My Solutions to Exercises for High-Dimensional Probability Second Edition by Roman Vershynin

Gallant Tsao

July 28, 2025

Contents

0 Appetizers

Exercise 1

(a)

$$\begin{aligned}\mathbb{E}[\|Z-\mathbb{E}[Z]\|_2^2] &= \mathbb{E}[\|Z\|_2^2 - 2\langle Z, \mathbb{E}[Z]\rangle + \|\mathbb{E}[Z]\|_2^2] \\ &= \mathbb{E}[\|Z\|_2^2] - 2\mathbb{E}[Z]^T\mathbb{E}[Z] + \|\mathbb{E}[Z]\|_2^2 \\ &= \mathbb{E}[\|Z\|]_2^2 - \|\mathbb{E}[Z]\|_2^2.\end{aligned}$$

(b)

From part (a),

$$\begin{aligned}\mathbb{E}[\|Z-\mathbb{E}[Z]\|_2^2] &= \mathbb{E}[\|Z\|]_2^2 - \|\mathbb{E}[Z]\|_2^2 \\ &= \frac{1}{2}\mathbb{E}[\|Z\|_2^2] - \mathbb{E}[Z]^T\mathbb{E}[Z] + \frac{1}{2}\mathbb{E}[\|Z\|_2^2] \\ &= \frac{1}{2}(\mathbb{E}[\|Z\|_2^2] - 2\mathbb{E}[Z^t]\mathbb{E}[Z'] + \frac{1}{2}\mathbb{E}[\|Z'\|_2^2]) \\ &= \frac{1}{2}(\mathbb{E}[\|Z\|_2^2] - 2\mathbb{E}[Z^TZ'] + \mathbb{E}[\|Z'\|_2^2]) \\ &= \frac{1}{2}\mathbb{E}[\|Z-Z'\|_2^2].\end{aligned}$$

Exercise 2

Let $\mu = \mathbb{E}[Z]$. First, notice that

$$\begin{aligned}
\mathbb{E}[\|Z-a\|_2^2] - \mathbb{E}[\|Z-\mu\|_2^2] &= \mathbb{E}[\|Z\|_2^2 - 2a^T Z + \|a\|_2^2 - \|Z\|_2^2 + 2\mu^T Z - \|\mu\|_2^2] \\
&= \|a\|_2^2 - 2(a^T - \mu^T)\mathbb{E}[Z] - \|\mu\|_2^2 \\
&= \|a\|_2^2 - 2a^T\mu + 2\|\mu\|_2^2 - \|\mu\|_2^2 \\
&= \|a-\mu\|_2^2.
\end{aligned}$$

From above, minimizing $\mathbb{E}[\|Z-a\|_2^2]$ in terms of a is the same as minimizing the term we have above as the second term does not depend on a. The expression above is minimized exactly at $a^* = \mu = \mathbb{E}[Z]$ as the quantity is always greater than or equal to 0, and reaches the value 0 if and only if $a = \mu$.

Exercise 3

$$
\begin{aligned}
\mathbb{E}\left[\left\|\sum_{j=1}^{k} Z_j\right\|_2^2\right] &= \mathbb{E}[(Z_1 + \cdots + Z_k)^T (Z_1 + \cdots + Z_k)] \\
&= \mathbb{E}\left[\sum_{j=1}^{k} \|Z_j\|_2^2 + \sum_{i \neq j} Z_i^T Z_j\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{k} \|Z_j\|_2^2\right] + \sum_{i \neq j} \mathbb{E}[Z_i]^T \mathbb{E}[Z_j] \\
&= \mathbb{E}\left[\sum_{j=1}^{k} \|Z_j\|_2^2\right] + 0 \qquad (\mathbb{E}[Z_i] = 0) \\
&= \mathbb{E}\left[\sum_{j=1}^{k} \|Z_j\|_2^2\right].
\end{aligned}
$$

Exercise 4

(a)

We can consider these points as being chosen randomly at uniform from the unit ball in n dimensions, i.e.

$$X_1, \dots, X_n \sim_{iid} \text{Unif}(B_1^n) \implies \mathbb{E}[X_i] = 0.$$

Then by exercise 3,

$$\mathbb{E}\left[\left\|\sum_{i=1}^{k} X_i\right\|_2^2\right] = \sum_{i=1}^{k} \mathbb{E}[\|X_i\|_2^2] \leq k.$$

Therefore there exists a realization $(x_1, \dots, x_n)$ for which

$$\left\|\sum_{i=1}^{n} x_i\right\|_2^2 \leq k \implies \left\|\sum_{i=1}^{n} x_i\right\|_2 \leq \sqrt{k}.$$

(b)

We are bounding $\mathbb{E}[\|X_i\|_2^2]$ by 1, which is a tight bound.

Exercise 5

Exercise 6

The first inequality comes as follows: we can see that

$$\frac{n}{k} \le \frac{n-i}{k-i}, \quad i = 1, 2, \ldots, k-1.$$

This is because by cross multiplication

$$n(k-i) = nk - ni \ge nk - ki = k(n-i).$$

Then

$$\left(\frac{n}{k}\right)^k = \frac{n}{k} \times \frac{n}{k} \times \cdots \times \frac{n}{k} \le \frac{n}{k} \times \frac{n-1}{k-1} \times \cdots \times \frac{n-k+1}{1} = \binom{n}{k}.$$

The second inequality is trivial as $k \ge 1$. For the third inequality, we get

$$\begin{aligned}
\sum_{j=0}^{k} \binom{n}{j} \cdot \left(\frac{k}{n}\right)^k &\le \sum_{j=0}^{k} \binom{n}{j} \cdot \left(\frac{k}{n}\right)^j \quad (k/n \le 1) \\
&\le \sum_{j=0}^{n} \binom{n}{j} \cdot \left(\frac{k}{n}\right)^j \quad (k/n \le 1) \\
&= \left(1 + \frac{k}{n}\right)^n \quad \text{(Binomial Theorem)} \\
&< e^k.
\end{aligned}$$

Exercise 7

Assume n is large so that the $5/n$ radius near the surface is valid. The inner ball has radius $\frac{1}{2} - \frac{5}{n}$. Then the volume of the inner ball is $(\frac{1}{2} - \frac{5}{n})^n$ times the volums of the outer unit ball. In particular, as $n \to \infty$,

$$\left(\frac{1}{2} - \frac{5}{n}\right)^n = \left(\frac{1}{2}\right)^n \left(1 - \frac{10}{n}\right)^n \to 0.$$

This means that most of the points will be concentrated towards the surface of the n-dimensional ball.

