4\delta_i(1-\delta_j) X_i X_j u_{ij}\right\|\right) \\
&= \mathbb{E}_{X,X'} F\left(\left\|\mathbb{E}_I \sum_{i\in I, j\in I^c} 4 X_i X'_j u_{ij}\right\|\right) \\
&\leq \mathbb{E}_I \mathbb{E}_{X,X'} F\left(4\left\|\sum_{i\in I, j\in I^c} X_i X'_j u_{ij}\right\|\right) \qquad \text{(Jensen's inequality)}
\end{aligned}
$$

So there exists $I_0 \in \{1, ..., n\}$, such that

$$
\begin{aligned}
\mathbb{E}_X F\left(\left\|\sum_{i\neq j} X_i X_j u_{ij}\right\|\right) &\leq \mathbb{E}_{X,X'} F\left(4\left\|\sum_{i\in I_0, j\in I_0^c} X_i X'_j u_{ij}\right\|\right) \\
&\leq \mathbb{E}_{X,X'} F\left(4\left\|\sum_{i.j} X_i X'_j u_{ij}\right\|\right) \qquad \text{(by the lemma we just proved)}
\end{aligned}
$$

Exercise 6.2.5 Note that $\left\|\frac{A+A^\top}{2}\right\| \leq \|A\|$ holds for any norm. Without loss of generality, we may assume A is symmetric. We define $S = X^\top A X$ for ease of notation. Suppose $A = \sum_i s_i u_i u_i^\top$, we may get

$$
S = X^\top \left(\sum_i s_i u_i u_i^\top\right) X = \sum_i s_i (X^\top u_i)^2.
$$

Since $X \sim N(0, I_n)$, $X^\top u_i$ are independent standard normal random variables. Then using Proposition 2.7.1, we obtain

$$\begin{aligned}
\mathbb{E}\exp(\lambda(S - \mathbb{E}\,S)) &= \prod_i \mathbb{E}\exp(\lambda s_i((X^\top u_i)^2 - \mathbb{E}\,(X^\top u_i)^2)) \\
&\le \prod_i \exp(C\lambda^2 s_i^2) \\
&= \exp(C\lambda^2 \|A\|_F),
\end{aligned}$$

provided $\lambda^2 s_i^2 \le C,\ \forall 1 \le i \le n$. The remaining part is straightforward. Specifically,

$$\begin{aligned}
\mathbb{P}(|S - \mathbb{E}\,S| \ge t) &= 2\mathbb{P}(S - \mathbb{E}\,S \ge t) \\
&= 2\mathbb{P}(\exp(\lambda(S - \mathbb{E}\,S)) \ge \exp(\lambda t)) \\
&\le 2\mathbb{E}\exp(\lambda(S - \mathbb{E}\,S))/\exp(\lambda t) \\
&\le 2\exp(C\lambda^2\|A\|_F - \lambda t) \\
&\le 2\exp(-c\min\{t^2/\|A\|_F^2, t/\|A\|\}) \quad \text{(optimizing over } 0 \le \lambda \le c/\|A\|).
\end{aligned}$$

Exercise 6.2.6 If X and g are independent, then given X, $g^\top BX \sim \mathcal{N}(0, \|BX\|_2^2)$, and $\mathbb{E}_g(\exp(\mu g^\top BX)) = \exp(\mu^2\|BX\|_2^2/2)$. Now setting $\lambda = \mu/\sqrt{2}$ we have

$$\mathbb{E}_X \exp(\lambda^2\|BX\|_2^2) = \mathbb{E}_X\mathbb{E}_g(\exp(\sqrt{2}\lambda g^\top BX)) = \mathbb{E}_g\mathbb{E}_X(\exp(\sqrt{2}\lambda g^\top BX))$$

Since $\|X\|_{\psi_2} \le K$, we know that given g, $\|\sqrt{2}g^\top BX\|_{\psi_2} \le \sqrt{2}\|Bg\|_2 K$. Thus we obtain

$$\mathbb{E}_g\mathbb{E}_X(\exp(\sqrt{2}\lambda g^\top BX)) \le \mathbb{E}_g \exp(CK^2\lambda^2\|Bg\|_2^2).$$

Supposing $B = \sum_i s_i u_i v_i^\top$, we obtain

$$\begin{aligned}
\mathbb{E}_g \exp(CK^2\lambda^2\|Bg\|_2^2) &= \prod_i \mathbb{E}_g \exp(CK^2\lambda^2 s_i^2 (g^\top v_i)^2) \\
&\le \prod_i \exp(C'K^2\lambda^2 s_i^2) \\
&= \exp(C'\lambda^2\|B\|_F^2),
\end{aligned}$$

provided $|\lambda| \leq c/K\|B\|$.

