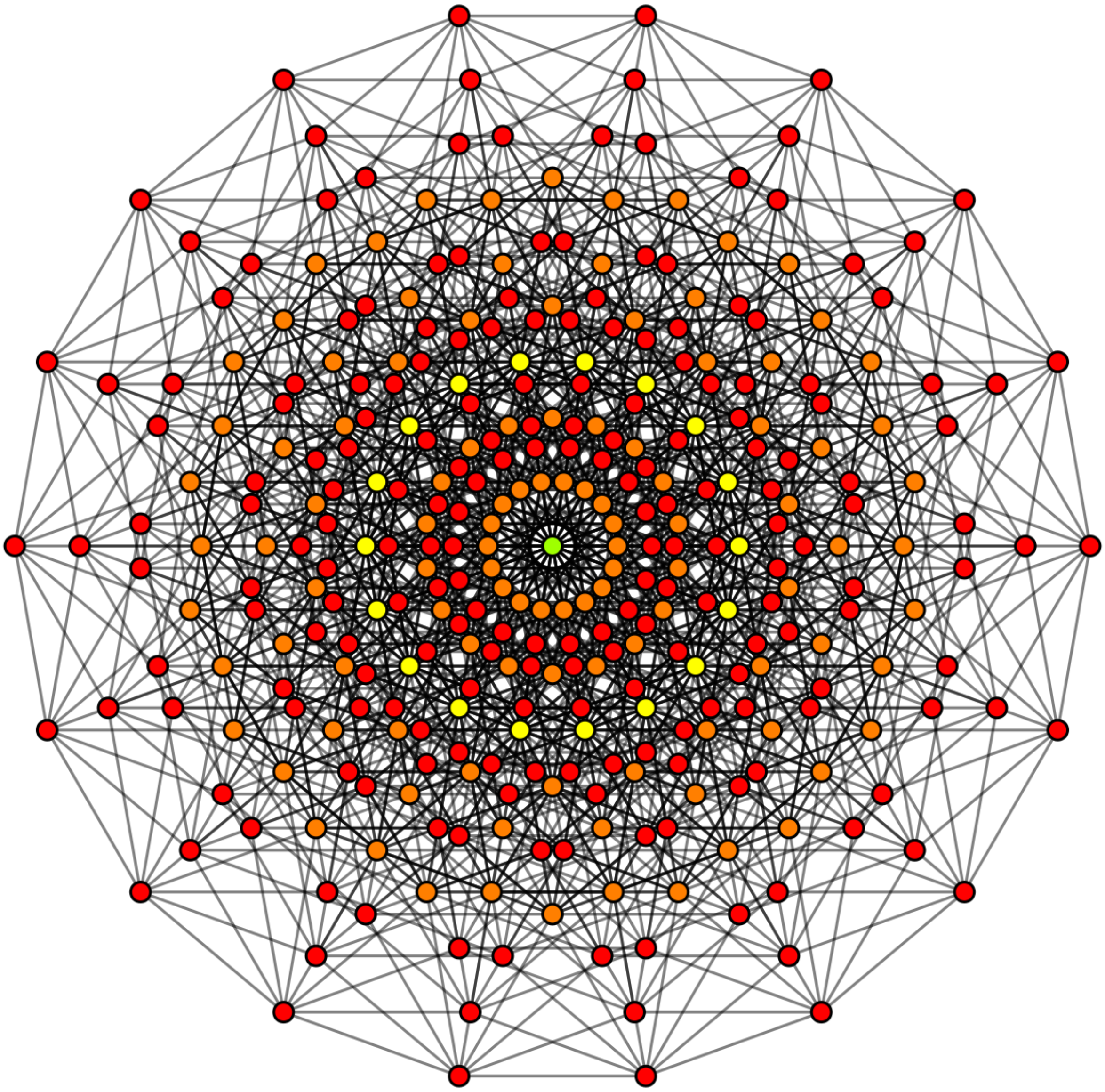
High-Dimensional Probability
Solution Manual

Pingbang Hu

June 8, 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. 2024

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

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

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

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

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

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

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

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

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

proving the result.

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

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

⊛

Problem (Exercise 0.0.5)**.** Prove the inequalities

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

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

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

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

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

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

⊛

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

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

suffice. Here C is a suitable absolute constant.

Answer. Omit. ⊛

Chapter 1

Preliminaries on random variables

1.1 Basic quantities associated with random variables

No Exercise!