Exercise 8

Let $X \sim \text{Unif}(B_1^n)$. Then the pdf of X is

$$f_X(x) = \frac{1}{\text{Vol}(B_1^n)}, \; x \in B_1^n.$$

Now let's consider the random variable $\|X\|_2$, i.e. the 2-norm of the random vector. Since the random vector is distributed uniformly in the n-dimensional ball, we can define its CDF as a function of the radius r:

$$F_{\|X\|_2}(r) = P(\|X\|_2 \leq r) = r^n.$$

Correspondingly, we can find the PDF by just taking the derivative of the CDF:

$$f_{\|X\|_2}(r) = nr^{n-1}, \; 0 \leq r \leq 1.$$

Then we can directly get that

$$\mathbb{E}[\|X\|_2] = \int_0^1 r \cdot nr^{n-1} \; dr = n \cdot \left[\frac{r^{n+1}}{n+1}\right]_0^1 = \frac{n}{n+1}.$$

1 A Quick Refresher on Analysis and Probability

Exercise 1.1

Let $x_1, x_2 \in \text{conv}(T)$, and $\lambda \in [0,1]$. Then there exists $j, k \in \mathbb{N}$ such that

$$x_1 = a_1 y_1 + \cdots + a_j y_j, a_i \geq 0, \sum_{i=1}^{j} a_i = 1,$$

$$x_2 = b_1 z_1 + \cdots + b_k z_k, b_I \geq 0, \sum_{i=1}^{k} b_i = 1.$$

Then we get

$$\lambda x_1 + (1-\lambda) x_2 = \lambda \sum_{i=1}^{j} a_i y_i + (1-\lambda) \sum_{i=1}^{k} b_i z_i.$$

From the formulation above,

$$a_i \geq 0 \implies \lambda a_i \geq 0, \ b_i \geq 0 \implies (1-\lambda) b_i \geq 0.$$

Moreover, when summing up the coefficients,

$$\lambda \sum_{i=1}^{j} a_i + (1-\lambda) \sum_{i=1}^{k} b_i = \lambda + (1-\lambda) = 1.$$

Therefore $_1 + (1-\lambda)x_2 \in \text{conv}(T)$. Here we assumed that without loss of generality, there are no shared points between x_1 and x_2. If there were to be shared points, it would not have affected our analysis because each coefficient in the convex combination will still be greater than 0, and also their sum will be 1.

Exercise 1.2

Let $f_1, \ldots, f_m$ be convex functions, and $g : K \to \mathbb{R}$ be defined as

$$g(x) = \max_x(f_1(x), \cdots, f_m(x)).$$

Let $x, y \in K$, and let $\lambda \in [0, 1]$. Then

$$\begin{aligned}
g(\lambda x + (1-\lambda)y) &= \max(f_1(\lambda x + (1-\lambda)y), \cdots, f_m(\lambda x + (1-\lambda)y)) \\
&\leq \max(\lambda f_1(x) + (1-\lambda)f_1(y), \cdots, \lambda f_m(x) + (1-\lambda)f_m(y)) \\
&\leq \max(\lambda f_1(x), \cdots, \lambda f_m(x)) + \max((1-\lambda)f_1(y), \cdots, (1-\lambda)f_m(y)) \\
&= \lambda \max(f_1(x), \cdots, f_m(x)) + (1-\lambda)\max(f_1(y), \cdots, f_m(y)) \\
&= \lambda g(x) + 1 - \lambda) g(y).
\end{aligned}$$

Therefore g is a convex function.

Exercise 1.3

(a)

($\implies$) Suppose that f is convex. For the base case, when $m = 2$, by the definition of convexity, the statement is true. For the inductive hypothesis, assume that for some $m \in \mathbb{N}$,

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leq \sum_{i=1}^{m} \lambda_i f(x_i), \lambda_1 \geq 0, \sum_{i=1}^{m} \lambda_i = 1.$$

With $\lambda_j \geq 0, \sum_{j=0}^{m+1} \lambda_j = 1$, without loss of generality assume that $\lambda_{m+1} < 1$ (if not we can switch to another λ that satisfies this condition).

$$\begin{aligned}
f\left(\sum_{i=1}^{m+1} \lambda_i x_i\right) &= f\left((1-\lambda_{m+1})\sum_{j=1}^{m} \frac{\lambda_j}{1-\lambda_{j+1}} x_j + \lambda_{m+1} x_{m+1}\right) \\
&\leq (1-\lambda_{m+1}) f\left((1-\lambda_{m+1})\sum_{j=1}^{m} \frac{\lambda_j}{1-\lambda_{j+1}} x_j\right) + \lambda_{m+1} f(x_{m+1}) \quad \text{(Base case)} \\
&\leq (1-\lambda_{m+1}) \sum_{j=1}^{m} \frac{\lambda_j}{1-\lambda_{m+1}} f(x_j) + \lambda_{m+1} f(x_{m+1}) \quad \text{(Inductive step)} \\
&= \sum_{j=1}^{m} \lambda_j f(x_j) + \lambda_{m+1} f(x_{m+1}) \\
&= \sum_{j=1}^{m+1} \lambda_j f(x_j).
\end{aligned}$$

($\impliedby$) Take $m = 2$ and we are done.

(b)

By the definition given for X,, let

$$P(X = x_i) = p_i, i = 1, \ldots, n, \ p_i \geq 0, \sum_{i=1}^{n} p_i = 1.$$

We can directly see from our construction that

$$\mathbb{E}[X] = \sum_{i=1}^{n} p_i x_i.$$

Then from part (a),

$$f(\mathbb{E}[X]) = f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i) = \mathbb{E}[f(X)].$$

Exercise 1.4

Let $x \in \operatorname{conv}(T)$. Then for some $m \in \mathbb{N}$,

$$x = \lambda_1 z_1 + \cdots + \lambda_m z_m,\ \ \lambda_i \geq 0, \sum_{i=1}^{m} \lambda_i = 1.$$

Then by Jensen's Inequality from Exercise 3,

$$f(x) = f\left(\sum_{i=1}^{m} \lambda_i z_i\right) \leq \sum_{i=1}^{m} \lambda_i f(z_i) \leq \sup_i f(z_i).$$

Therefore we get

$$\sup_{x \in \operatorname{conv}(T)} f(x) \leq \sup_{x \in T} f(x).$$

The other side ("$\geq$") is obvious because $T \subseteq \operatorname{conv}(T)$. Therefore we get the equality.

Exercise 1.5

We'll proceed via proof by induction. For the base case when $n = 1$, let $x \in [-1.1]$. Then x can be written as a combination via

$$x = \frac{1+x}{2} \cdot 1 + \frac{1-x}{2} \cdot (-1).$$

For the inductive step, assume if $x \in [-1,1]^n$, $x \in \text{conv}(\{-1,1\}^n)$. Now let's consider $x \in [-1,1]^{n+1} = (x_1, \cdots, x_{n+1})$. For a fixed value of $x_{n+1} \in [-1,1]$, from the induction hypothesis, $x \in \text{conv}(\{-1,1\}^n)$. Then

$$\begin{aligned} &x x_1, \cdots, x_{n+1}) \\ &= \frac{1+x_{n+1}}{2}(x_1, \cdots, x_n, 1) + \frac{1-x_{n+1}}{2}(x_1, \cdots, x_n, -1). \end{aligned}$$

Therefore x is a convex combination of points from a convex combination (we can acheive that via normalizing), hence $x \in \text{conv}(\{-1,1\}^{n+1})$ so $[-1,1]^n \subseteq \text{conv}(\{-1,1\}^n)$.
For the other side of the proof, let $x \in \text{conv}(\{-1,1\}^n)$. Then $\exists m \leq 2^n$ such that

$$x = \lambda_1 z_1 + \cdot + \lambda_m z_m, z_i \in \text{conv}(\{-1,1\}^n), \lambda_i \geq 0, \sum_{i=1}^{m} \lambda_1 = 1.$$

Each entry x_1 satisfies $-1 \leq x_i \leq 1$ and equality occurs when all corresponding entries in z_i are either 1 or -1, hence $\text{conv}(\{-1,1\}^n) \subseteq [-1,1]^n$.
Finally we conclude that $\text{conv}(\{-1,1\}^n) = [-1,1]^n$.

