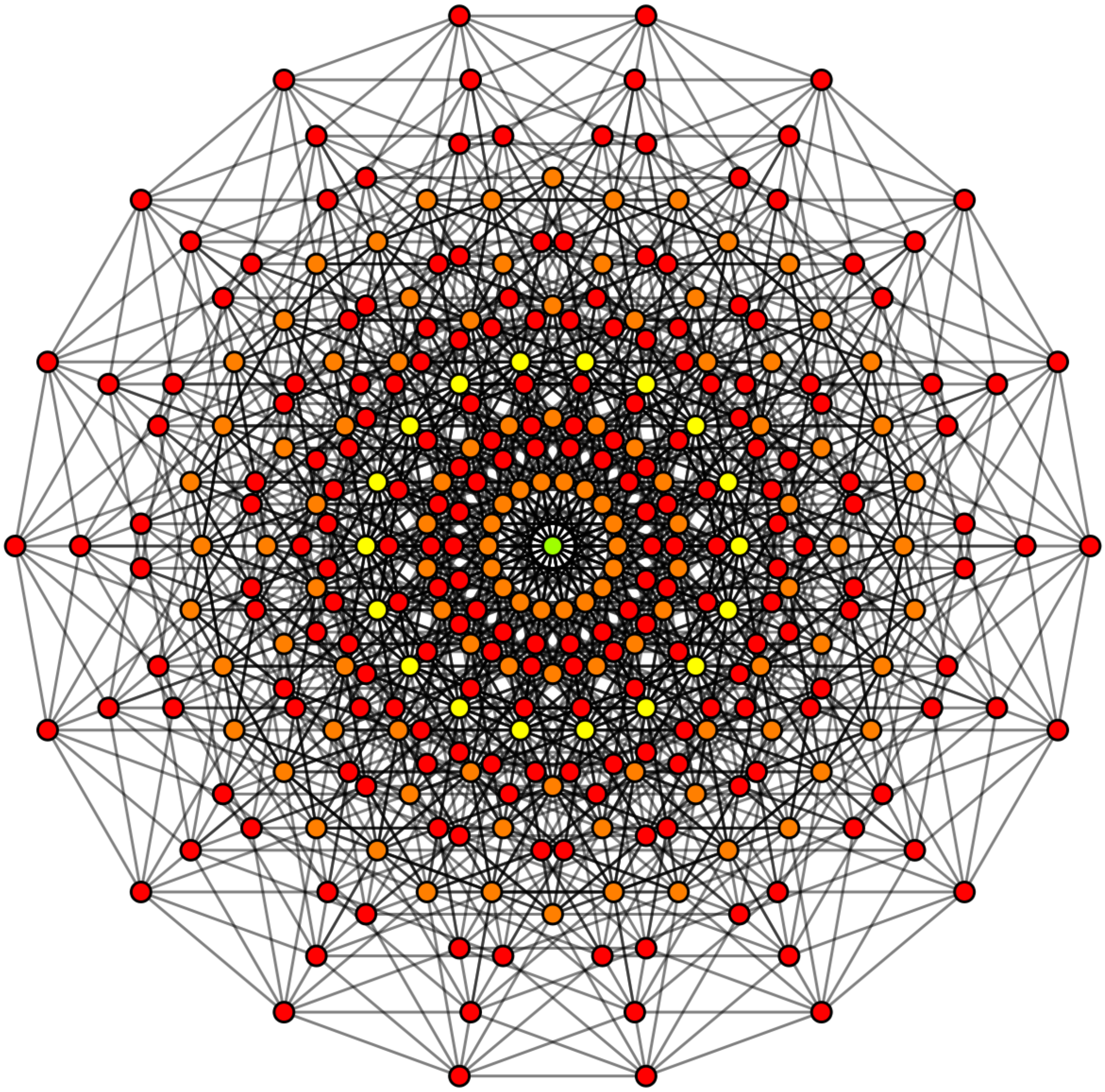
High-Dimensional Probability
Solution Manual

Pingbang Hu

June 7, 2024

Abstract

This is the solution I write for the reading group on Roman Vershynin's *High Dimensional Probability* [Ver24], where I serve as the lead. It may contain factual and/or typographic errors, and some exercises are omitted.



The reading group is held 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.

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

(a) Let $Z_1, \ldots, Z_k$ 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 $Z_1, \ldots, Z_k$ 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

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

⊛

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,

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

⊛

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

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

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.

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 $X_1, X_2, \dots$ 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.