1.2 Some classical inequalities

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

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

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

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

⊛

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. 2024

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

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

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

⊛

1.3 Limit theorems

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

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

Answer. We see that

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

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

Chapter 2

Concentration of sums of independent random variables

Week 3: More Powerful Concentration Inequalities

2.1 Why concentration inequalities?

2 Feb. 2024

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

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

Answer. Denote the standard normal density as

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

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

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

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

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

implying that

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

which gives the second inequality. ⊛

2.2 Hoeffding's inequality

Problem (Exercise 2.2.3). Show that

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

Answer. Omit. ⊛

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

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

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

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

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

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

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

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

Hence, we see that

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

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

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

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

Proof. Since

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

■

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

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

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

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

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

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

and

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

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

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

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

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

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

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

raising both sides by e shows the desired result. ⊛

Say given $X_i \in [m_i, M_i]$ for every i, then $X_i - \mathbb{E}\left[X_i\right] \in [m_i - \mathbb{E}\left[X_i\right], M_i - \mathbb{E}\left[X_i\right]]$ 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 $\mathbb{E}\left[X_i\right]=\frac{1}{2}+\delta$.

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

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

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

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

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

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

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

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

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

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

(b) Consider gathering k estimator from the above procedure, i.e., we now have $\hat{\mu}_1,\dots,\hat{\mu}_k$ such that each are an ϵ-accurate mean estimator with probability at least $3/4$. This requires $k\cdot 4\sigma^2/\epsilon^2=O(k\sigma^2/\epsilon^2)$ samples. We claim that the median $\hat{\mu}\coloneqq\mathrm{median}(\hat{\mu}_1,\dots,\hat{\mu}_k)$ is an ϵ-accurate mean estimator with probability at least $1-\delta$ for some k (depends on δ). Consider a series of indicators $X_i=\mathbb{1}_{\lvert\hat{\mu}_i-\mu\rvert>\epsilon}$, 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

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

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

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

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

⊛

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

(a) Show that the MGF of X_i satisfies

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

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

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

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

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

(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 $X \sim \operatorname{Pois}(\lambda)$. 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 $X \sim \operatorname{Pois}(\lambda)$, i.e., as $N \to \infty$, $\max_{i\leq N} p_{N,i} \to 0$ and $\lambda_N := \mathbb{E}\left[S_N\right] \to \lambda < \infty$, $S_N \to \operatorname{Pois}(\lambda)$. From Chernoff's inequality, for any $t > \lambda_N$,

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

We then see that

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

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

Week 4: Chernoff's Inequality and Degree Concentration

7 Feb. 2024

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

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

where $c > 0$ is an absolute constant.

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

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

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

Claim. For all $x > 0$,

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

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

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

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

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

⊛

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

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

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

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

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

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

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

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

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

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

Combining two tails, we then see that

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

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

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

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

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

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 $X \sim \mathrm{Pois}(\lambda)$, i.e., as $N \to \infty$, $\max_{i \leq N} p_{N,i} \to 0$ and $\lambda_N \coloneqq \mathbb{E}[S_N] \to \lambda < \infty$, $S_N \to \mathrm{Pois}(\lambda)$. From multiplicative form of Chernoff's inequality, for $t_N \coloneqq \delta\lambda_N$,

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

It then follows that from the Poisson limit theorem,

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

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

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

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

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

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

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

2.4 Application: degrees of random graphs

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

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

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

Hence, from union bound, we have

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

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

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

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

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

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

for some $C > 0$. This implies that

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

Now, considering $C = M$, we have

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

Taking logarithm, we observe that

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

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

Answer. Omit. ⊛

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

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

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

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

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

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

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

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 $\operatorname{Bin}(n/2, q)$ with $X = \sum_{i\in A} X_i$ and mean $nq/2$, variance $nq(1-q)/2$. From Chebyshev's inequality,

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

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

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

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

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

with overwhelming probability. ⊛

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

Week 5: Sub-Gaussian Random Variables

2.5 Sub-gaussian distributions

16 Feb. 2024

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

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

Deduce that

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

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

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

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

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

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

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

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

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

where we recall that

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

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

Lemma 2.5.1. We have that for $p \geq 1$,

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

Proof. Consider the Legendre duplication formula, i.e.,

$$\Gamma(z)\Gamma(z+1/2) = 2^{1-2z}\sqrt{\pi}\Gamma(2z).$$

We see that for p being even, $(1+p)/2 = p/2 + 1/2$, by letting $z \coloneqq p/2 \in \mathbb{N}$,