Exercise 1.6

Let $x \in B_1^n$ so $\sum_{i=1}^n |x_i| \leq 1$. Define the following sets:

$$I_+ = \{i \in \{1, \cdots, n\} : x_i > 0\}, I_- = \{i \in \{1, \cdots, n\} : x_i < 0.$$

Without loss of generality, assume either $|I_+| > 0$ or $I_- > 0$. If both are zero, x has to be the origin, which finding a convex combination from the standard bases vectors would be very easy. Define

$$\lambda_{i+} = \begin{cases} |x_i| & \text{if } \ i \in I_+, \\ 0 & \text{if } \ i \in I_-, \\ \frac{1}{2(|I_-|+|I_+|)}\left(1 - \sum_{i \in I_- \cup I_+} |x_i|\right) & \text{otherwise} \end{cases},$$

$$\lambda_{i+} = \begin{cases} |x_i| & \text{if } \ i \in I_-, \\ 0 & \text{if } \ i \in I_+, \\ \frac{1}{2(|I_-|+|I_+|)}\left(1 - \sum_{i \in I_- \cup I_+} |x_i|\right) & \text{otherwise} \end{cases}.$$

Then,

$$x = \sum_{i=1}^n \lambda_{i+} e_i + \lambda_{i-}(-e_i), \ \lambda_{i+}, \lambda_{i-} \geq 0, \ \sum_{i=1}^n \lambda_{i+} + \lambda_{i-} = 1.$$

Hence $x \in \operatorname{conv}(\{\pm e_1, \cdots, \pm e_n\})$ so $B_1^n \subseteq \operatorname{conv}(\{\pm e_1, \cdots, \pm e_n\})$.
Now let $x \in \operatorname{conv}(\{\pm e_1, \cdots, \pm e_n\})$. Then $\exists \lambda_{i+}, \lambda_{i-} \geq 0$ and summing to 1 such that

$$\begin{aligned} x &= \lambda_{1+} e_1 + \cdots + \lambda_{n+} e_n + \lambda_{1-}(-e_1) + \cdots + \lambda_{n-}(-e_n) \\ &\leq |\lambda_{1+} e_1| + \cdot + |\lambda_{n+} e_n| + |\lambda_{1-} e_1| + |\lambda_{n-} e_n| \\ &= \sum_{i=1}^n |\lambda_{i+}| + |\lambda_{i-}| \\ &= 1. \end{aligned}$$

Therefore $x \in B_1^n$ so $\operatorname{conv}(\{\pm e_1, \cdots, \pm e_n\}) \in B_1^n$. We conclude that $B_1^n = \operatorname{conv}(\{\pm e_1, \cdots, \pm e_n\})$.

Exercise 1.7

Denote E_i =event that freshman i has no friends, X =number of freshman. Then we are bounding

$$\begin{aligned}
\sum_{n=0}^{\infty} P\left(\bigcup_{i=1}^{X} E_i \middle| X = n\right) P(x = n) &= \sum_{n=0}^{\infty} P\left(\bigcup_{i=1}^{n} E_i\right) P(x = n) \\
&\leq \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \sum_{i=1}^{n} P(E_i) \\
&= \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \cdot n(1-p)^{n-1} \\
&= \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!} (1-p)n - 1 \\
&= \lambda \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} (1-p)^n \\
&= \lambda e^{-p\lambda}.
\end{aligned}$$

From the question, since $p \geq 2 \ln \lambda / \lambda$,

$$\lambda e^{-p\lambda} \leq \lambda e^{-2 \ln \lambda} = \frac{1}{\lambda}.$$

Exercise 1.8

Exercise 1.9

Let E_i = the event that student i has no friends, and $B = \{\text{there exists a friendless student}\}$. We are bounding the probability

$$P(B) = P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i) = n(1-p_n)^{n-1}$$

Now when we take the limit,

$$\lim_{n\to\infty} n(1-p_n)^{n-1} < \lim_{n\to\infty} n\left(\frac{(1+\varepsilon)\ln n}{n}\right)^{n-1} \to 0 \text{ as } n \to \infty.$$

Exercise 1.10

(a)

From definition,

$$\mathbb{E}[X_i] = (1-p_n)^{n-1} \implies \mathbb{E}[S_n] = n(1-p_n)^{n-1}.$$

We have that

$$\begin{aligned}
\mathbb{E}[S_n] &= n(1-p_n)^{n-1} \\
&< n\left(1 - \frac{(1-\varepsilon)\ln n}{n}\right)^{n-1} \\
&\to n \cdot e^{-(1-\varepsilon)\ln n} \\
&= \lim_{n\to\infty} n^{\varepsilon} \\
&= +\infty.
\end{aligned}$$

(b)

First, let's look at the random variable $X_i X_j$. This is the indicator variable that both student i and j are friendless. For student i and j, to be friendless, they cannot be friends with each of the the other $n-1$ students, but we double counted the relationship between student i and j! Therefore,

$$\mathbb{E}[X_i X_j] = (1-p_n)^{2n-3} \quad (i \neq j).$$

The random variable X_i^2 is the same as X_i as both represent the indicator that student i is friendless. Therefore, we have

$$\begin{aligned}
\mathbb{E}\left[S_n^2\right] &= \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + \sum_{i\neq j} \mathbb{E}[X_i X_j] \\
&= n(1-p_n)^{n-1} + n(n-1)(1-p_n)^{2n-3}.
\end{aligned}$$

Then, we can calculate the variance of S_n:

$$\begin{aligned}
\mathrm{Var}(S_n) &= \mathbb{E}\left[S_n^2\right] - (\mathbb{E}[S_n])^2 \\
&= n(1-p_n)^{n-1} + n(n-1)(1-p_n)^{2n-3} - n^2(1-p_n)^{2n-2} \\
&= n(1-p_n)^{n-1} - n(np_n - p_n + 1)(1-p_n)^{2n-2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\mathrm{Var}(S_n)}{\mu_n^2} &= \frac{n(1-p_n)^{n-1} - n(np_n - p_n + 1)(1-p_n)^{2n-2}}{n^2(1-p_n)^{2n-2}} \\
&= \frac{1}{\mathbb{E}[S_n]} - \frac{np_n - p_n + 1}{n} \\
&\to 0 - 0 \\
&= 0.
\end{aligned}$$

(c)

$$\begin{aligned}
P(S_n = 0) &= P(S_n - \mu_n = \mu_n) \\
&\leq P(|S_n - \mu_n| \geq \mu_n) \\
&\leq \frac{\mathrm{Var}(S_n)}{\mu_n} \quad \text{(By Chebyshev's inequality)} \\
&\to 0.
\end{aligned}$$

Therefore, there exists at least at least one friendless student with probability that converges to 1 as $n \to \infty$. In fact, this problem (and the previous problem) demonstrate that $\ln n/n$ is an *evolution threshold* for isolated vertices for the Erdős–Rényi model.

Exercise 1.11

(a)

First of all, note that the function $f(x) = x^{q/p}$ is convex. By Jensen's inequality,

$$\mathbb{E}\left[|X|^p\right]^{q/p} \leq \mathbb{E}\left[(|X|^p)^{q/p}\right] = \mathbb{E}\left[|X|^q\right].$$

Taking $(1/q)$th powers on both sides of the inequality gives the result.