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 $\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Then 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 2.2.7). Prove the Hoeffding's inequality for general bounded random variables, possibly with some absolute constant instead of 2 in the tail. For convenience, we restate the theorem: Let $X_1, \dots, X_N$ be independent random variables. Assume that $X_i \in [m_i, M_i]$ for every

i. Then, for any $t > 0$, we have

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

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

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

or equivalently,

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

⊛

Say given $X_i \in [m_i, M_i]$ for every i, then $X_i - \mathbb{E}\left[X_i\right] \in [m_i - \mathbb{E}\left[X_i\right], M_i - \mathbb{E}\left[X_i\right]]$ 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 Chebyshev's inequality,

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

By requiring $\sigma^2/(N\epsilon^2) \leq 1/4$, we're done, which is equivalent to requiring $N \geq 4\sigma^2/\epsilon^2 = O(\sigma^2/\epsilon^2)$.

(b) Consider we gather 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

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

as $\mathbb{E}[X_i] = 1/4$. From Hoeffding's inequality, the above probability is bounded above by

$\exp\left(-2(k/4)^2/k\right)$, 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 $X_1, \ldots, X_N$ 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 $t > 0$,

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

(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 $X \sim \text{Pois}(\lambda)$. 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 $X \sim \text{Pois}(\lambda)$, i.e., as $N \to \infty$, $\max_{i \leq N} p_{N,i} \to 0$ and $\lambda_N := \mathbb{E}\left[S_N\right] \to \lambda < \infty$, $S_N \to \text{Pois}(\lambda)$. 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.

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 $\ln(1-x) = -x - x^2/2 - x^3/3 - \ldots$. ⊛

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. 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

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

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 2.4.5). 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

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

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] Possible 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 4: Sub-Gaussian Random Variables

2.5 Sub-gaussian distributions

16 Feb.

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

$$
\begin{aligned}
(\mathbb{E}[|X|^p])^{1/p} &= \left(\int_{-\infty}^{\infty} |x|^p \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, \mathrm{d}x\right)^{1/p} \\
&= \left(2\int_{0}^{\infty} |x|^p \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, \mathrm{d}x\right)^{1/p} && \text{symmetric around } 0 \\
&= \left(2\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} u^{p/2} e^{-u/2} \frac{1}{2\sqrt{u}} \, \mathrm{d}u\right)^{1/p} && x^2 =: u \\
&= \left(\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} u^{(p-1)/2} e^{-u/2} \, \mathrm{d}u\right)^{1/p} \\
&= \left(\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} (2t)^{(p-1)/2} e^{-t} 2 \, \mathrm{d}t\right)^{1/p} && u/2 =: t \\
&= \left(\frac{1}{\sqrt{2\pi}} \cdot 2^{(p-1)/2} \cdot 2 \int_{0}^{\infty} t^{(p-1)/2} e^{-t} \, \mathrm{d}t\right)^{1/p} \\
&= \left(\frac{1}{\sqrt{2\pi}} 2^{(p+1)/2} \Gamma\left(\frac{p+1}{2}\right)\right)^{1/p} \\
&= \left(\frac{1}{\sqrt{2}} \sqrt{2}^{p+1} \frac{\Gamma((p+1)/2)}{\Gamma(1/2)}\right)^{1/p} && \Gamma(1/2) = \sqrt{\pi} \\
&= \sqrt{2} \left(\frac{\Gamma((p+1)/2)}{\Gamma(1/2)}\right)^{1/p},
\end{aligned}
$$

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 $z := p/2 \in \mathbb{N}$,

$$\begin{aligned}\Gamma((1+p)/2) = \frac{2^{1-p}\sqrt{\pi}\Gamma(p)}{\Gamma(p/2)} &= 2^{1-p}\sqrt{\pi}\frac{(p-1)!}{(p/2-1)!} \\ &= 2^{1-p}\sqrt{\pi}\frac{(p-1)!}{(1/2)^{p/2-1}(p-2)!!} = 2^{-p/2}\sqrt{\pi}(p-1)!!.\end{aligned}$$

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 $\mathbb{E}[\exp(\lambda X)] \leq \exp\left(K_5^2\lambda^2\right)$ for all $\lambda \in \mathbb{R}$, we see that from Jensen's inequality,

$$\exp(\mathbb{E}[\lambda X]) \leq \mathbb{E}[\exp(\lambda X)] \leq \exp\left(K_5^2\lambda^2\right),$$

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 $\lambda \mapsto \mathbb{E}[\exp\left(\lambda^2X^2\right)]$ is only finite in some bounded neighborhood of zero.

(b) Suppose that some random variable X satisfies $\mathbb{E}[\exp\left(\lambda^2X^2\right)] \leq \exp\left(K\lambda^2\right)$ 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

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

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,

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

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),

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

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

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

⊛

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 $Y_i \coloneqq |X_i|/K\sqrt{1 + \log i}$ (which is always positive) for all $i \geq 1$. Then we see that for all $t \geq 0$,

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

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

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

⊛

Problem (Exercise 2.5.11). Show that the bound in Exercise 2.5.10 is sharp. Let $X_1, X_2, \dots, X_N$ 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.,

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

so

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

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. ⊛

Bibliography

[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.