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

For odd p, recall the identity $\Gamma(z+1) = z\Gamma(z)$. We then have

$$\begin{aligned} \Gamma((1+p)/2) =& \frac{p-1}{2} \cdot \Gamma((p-1)/2) \\ =& \frac{(p-1)(p-3)}{2^2} \cdot \Gamma((p-3)/2) \\ &\vdots \\ =& \frac{(p-1)(p-3)\ldots(p-(p-2))}{2^{(p-1)/2}} \cdot \Gamma(1) && 2 = (p-(p-2)) \\ =& 2^{-(p-1)/2}(p-1)(p-3)\ldots(2) \\ =& 2^{-(p-1)/2}(p-1)!!. \end{aligned}$$

■

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

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

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

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

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 $t := \|X\|_{\psi_2} + \|Y\|_{\psi_2}$, the above implies

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

hence the triangle inequality is verified. ⊛

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

Answer. Omit. ⊛

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

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

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

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

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

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

as $K \coloneqq \max_i \|X_i\|_{\psi_2}^2$. 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

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

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 $C' \coloneqq \sqrt{2}C$,

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

which is exactly what we want. ⊛

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

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

Answer. Again, let's first write

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

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

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

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

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

so

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

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

Week 6: Hoeffding's and Khintchine's Inequalities

2.6 General Hoeffding's and Khintchine's inequalities

21 Feb. 2024

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

Answer. Omit. ⊛

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

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

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

Answer. From Jensen's inequality,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

implying the result. ■

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

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

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

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

implying

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

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

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

Observe that from Exercise 2.6.5:

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

hence

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

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

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

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

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

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

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

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

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

Hence, we have that

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

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

Week 7: Sub-Exponential Random Variables

2.7 Sub-exponential distributions

1 Mar. 2024

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

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

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

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

(a) The tails of X satisfy

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

(b) The moments of X satisfy

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

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

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

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

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

Claim. (a) $\Rightarrow$ (b)

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

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

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

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

as we desired. ⊛

Claim. (b) $\Rightarrow$ (c)

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

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

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

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

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

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

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

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

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

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

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

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

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

Claim. (c) ⇒ (d)

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

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

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

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

⊛

Claim. (d) ⇒ (a)