(b)

First let $q < \infty$, and definte $a := \frac{p+q}{2}$. Let X be the random variable with pdf

$$f_X(x) = \frac{a}{x^{a+1}}, x \geq 1.$$

Then

$$\begin{aligned} \|X\|_{L^p}^p &= \int_1^\infty x^p \cdot \frac{a}{x^{a+1}}\ dx \\ &= \int_1^\infty ax^{(p-q)/2-1}\ dx \\ &< \infty \quad (\frac{p-q}{2} - 1 < -1). \end{aligned}$$

Also,

$$\begin{aligned} \|X\|_{L^q}^q &= \int_1^\infty x^q \cdot \frac{a}{x^{a+1}}\ dx \\ &= \int_1^\infty ax^{(q-p)/2-1}\ dx \\ &= +\infty \quad (\frac{q-p}{2} - 1 \geq -1). \end{aligned}$$

By taking the limit for $q \to \infty$, we can also see that the bound is true for when $q = \infty$.

Exercise 1.12

We have

$$\begin{aligned}
\|X\|_{L^p}^p &= \mathbb{E}\left[|X| \cdot |X|^{p-1}\right] \\
&= \|X \cdot X^{p-1}\|_{L^1} \\
&\leq \|X\|_{L^1} \|X^{p-1}\|_{L^\infty} \quad \text{(Hölder's inequality)} \\
&= \|X\|_{L^1} \|X\|_{L^\infty}^{p-1}.
\end{aligned}$$

Taking both sides to the $1/p$th power gives the result.

Exercise 1.13

Exercise 1.14

Consider the random vector $X = (X_1, \ldots, X_n)$. For the first inequality,

$$\begin{aligned}\left(\sum_{i=1}^{n} (\mathbb{E}[X_i])^p\right)^{1/p} &= \|\mathbb{E}[X]\|_p \\ &\leq \mathbb{E}[\|X\|_p] \quad \text{(By Jensen's inequality)} \\ &= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i^p\right)^{1/p}\right].\end{aligned}$$

For the second inequality,

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^{n} X_i^p\right)^{1/p}\right]^p &\leq \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i^p\right)^{1/p \cdot p}\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{n} X_i^p\right].\end{aligned}$$

Taking the pth root on both sides of the equation gives the inequality.

Exercise 1.15

(a)

First note that, for every $x \in \mathbb{R}$, we can express x as follows:

$$x = \int_0^x 1 \; dt = \int_0^\infty \mathbf{1}_{\{t<x\}} \; dt - \int_{-\infty}^0 \mathbf{1}_{\{t>x\}} \; dt.$$

Therefore, by plugging the random variable X in and taking expectations on both sides of the equation, we get

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

(b)

By definition, $f(x)$ has to be nonnegative. Then we can express it as

$$f(x) = \int_0^x f'(t) \; dt = \int_0^\infty \mathbf{1}_{\{t<x\}} f'(t) \; dt.$$

Then by plugging in the random variable X and taking expectations on both sides of the equation, we get

$$\mathbb{E}[f(X)] = \int_0^\infty P(X > t) f'(t) \; dt.$$

(c)

Note that $|X|$ is nonnegative, and the function $f : \mathbb{R}_+ \to \mathbb{R}_+$ defined by $f(x) = x^p$ is increasing and differentiable for every $p \in (0, \infty)$, and $f(0) = 0$. By applying part (b),

$$\mathbb{E}[|X|^p] = \int_0^\infty P(|X| > t) f'(t) \; dt = \int_0^\infty P(|X| > t) p t^{p-1} \; dt.$$

Exercise 1.16

We can rewrite the expectation as follows:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[X\mathbf{1}_{\{X\le\varepsilon\mathbb{E}[X]\}}\right] + \mathbb{E}\left[X\mathbf{1}_{\{X>\varepsilon\mathbb{E}[X]\}}\right] \\ &\le \varepsilon\mathbb{E}[X] + \mathbb{E}[X]^{1/2} P(X > \varepsilon\mathbb{E}[X])^{1/2} \quad \text{(By Cauchy-Schwartz).}\end{aligned}$$

Then, move all terms with $\mathbb{E}[X]$ to one side and squaring the inequality gives

$$(1-\varepsilon)^2\mathbb{E}[X]^2 \le \mathbb{E}\left[X^2\right] P(X > \varepsilon\mathbb{E}[X]).$$

Then, dividing $\mathbb{E}\left[X^2\right]$ on both sides of the inequality above gives

$$P(X > \varepsilon\mathbb{E}[X]) \ge (1-\varepsilon)^2\frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]},$$

hence the result is proven.

Exercise 1.17

(a)

For the first equality, without loss of generality we can assume that $\|x\|_p = 1$. Then we have that

$$|x_i| \leq 1 \text{ for all } i \implies |x_i|^q \leq |x_i|^p \text{ for all } q \geq p.$$

This gives

$$\begin{aligned}\|x\|_q &= \left(\sum_{i=1}^n |x_i|^q\right)^{1/q} \\ &\leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/q} \\ &= \|x\|_p^{p/q}.\end{aligned}$$

Taking qth powers on both sides of the equation gives the first inequality.
For the second inequality, let us denote

$$y := (|x_1|^p, \ldots, |x_n|^p)m, \ z := \mathbf{1}_n.$$

Consider the conjugate exponents $\frac{q}{p}$ and $\frac{q}{q-p}$ (it's pretty easy to check that this pair is indeed a pair of conjugate exponents). We have

$$\begin{aligned}\|x\|_p^p &= \sum_{i=1}^n |x_i|^p \\ &= \langle y, z \rangle \\ &\leq \left(\sum_{i=1}^n y_i^{q/p}\right)^{p/q} \cdot \left(\sum_{i=1}^n z_i^{q/(q-p)}\right)^{(q-p)/q} && \text{(By Hölder's inequality)} \\ &= \|x\|_q^p \cdot n^{1-\frac{p}{q}}.\end{aligned}$$

Taking the pth root on both sides of the equation gives

$$\|x\|_p \leq n^{\frac{1}{p}-\frac{1}{q}} \|x\|_q,$$

which is exactly what we're looking for.

(b)

Consider $x = e_i$, the ith standard basis vector in $\mathbb{R}^n$. Then

$$\|x\|_p = (0^p + \cdots + 0^p + 1^p)^{1/p} = 1 = \|x\|_q.$$

Now consider $y = \mathbf{1}_n$, the vector of all ones in $\mathbb{R}^n$. Then

$$\|y\|_p = n^{1/p} = n^{1/p-1/q+1/q} = n^{1/p-1/q}\|y\|_q.$$

Therefore the two bounds in part (a) can both be tight.

Exercise 1.18

(a)

Take $q = \infty$ from Exercise 1.17 to get

$$\|x\|_\infty \le \|x\|_p \le n^{1/p}\|x\|_\infty.$$

As $p \to \infty$, $n^{1/p} \to 1$ hence $\|x\|_p \to \|x\|_\infty$.