Exercise 6.3.5 Recall that in Exercise 6.2.6 we prove that $\mathbb{E}\exp(\lambda^2\|BX\|_2^2) \leq \exp(C^2K^2\lambda^2\|B\|_F^2)$, provided $|\lambda| \leq c/K\|B\|$. It follows that

$$\begin{aligned}
\mathbb{P}(\|BX\|_2 \geq CK\|B\|_F + t) &\leq \mathbb{P}(\|BX\|_2^2 \geq (CK\|B\|_F + t)^2) \\
&= \mathbb{P}(\exp(\lambda^2\|BX\|_2^2) \geq \exp(\lambda^2(CK\|B\|_F + t)^2)) \\
&\leq \mathbb{E}(\exp(\lambda^2\|BX\|_2^2))/\exp(\lambda^2(CK\|B\|_F + t)^2) \\
&\leq \exp(\lambda^2C^2K^2\|B\|_F^2 - \lambda^2(CK\|B\|_F + t)^2) \\
&= \exp(-\lambda^2(2tCK\|B\|_F + t^2)) \\
&\leq \exp(-\lambda^2t^2)
\end{aligned}$$

Setting $\lambda = c/K\|B\|$ yields

$$\mathbb{P}(\|BX\|_2 \geq CK\|B\|_F + t) \leq \exp(-c^2t^2/K^2\|B\|^2).$$

Exercise 6.3.6 A simple example is that we flip a fair coin, if the result is head we set $X = 0$, or we sample X according to the uniform distribution on $\sqrt{2n}\mathbb{S}^{n-1}$.

Exercise 6.5.4 See Theorem 3.2 of [9].

Exercise 6.7.5 Suppose $X_i \equiv e_i \in \mathbb{R}^N$, then $\mathbb{E}\left\|\sum_i X_i\right\|_\infty = 1$, and $\mathbb{E}\left\|\sum_i g_iX_i\right\|_\infty = C\sqrt{\log N}$.

Exercise 6.7.7, 6.7.8 One may refer to the proof of Theorem 7 in Prof. John Duchi's Note *Probability Bounds*.

7 Random processes

Exercise 7.1.9 Following the proof of Lemma 6.4.2, we can get the upper bound. However, the lower bound does not hold. Let ξ be 1 with probability $1 - 1/n$ and $-(n-1)$ with probability $1/n$. Suppose that $N = 1$, $T = \{1, 2, \ldots, k\}$, and $X_1(t)$ are independent copies of ξ. Then we have

$\mathbb{E}\sup_{t\in T} X_1(t) \le 1$ and

$$\begin{aligned}\mathbb{E}\sup_{t\in T}\epsilon_1 X_1(t) &= (n-1)\left[1-\left(\frac{2n-1}{2n}\right)^k\right] + \left(\frac{2n-1}{2n}\right)^k - \frac{1}{2^k} - \left[\frac{1}{2^k} - \frac{1}{(2n)^k}\right] - \frac{n-1}{(2n)^k} \\ &\ge (n-1)\left[1-\left(\frac{2n-1}{2n}\right)^k - \frac{1}{(2n)^k}\right].\end{aligned}$$

For a fix n, $\mathbb{E}\sup_{t\in T}\epsilon_1 X_1(t) \ge \frac{n-1}{2}$ for a sufficiently large k.

Exercise 7.2.14 One can check that $\frac{\partial^2 f}{\partial x_{ij}\partial x_{il}}(x) \le 0$ and $\frac{\partial^2}{\partial x_{ij}\partial x_{kl}} f(x) \ge 0$ for $i \ne k$. Define $Z(u) = \sqrt{u}X + \sqrt{1-u}Y$ and $\Sigma_{ij,kl} = \mathrm{Cor}\,(X_{ij}, X_{kl})$. Following the proof of Lemma 7.2.7, we can obtain

$$\frac{d}{du}\mathbb{E}\, f(Z(u)) = \frac{1}{2}\sum_{i,j,k,l}\left(\Sigma^X_{ij,kl} - \Sigma^Y_{ij,kl}\right)\mathbb{E}\,\frac{\partial^2}{\partial x_{ij}\partial x_{kl}}(Z(u)) \le 0.$$

Following the proof of Theorem 7.2.9, we obtain the desired inequality.

Exercise 7.3.2

$$\begin{aligned}\|uv^\top - wz^\top\|_F^2 &= \sum_{i,j}(u_i v_j - w_i z_j)^2 \\ &= \sum_{i,j}[u_i(v_j - z_j) + (u_i - w_i)z_j]^2 \\ &= \sum_{i,j}\left[u_i^2(v_j - z_j)^2 + (u_i - w_i)^2 z_j^2 + 2u_i(u_i - w_i)z_j(v_j - z_j)\right] \\ &= \|u-w\|_2^2 + \|v-z\|_2^2 + 2(1-\langle u, w\rangle)(\langle z, v\rangle - 1) \\ &\le \|u-w\|_2^2 + \|v-z\|_2^2\end{aligned}$$

Exercise 7.3.4 Define $X_{uv} = \langle Au, v\rangle$ and $Y_{uv} = \langle g, u\rangle + \langle h, v\rangle$ where $g \sim N(0, I_n)$, $h \sim N(0, I_m)$ are independent. By Exercise 7.3.2, $\mathbb{E}\,(X_{uv} - X_{wz})^2 \le \mathbb{E}\,(Y_{uv} - Y_{wz})^2$ and $\mathbb{E}\,(X_{uv} - X_{uz})^2 = \mathbb{E}\,(Y_{uv} - Y_{uz})^2$. Gordon's inequality without the requirement of equal variance implies the first inequality. Following the proof of Corollary 7.3.3, we obtain the second inequality.