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

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

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

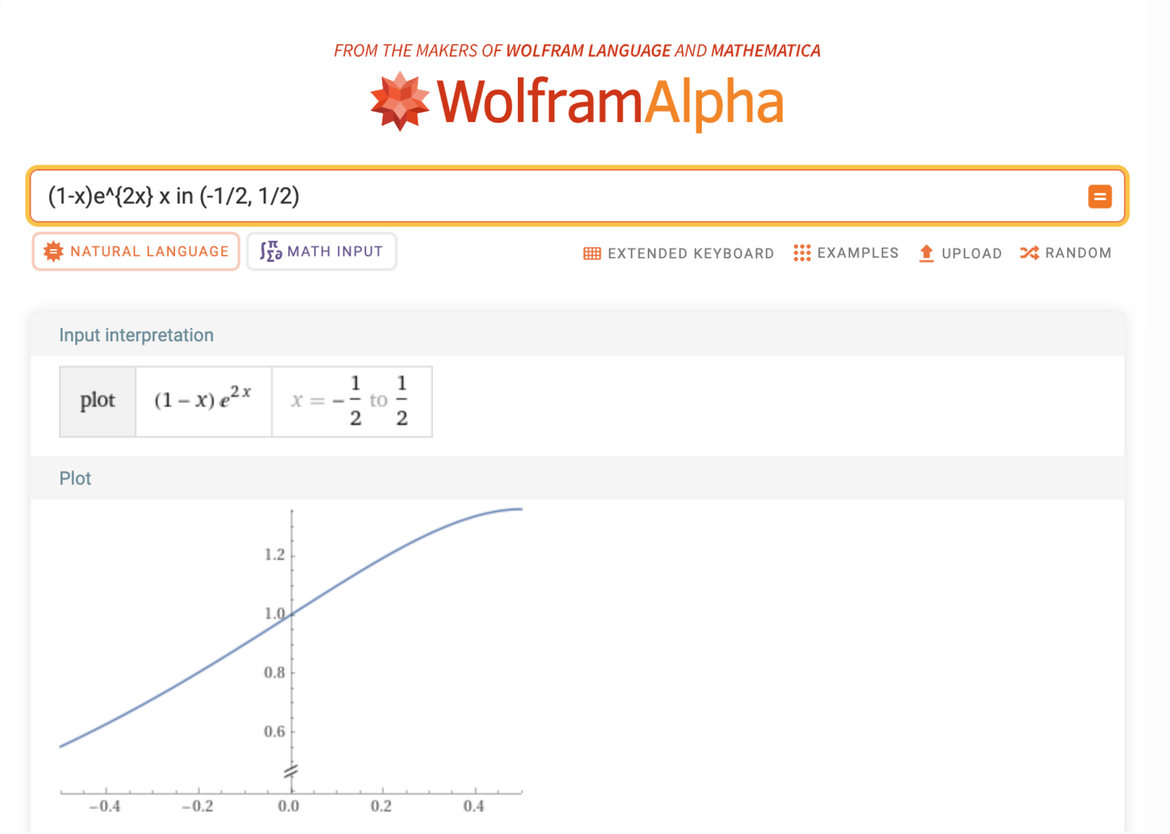
⊛

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

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

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

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



⊛

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

$$\|X - \mathbb{E}[X]\|_{\psi_1} \leq C\|X\|_{\psi_1}.$$

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

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

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

$$\|X\|_{L^1} \leq K_2 \cong \|X\|_{\psi_1}$$

since $K_i \cong \|X\|_{\psi_1} = K_4$. ⊛

Week 8: Bernstein's Inequality

6 Mar. 2024

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

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

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

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

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

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

2.8 Bernstein's inequality

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

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

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

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

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

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

Proof. From Taylor's expansion,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

⊛

Chapter 3

Random vectors in high dimensions

Week 9: Concentration Inequalities of Random Vectors

15 Mar. 2024

3.1 Concentration of the norm

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

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

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

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

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

from Theorem 3.1.1 and

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

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

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

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

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

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

Expanding the variance, we see that

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

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

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

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

⊛

Problem (Exercise 3.1.5). Deduce from Theorem 3.1.1 that

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

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

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

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

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

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

⊛

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

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

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

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

proving the result. ⊛

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

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

Answer. We want to bound

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

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

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

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

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

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

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

by letting $C := \sqrt{\pi}e/2$. ⊛

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

3.2 Covariance matrices and principal component analysis

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

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

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

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

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

is an isotropic, mean zero random vector.

Answer. (a) Firstly,

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

Moreover,

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

as Σ is positive-semidefinite.

(b) Similarly,

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

and moreover,

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

hence Z is also isotropic.

⊛

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

Check that

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

Answer. This directly follows from

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

⊛

Week 10: Common High-Dimensional Distributions

3.3 Examples of high-dimensional distributions

20 Mar. 2024

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

from (a).

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

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

as desired.

⊛

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

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

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

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

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

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

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

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

Answer. (a) It's because

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

from the fact that X is isotropic.

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

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

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

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

⊛

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

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

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

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

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

as u, v are unit orthogonal vectors. ⊛

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

$$g = r\theta$$

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

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

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

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

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

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

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

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

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

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

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

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

Such a separation of probability proves the independence.

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

⊛

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

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

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

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

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

- The left-hand side:

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

- The right-hand side:

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

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

Week 11: High-Dimensional Sub-Gaussian Distributions

3.4 Sub-gaussian distributions in higher dimensions

29 Mar. 2024

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

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

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

Answer.

1. We see that

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

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

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

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

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

⊛

Problem (Exercise 3.4.4). Show that

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

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

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

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

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

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

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

Hence, for any $t > 0$,

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

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

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

meaning that

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

⊛

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

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

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

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

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

is sub-gaussian, and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

⊛

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

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

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

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

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

With this construction, X is isotropic since

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

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

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

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

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

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

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

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

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

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

⊛

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.

[Vla+20] Mariia Vladimirova et al. "Sub-Weibull distributions: Generalizing sub-Gaussian and sub-Exponential properties to heavier tailed distributions". In: *Stat* 9.1 (Jan. 2020). ISSN: 2049-1573. DOI: 10.1002/sta4.318. URL: http://dx.doi.org/10.1002/sta4.318.