(b)

Again, take $q = \infty$ to get

$$\|x\|_\infty \le \|x\|_p \le n^{1/p}\|x\|_\infty.$$

On the right hand side of the inequality,

$$n^{1/p}\|x\|_\infty = e^{\ln n/p}\|x\|_\infty \le e\|x\|_\infty \text{ only if } p > \ln n.$$

Exercise 1.19

(a)

2 Sums of Independent Random Variables

Exercise 2.1

First note that by independence,

$$\mathbb{E}[Y_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n] = 0.5^n.$$

For the first inequality, denote $A = \{X_i \geq 0.5 \quad \forall i\}$. Then we can get that $P(A) \leq P(Y_n \geq \mathbb{E}[Y_n])$ because $A \subseteq \{Y_n \geq \mathbb{E}[Y_n]\}$. By independence, $P(A) = 0.5^n$ and we are done. For the second inequality, since all X_i are nonnegative, by Markov's inequality we get

$$P(Y_n \geq \mathbb{E}[Y_n]) = P(\sqrt{Y_n} \geq \sqrt{\mathbb{E}[Y_n]}) \leq \frac{\mathbb{E}[\sqrt{Y_n}]}{\sqrt{\mathbb{E}[Y_n]}}.$$

The denominator is just $\sqrt{(1/2)^n}$ from earlier. For the numerator, since X_i are iid,

$$\mathbb{E}[\sqrt{Y_n}] = (\mathbb{E}[\sqrt{X_i}])^n.$$

Now this problem amounts to finding the expected value of $\sqrt{X_i}$.

$$\begin{aligned} F_{\sqrt{X_i}}(x) &= P(\sqrt{X_i} \leq x) \\ &= P(X_i \leq x^2) \\ &= F_{X_i}(x^2) \\ &= x^2. \end{aligned}$$

Then the pdf of $\sqrt{X_i}$ is

$$f_{\sqrt{X_i}}(x) = 2x, \quad x \in [0, 1].$$

Finally, we get that

$$\mathbb{E}[\sqrt{X_i}] = \int_0^1 2x^2 \, dx = \frac{2}{3} \implies P(Y_n \geq \mathbb{E}[Y_n]) \leq \frac{(2/3)^n}{(1/2)^{n/2}} = \left(\frac{2\sqrt{2}}{3}\right)^n \leq 0.95^n.$$

Exercise 2.2

We'll approach this problem via a more calculus-based flavor. Define the function

$$\begin{aligned} f(t) &:= P(g \geq t) - \frac{t}{t^2+1} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \\ &= \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx. \end{aligned}$$

First we can immediately see that $f(x) = \frac{1}{2} > 0$. Moreover,

$$\begin{aligned} f'(t) &= \frac{1}{\sqrt{2\pi}} e^{-t^2/2} + \frac{t^4+2t^2-1}{(t^2+1)^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \\ &= -\frac{2}{(t^2+1)^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \\ &< 0 \quad \text{for all } t > 0. \end{aligned}$$

We can also see that

$$\lim_{t\to\infty} f(t) = 0 - 0 = 0.$$

From these three facts, $f(t) \geq 0$ for all $t > 0$ hence the inequality follows.

Exercise 2.3

(a)

$$f'(x) = (-x) \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = -xf(x) \text{ for all } x.$$

(b)

Using integration by parts and the property from part (a),

$$\begin{aligned}
\int_t^\infty f(x)\ dx &= \int_t^\infty -\frac{f'(x)}{x}\ dx \\
&= \left[-\frac{f(x)}{x}\right]_t^\infty - \int_t^\infty f(x)\ d(-\frac{1}{x}) \\
&= \frac{f(t)}{t} - \int_t^\infty \frac{f(x)}{x^2}\ dx \\
&= \frac{f(t)}{t} - \int_t^\infty -\frac{f'(x)}{x^3}\ dx \\
&= \frac{f(t)}{t} - \left[\left[-\frac{f(x)}{x^3}\right]_t^\infty - \int_t^\infty f(x)\ d(-\frac{1}{x^3})\right] \\
&= \frac{f(t)}{t} - \frac{f(t)}{t^3} + 3\int_t^\infty \frac{f(x)}{x^4}\ dx \quad (*) \\
&= \frac{f(t)}{t} - \frac{f(t)}{t^3} + 3\int_t^\infty -\frac{f'(x)}{x^5}\ dx \\
&= \frac{f(t)}{t} - \frac{f(t)}{t^3} + 3\left[-\frac{f(x)}{x^5}\right]_t^\infty - \int_t^\infty f(x)\ d(-\frac{1}{x^5}) \\
&= \frac{f(t)}{t} - \frac{f(t)}{t^3} + \frac{3f(t)}{t^5} - 5\int_t^\infty \frac{f(x)}{x^6}\ dx. \quad (**)
\end{aligned}$$

Since $f(t) > 0$, dividing $(*)$ by $f(t)$ we will get the first inequality, and dividing $(**)$ by $f(t)$ results in the second inequality.

Exercise 2.4

(a)

$$\begin{aligned}
\mathbb{E}[g\mathbf{1}_{g>t}] &= \int_t^\infty x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\ dx \\
&= \left[-\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right]_t^\infty \\
&= \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.
\end{aligned}$$

(b)

$$\begin{aligned}
\mathbb{E}[g^2\mathbf{1}_{g>t}] &= \int_t^\infty x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\ dx \\
&= \left[-\frac{x}{\sqrt{2\pi}} e^{-x^2/2} \right]_t^\infty + \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\ dt \\
&\le \frac{t}{\sqrt{2\pi}} e^{-t^2/2} + \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad \text{(prop 2.1.2)} \\
&= \left(t + \frac{1}{t} \right) \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.
\end{aligned}$$

Exercise 2.5

We start by expanding both quantities into their respective Taylor series representations:

$$\begin{aligned}
\cosh x &= \frac{e^x + e^{-x}}{2} \\
&= \frac{1}{2}\sum_{n=0}^{\infty} \frac{x^n}{n!} + \frac{1}{2}\sum_{n=0}^{\infty} (-1)^n \frac{x^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \quad (x \in \mathbb{R}),
\end{aligned}$$

$$\exp(x^2/2) = \sum_{n=0}^{\infty} \frac{(x^2/2)^n}{n!} = \sum_{n=0}^{\infty} \frac{x^{2n}}{n!2^n} \quad (x \in \mathbb{R}).$$

From the above, if we subtract the Taylor series of $\cosh x$ from that of $\exp(x^2/2)$, to show that this quantity is nonnegative, it is enough to show that

$$n!2^n \leq (2n)! \quad \forall n \in \mathbb{N}_0.$$

We'll proceed via proof by induction. For the base case, when $n = 0$,

$$0!2^0 = 1 \leq 1 = (2 \cdot 0)!$$

For the inductive step, assume that for some $n \in \mathbb{N}_0$, the statement is true.

$$\begin{aligned}
(n+1)!2^{n+1} &= 2(n+1)n!2^n \\
&\leq 2(n+1) \cdot (2n)! \\
&\leq (2n+2)(2n+1)(2n)! \\
&= (2(n+1))!
\end{aligned}$$

Therefore we are done.