Exercise 7.3.5 Define $X_u = \langle Au, u\rangle$ and $Y_u = 2\langle g, u\rangle$ where $g \sim N(0, I_n)$. One can check that $\mathbb{E}\,(X_u - X_v)^2 = 4 - 4\langle u, v\rangle^2$ and $\mathbb{E}\,(Y_u - Y_v)^2 = 8 - 8\langle u, v\rangle$. It follows that $\mathbb{E}\,(X_u - X_v)^2 \le \mathbb{E}\,(Y_u - Y_v)^2$.

Following the proof of Corollary 7.3.3, we obtain the desired bounds.

Exercise 7.6.1 The lower bound follows from Jensen's inequality. To prove the upper bound, consider the function $f(g) = \sup_{x,y\in T}\langle g, x-y\rangle$. Without loss of generality, we assume $\mathrm{diam}(T) = 1$. For any $g, g' \in \mathbb{R}^n$, $f(g) \le \sup_{x,y\in T}\langle g - g', x - y\rangle + f(g') \le \|g - g'\| + f(g')$. By symmetry, $f(g)$ is 1-Lipschitz. Note that $\mathbb{E}\, f(g) = w(T - T) \le h(T - T)$. The upper bound follows from Theorem 5.2.2.

Exercise 7.6.9

$$\gamma(T) = \mathbb{E}\sup_{x\in T}|\langle g, x\rangle| \le \mathbb{E}\sup_{x\in T}|\langle g, x-y\rangle| + \mathbb{E}\,|\langle g, y\rangle| \le \gamma(T-T) + \frac{2}{\pi}\|y\|_2 \le 2\,[w(T) + \|y\|_2].$$

$$w(T) + \|y\|_2 = \mathbb{E}\sup_{x\in T}\langle g, x\rangle + \sqrt{\frac{\pi}{2}}\,\mathbb{E}\,|\langle g, y\rangle| \le 3\,\mathbb{E}\sup_{x\in T}|\langle g, x\rangle| = 3\gamma(T).$$

Exercise 7.7.4 Suppose that P is an orthogonal projection and T is a closed set. Then we can decompose P as $P = \sum_{i=1}^m u_i u_i^\top$ and find $x \in T - T$ such that $\|x\|_2 = \mathrm{diam}(T)$. It follows that

$$\mathbb{E}\,\mathrm{diam}(PT) = \mathbb{E}\sup_{x\in T-T}\|Px\| \ge \mathbb{E}\sup_{x\in T-T}|\langle v_1, x\rangle| = w_s(T-T) = 2w_s(T)$$

On the other hand, Lemma 5.3.2 implies that

$$\mathbb{E}\,\mathrm{diam}(PT) \ge \mathbb{E}\,\|Px\|_2 \ge c\sqrt{\frac{m}{n}}\|x\|_2 = c\sqrt{\frac{m}{n}}\mathrm{diam}(T)$$

for some $c > 0$. This completes the proof.

Exercise 7.7.5 Note that $\mathrm{diam}(P(AB_2^k)) = 2\|PA\|$, and $w_s(AB_2^k) = \mathbb{E}\sup_{\|x\|_2\le 1}\langle\theta, Ax\rangle = \mathbb{E}\,\|A^\top\theta\|_2 \le \sqrt{\mathbb{E}\,\mathrm{tr}(A^\top\theta\theta^\top A)} = \|A\|_F/\sqrt{n}$. Theorem 7.7.1 implies part (a). Similarly, we can obtain part (b).

References

[1] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence.* OUP Oxford, 2013.

[2] D. de Caen. An upper bound on the sum of squares of degrees in a graph. *Discrete Mathematics*, 185(1):245–248, 1998.

[3] R. Durrett. *Probability: Theory and Examples.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

[4] Gautam Kamath. Bounds on the expectation of the maximum of samples from a gaussian. 2015.

[5] Edward Neuman. Inequalities and bounds for the incomplete gamma function. *Results in Mathematics*, 63(3-4):1209–1214, 2013.

[6] S. Y. Novak. Poisson approximation. *arXiv: Probability*, 2018.

[7] Francesco Orabona and David Pal. Optimal non-asymptotic lower bound on the minimax regret of learning with expert advice. *arXiv preprint arXiv:1511.02176*, 2015.

[8] Martin Raab and Angelika Steger. "balls into bins"—a simple and tight analysis. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 159–170. Springer, 1998.

[9] YOAV SEGINER. The expected norm of random matrices. *Combinatorics, Probability and Computing*, 9(2):149–166, 2000.

[10] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.