Exercise 2.6

As per usual, let $\lambda > 0$. By the typical procedure for the exponential moment method,

$$\begin{aligned} P(g \geq t) &= P(\lambda g \geq \lambda t) \\ &= P(\exp(\lambda g) \geq \exp(\lambda t)) \\ &\leq e^{-\lambda t}\mathbb{E}[\exp(\lambda g)] \\ &= \exp(-\lambda t + \lambda^2/2). \end{aligned}$$

Defining the quantity above as a function $f(\lambda)$,

$$f'(\lambda) = (-t + \lambda)\exp(-\lambda t + \lambda^2/2) = 0 \implies \lambda^* = t.$$

Moreover,

$$f''(\lambda^*) = \exp(-\lambda t + \lambda^2/2) > 0.$$

Therefore by the second derive test, we have found a minimizer $\lambda^* = t$ for the quantity on the RHS bound. Plugging back gives the result.

Exercise 2.7

Again, we'll use the exponential moment method so let $\lambda > 0$.

$$\begin{aligned}
P\left(\sum_{i=1}^{N} X_i \le \varepsilon N\right) &= P\left(\sum_{i=1}^{N} -(X_i/\varepsilon) \ge -N\right) \\
&= P\left(\exp\left(\sum_{i=1}^{N} -(X_i/\varepsilon) \ge \exp(-\lambda N)\right)\right. \\
&\le e^{\lambda N}\mathbb{E}[\exp\left(\lambda \sum_{i=1}^{n} -X_i/\varepsilon\right)] \\
&= e^{\lambda N}\prod_{i=1}^{N}\mathbb{E}[\exp(-\lambda X_i/\varepsilon)].
\end{aligned}$$

Now fix i. Since X_i is uniformly bounded by K, we have

$$\begin{aligned}
\mathbb{E}[\exp(-\lambda X_i/\varepsilon)] &= \int_0^\infty e^{-\lambda x/\varepsilon} f_X(x)\ dx \\
&\le \int_0^\infty K e^{-\lambda x/\varepsilon}\ dx \\
&= K\left[-\frac{\varepsilon}{\lambda}e^{-\lambda x/\varepsilon}\right]_0^\infty \\
&= \frac{K\varepsilon}{\lambda}.
\end{aligned}$$

Combining the above gives

$$P\left(\sum_{i=1}^{N} X_i \le \varepsilon N\right) \le e^{\lambda N}\left(\frac{K\varepsilon}{\lambda}\right)^N = (K\varepsilon)^N e^{\lambda N}\lambda^{-N}.$$

Defining the result above as $f(\lambda)$ and differentiate, we get

$$f'(\lambda) = (K\varepsilon)^N(N\lambda - N)e^{\lambda N}\lambda^{-N-1} = 0 \implies \lambda^* = 1.$$

Moreover,

$$f''(\lambda^*) = (K\varepsilon)^N \cdot Ne^{\lambda N} > 0.$$

Therefore $\lambda^* = 1$ is the minimizer of our bound, and plugging back gives the result.

Exercise 2.8

The function $f(x) = e^{\lambda x}$ is convex because $f''(x) = \lambda^2 e^{\lambda x} \geq 0$. By Jensen's inequality, for any $a, b \in \mathbb{R}$ and $p \in [0, 1]$,

$$\begin{aligned} f(pa + (1-p)b) &= e^{\lambda(pa+(1-p)b)} \\ &\leq pe^{\lambda a} + (1-p)e^{\lambda b} \\ &= pf(a) + (1-p)f(b). \end{aligned}$$

In particular, this means for any $x \in [a, b]$,

$$e^{\lambda x} \leq e^{\lambda a} + (1-p)e^{\lambda b} \implies e^{\lambda X} \leq e^{\lambda a} + (1-p)e^{\lambda b}.$$

Taking expectations on both sides,

$$\mathbb{E}[e^{\lambda X}] \leq e^{\lambda a} + (1-p)e^{\lambda b} = \mathbb{E}[e^{\lambda Y}].$$

Exercise 2.9

(a)

We can assume WLOG X has mean zero because we can define another random variable $Y = X - \mathbb{E}[X]$, then Y will take values between $[a - \mathbb{E}[X], b - \mathbb{E}[X]]$, which does not affect the analysis.
We can assume WLOG $b - a = 1$ because we can define another random variable $Y = X/(b-a)$. Then Y takes values in $[a/(b-a), b/(b-a)]$, which again does not affect the analysis.
We can assume that X takes values in $\{a, b\}$ because from Exercise 2.8 (add link), if we prove that the bound is true for the discrete version, we have also effectively proven it for the continuous version.

(b)

Without loss of generality assume X satisfies everything in part (a). Define $P(X = a) = p$. Then from the expectation and range conditions,

$$\begin{cases} pa + (1-p)b & = 0 \\ b - a & = 1 \end{cases} \implies p = a + 1 = b, \ 1 - p = -a.$$

After finding p, we can solve for the cumulant generating function $K(\lambda)$:

$$\begin{aligned} K(\lambda) &= \ln\left(\mathbb{E}[e^{\lambda X}]\right) \\ &= \ln\left(be^{\lambda a} + (-a)e^{\lambda b}\right) \\ &= \ln\left(be^{\lambda a} - ae^{\lambda(a+1)}\right) \\ &= \ln e^{\lambda a}(b - ae^{\lambda}) \\ &= \lambda a + \ln\left(b - ae^{\lambda}\right). \end{aligned}$$

We can see clearly that $K(0) = 0 + \ln(b - a) = 0$. We also get

$$K'(\lambda) = a - \frac{ae^{\lambda}}{b - ae^{\lambda}} \implies K'(0) = a - a = 0.$$

Moreover,

$$\begin{aligned} K''(\lambda) &= -\frac{ae^{\lambda}(b - ae^{\lambda}) - ae^{\lambda} \cdot (-ae^{\lambda})}{(b - ae^{\lambda})^2} \\ &= -\frac{abe^{\lambda}}{(b - ae^{\lambda})^2} \\ &\leq \frac{(-ae^{\lambda} + b)^2}{2^2(b - ae^{\lambda})^2} \quad \text{(AM-GM inequality with } x = -ae^{\lambda}, y = b\text{)} \\ &= \frac{1}{4}. \end{aligned}$$

Then by Taylor's Theorem,

$$K(\lambda) = K(0) + \lambda K'(0) + \frac{\lambda^2 K''(\xi)}{2!} \leq 0 + 0 + \frac{\lambda^2}{4 \cdot 2!} = \frac{\lambda^2}{8} \text{ for all } \lambda \in \mathbb{R}.$$

Exercise 2.10

We're going to use the exponential moment method. Let $\lambda > 0$. Then

$$
\begin{aligned}
P\left(\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right) &= P\left(\exp\left(\lambda\sum_{i=}^{N}(X_i - \mathbb{E}[X_i])\right) \geq \exp(\lambda t)\right) \\
&\leq e^{-\lambda t}\prod_{i=1}^{N}\mathbb{E}\left[\exp\left(\lambda X_i - \mathbb{E}[X_i]\right)\right] \quad \text{(By Markov + independence)} \\
&= e^{-\lambda t}\prod_{i=1}^{N}\exp\left(\frac{\lambda^2(b_i - a_i)^2}{8}\right) \quad \text{(By Exercise 2.9)} \\
&= \exp\left(-\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^{N}(b_i - a_i)^2\right).
\end{aligned}
$$

Again, we set the right hand side to be a function of λ and minimize it.

$$
f'(\lambda) = \left(-t + \frac{\lambda}{4}\sum_{i=1}^{N}(b_i - 1_i)^2\right) f(\lambda) = 0 \implies \lambda^* = \frac{4t}{\sum_{i=1}^{N}(b_i - a_i)^2}.
$$

Moreover,

$$
\begin{aligned}
f''(\lambda) &= \frac{1}{4}\sum_{i=1}^{N}(b_i - a_i)^2 f(\lambda) + \left(-t + \frac{\lambda}{4}\sum_{i=1}^{N}(b_i - a_i)^2\right) f(\lambda) \\
\implies f''(\lambda^*) &= \frac{1}{4}\sum_{i=1}^{N}(b_i - a_i)^2 \cdot \exp\left(-\frac{2t^2}{\sum_{i=1}^{N}(b_i - a_i)^2}\right) \geq 0.
\end{aligned}
$$

By the second derivative test, $\lambda*$ indeed minimizes the expression that we found earlier. Lastly, plug λ^* back to the original equation above gives the result.

Exercise 2.11

We'll use the exponential moment method. Let $\lambda > 0$. Then

$$\begin{aligned} P(S_N \leq t) &= P(-S_N \geq -t) \\ &= P(\exp(-\lambda S_N) \geq \exp(-\lambda t)) \\ &\leq e^{\lambda t}\mathbb{E}\left[\exp(-\lambda S_N)\right] \quad \text{(By Markov's inequality)} \\ &= e^{\lambda t}\prod_{i=1}^{N}\mathbb{E}\left[\exp(-\lambda X_i)\right] \quad \text{(By independence)} \\ &\leq \exp\left(\lambda t + (e^{-\lambda} - 1)\mu\right), \quad (*) \end{aligned}$$

where for the last inequality we used the definition of expectation:

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

Then we get

$$\prod_{i=1}^{N}\mathbb{E}\left[\exp(-\lambda X_i)\right] \leq \exp\left((e^{-\lambda} - 1)\sum_{i=1}^{N}p_i\right) = \exp\left((e^{-\lambda} - 1)\mu\right).$$

We set the right hand side of $(*)$ as a function of λ and minimize it.

$$f'(\lambda) = (t - \mu e^{-\lambda})f(\lambda) = 0 \implies e^{-\lambda} = \frac{t}{\mu} \implies \lambda^* = \ln\left(\frac{\mu}{t}\right).$$

Moreover,

$$f''(\lambda) = \mu e^{-\lambda}f(\lambda) + (t - \mu e^{-\lambda})^2 f(\lambda) \implies f''(\lambda^*) = t \cdot \exp\left(t\ln\left(\frac{\mu}{t}\right) + \left(\frac{t}{\mu} - 1\right)\mu\right) > 0.$$

By the second derivative test, λ^* indeed minimizes $(*)$. Plugging λ^* back to $(*)$ then gives the result.

Exercise 2.12

From the pmf of a binomial random variable,

$$\begin{aligned}
P(S_N = t) &= \binom{N}{t} \left(\frac{\mu}{N}\right)^t \left(1 - \frac{\mu}{N}\right)^{N-t} \\
&\geq \left(\frac{N}{t}\right)^t \left(\frac{\mu}{N}\right)^t \left(1 - \frac{\mu}{N}\right)^{N-t} \qquad \text{(Exercise 0.6)} \\
&\geq \left(\frac{N}{t}\right)^t \left(\frac{\mu}{N}\right)^t \left(1 - \frac{\mu}{N}\right)^{N-\mu} \qquad (\mu \leq t) \\
&\geq \left(\frac{N}{t}\right)^t \left(\frac{\mu}{N}\right)^t e^{-\mu} \\
&= e^{-\mu} \left(\frac{\mu}{t}\right)^t.
\end{aligned}$$

Exercise 2.13

(a)

We'll use the exponential moment method. Let $\lambda > 0$. Then

$$\begin{aligned} P(X \geq t) &= P(e^{\lambda X} \geq e^{\lambda t}) \\ &\leq e^{-\lambda t}\mathbb{E}\left[e^{\lambda X}\right] \quad \text{(By Markov's inequality)} \\ &= \exp\left(-\lambda t + \mu(e^{\lambda} - 1)\right). \quad (*) \end{aligned}$$

We set the right hand side of $(*)$ as a function of λ and minimize it.

$$f'(\lambda) = (-t + \mu e^{\lambda})f(\lambda) = 0 \implies \lambda^* = \ln\left(\frac{t}{\mu}\right).$$

Moreover,

$$f''(\lambda) = \mu e^{\lambda} f(\lambda) + (-t + \mu e^{\lambda})^2 f(\lambda) \implies f''(\lambda^*) = t\exp\left(-t\ln\left(t/\mu\right) + t - \mu\right) \geq 0.$$

By the second derivative test, λ^* indeed minimizes $(*)$. Plugging λ^* back to $(*)$ then gives the result.

(b)

We'll use the exponential moment method. Let $\lambda > 0$. Then

$$\begin{aligned} P(X \leq t) &= P(-X \geq -t) \\ &= P(e^{-\lambda X} \geq e^{-\lambda t}) \\ &\leq e^{\lambda t}\mathbb{E}\left[e^{-\lambda X}\right] \quad \text{(By Markov's inequality)} \\ &= \exp\left(\lambda t + \mu(e^{-\lambda} - 1)\right). \end{aligned}$$

We set the right hand side of $(*)$ as a function of λ and minimize it.

$$f'(\lambda) = (t - \mu e^{-\lambda})f(\lambda) = 0 \implies \lambda^* = \ln\left(\frac{\mu}{t}\right).$$

Moreover,

$$f''(\lambda) = \mu e^{-\lambda} f(\lambda) + (t - \mu e^{-\lambda})^2 f(\lambda) \implies f''(\lambda^*) = t \cdot \exp\left(t\ln\left(\mu/t\right) + \mu(t/\mu - 1)\right) \geq 0.$$

By the second derivative test, λ^* indeed minimizes $(*)$. Plugging λ^* back to $(*)$ then gives the result.

(c)

By Theorem 2.3.1 with $t = (1+\delta)\mu$,

$$\begin{aligned} P(X - \mu \geq \delta\mu) &= P(X \geq (1+\delta)\mu) \\ &\leq e^{-\mu}\left(\frac{e\mu}{(1+\delta)\mu}\right)^{(1+\delta)\mu} \quad \text{(By Chernoff's inequality)} \\ &= e^{-\mu + (1+\delta)\mu} \cdot e^{-(1+\delta)\ln(1+\delta)\mu} \\ &= \exp\left(-\mu((1+\delta)\ln(1+\delta) - \delta)\right). \end{aligned}$$

From the proof of Corollary 2.3.4, we already proved that

$$(1+\delta)\ln(1+\delta) - \delta \geq \frac{\delta^2}{3}.$$

Therefore, by plugging this into our original bound, we get

$$P(X - \mu \geq \delta\mu) \leq \exp\left(-\frac{\delta^2\mu}{3}\right).$$

For the other side, we apply Remark 2.3.2 and get

$$\begin{aligned}
P(X - \mu \leq -\delta\mu) &= P(X \leq (1-\delta)\mu) \\
&\leq e^{-\mu}\left(\frac{e}{(1-\delta)\mu}\right)^{(1-\delta)\mu} \qquad \text{(By Remark 2.3.2)} \\
&= \exp\left(-\mu((1-\delta)\ln(1-\delta)+\delta)\right).
\end{aligned}$$

Again, from the proof of Corollary 2.3.4, we already have

$$(1-\delta)\ln(1-\delta)+\delta \geq \frac{\delta^2}{2} \geq \frac{\delta^2}{3},$$

hence plugging this bound into our original bound gives

$$P(X - \mu \leq -\delta\mu) \leq \exp\left(-\frac{\delta^2\mu}{3}\right).$$

Combining the two results above gives the statement.

(d)

$$P(X = t) = \frac{e^{-\mu}\mu^t}{t!} \geq \frac{e^{-\mu}\mu^t}{t^t} = e^{-\mu}\left(\frac{\mu}{t}\right)^t.$$

Exercise 2.14

For this question, we'll use the following identity:

$$\ln(1+x) \geq \frac{x}{1+x/2} \text{ for all } x \geq 0.$$

This is above is true since we can define the function

$$g(x) := \ln(1+x) - \frac{x}{1+x/2}.$$

Then

$$g(0) = 0, \; g'(x) = \frac{1}{1+x} - \frac{1}{(1+x/2)^2} \geq 0.$$

Ok great! Let's go back to the problem. Just like Exercise 2.13, we have to show two sides of the bound, then combine together.

For the upper bound, again by applying Theorem 2.3.1 with $t = (1+\delta)\mu$,

$$\begin{aligned}
P(S_N - \mu \geq \delta\mu) &= P(S_N \geq (1+\delta)\mu) \\
&\leq \exp\left(-\mu((1+\delta)\ln(1+\delta) - \delta)\right) \\
&\leq \exp\left(-\frac{\delta^2\mu}{2+\delta}\right)
\end{aligned}$$

where in the last inequality, we used that fact that (due to the identity at the beginning),

$$\begin{aligned}
(1+\delta)\ln(1+\delta) - \delta &\geq \frac{(1+\delta)\delta}{1+\delta/2} - \delta \\
&= \frac{\delta^2/2}{1+\delta/2} \\
&= \frac{\delta^2}{2+\delta}.
\end{aligned}$$

For the lower bound, the case where $\delta > 1$ is trivial. For the other case where $\delta \leq 1$, from the proof of Corollary 2.3.4, just like the previous exercise,

$$\begin{aligned}
P(X - \mu \leq -\delta\mu) &= P(X \leq (1-\delta)\mu) \\
&\leq e^{-\mu}\left(\frac{e}{(1-\delta)\mu}\right)^{(1-\delta)\mu} \qquad \text{(By Remark 2.3.2)} \\
&= \exp\left(-\mu((1-\delta)\ln(1-\delta) + \delta)\right).
\end{aligned}$$

Moreover, we have

$$(1-\delta)\ln(1-\delta) + \delta \geq \frac{\delta^2}{2} \geq \frac{\delta^2}{2+\delta},$$

and therefore

$$P(X - \mu \leq -\delta\mu) \leq \exp\left(-\frac{\delta^2\mu}{2+\delta}\right).$$

Combining the two bounds completes the proof.

Exercise 2.15

When using the exponential moment method in proving the Chernoff inequalities, the key chain of inequalities that we relied on is as follows:

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

From Exercise 2.8, if X, Y are random variables such that X takes values in $[a, b]$ and Y takes values in $\{a, b\}$, then

$$\mathbb{E}\left[\exp(\lambda X_i)\right] \leq \mathbb{E}\left[\exp(\lambda Y_i)\right] \text{ for all } \lambda \in \mathbb{R}.$$

Therefore if the bound holds true for independent Bernoulli random variables, they have to hold for any independent random variables taking values on $[0, 1]$.

Exercise 2.21

Let $X_i = \{\text{Algorithm at ith time is wrong}\}$. Then $X_i \sim \text{Ber}(\frac{1}{2} - \varepsilon)$. By Hoeffding's inequality (Theorem 2.2.6), the probability that the answer is wrong is

$$\begin{aligned} P\left(\sum_{i=1}^{N} X_i \geq \frac{N}{2}\right) &= P\left(\sum_{i=1}^{N} X_i - \left(\frac{N}{2} - N\varepsilon\right) \geq \frac{N}{2} - \left(\frac{N}{2} - N\varepsilon\right)\right) \\ &= P\left(\sum_{i=1}^{N} \left(X_i - \left(\frac{1}{2} - \varepsilon\right)\right) \geq N\varepsilon\right) \\ &\leq \exp\left(-\frac{2N^2\varepsilon^2}{N}\right) \\ &= \exp\left(-2N\varepsilon^2\right) \\ &\leq \delta. \end{aligned}$$

Solving the inequality above for N gives

$$-2N\varepsilon^2 \leq \ln \delta \implies N \geq \frac{1}{2\varepsilon^2} \ln\left(\frac{1}{\delta}\right).$$

Exercise 2.22

(a)

The pdf of the standard normal distribution is

$$f_g(s) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \ x \in \mathbb{R}.$$

Then

$$\begin{aligned}
\mathbb{E}\left[|g|^p\right] &= \int_{-\infty}^{\infty} |x|^p \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ dx \\
&= \int_{-\infty}^{0} (-x)^p \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ dx + \int_{0}^{\infty} x^p \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ dx \\
&= 2 \int_{0}^{\infty} x^p \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ dx \\
&= 2 \int_{0}^{\infty} (2t)^{(p-1)/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-t} \ dt \quad (t = x^2/2) \\
&= \frac{2^{p/2}}{\sqrt{\pi}} \int_{0}^{\infty} t^{(p-1)/2} e^{-t} \ dt \\
&= \frac{2^{p/2}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right).
\end{aligned}$$

(b)

From part (a),

$$\begin{aligned}
\|g\|_{L^p} &= \left[\frac{2^{p/2}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right)\right]^{1/p} \\
&\approx \left[\frac{2^{p/2}}{\sqrt{\pi}} \cdot \sqrt{2\pi(p+1)/2} \cdot \left(\frac{p-1}{2e}\right)^{(p-1)/2}\right]^{1/p} \\
&= \left[\sqrt{2(p-1)} \cdot \left(\frac{p-1}{e}\right)^{(p-1)/2}\right]^{1/p} \\
&= \left[\sqrt{2e}\left(\frac{p}{e}\right)^{p/2}\right]^{1/p} \\
&= \sqrt{\frac{p}{e}} \cdot (2e)^{1/p} \\
&= \sqrt{\frac{p}{e}}(1 + o(1)) \text{ as } p \to \infty.
\end{aligned}$$

3 Random Vectors in High Dimensions

4 Random Matrices

Exercise 4.1

(a)

This is just note true in general. We can only do this if A is symmetric, as in general the left and right singular vectors cannot cancel out.

(b)

By using the matrix form of the SVD,

$$\begin{aligned} A^{-1} &= (U\Sigma V^T)^{-1} \\ &= V^{-T}(U\Sigma)^{-1} \\ &= V\Sigma^{-1}U^T \\ &= \sum_{i=1}^{n} \sigma_i^{-1} v_i u_i^T. \end{aligned}$$

Exercise 4.2

(a)

From Definition 4.1.8, we check the criterion for a norm, namely:
For positive definiteness, we have

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \geq 0, \text{ and } = 0 \iff \|Ax\|_2 = 0 \iff A = 0_{m \times n}.$$

For absolute homegeneity, we have (for $c \in \mathbb{R}$)

$$\|cA\| = \max_{x \neq 0} \frac{\|cAx\|_2}{\|x\|_2} = \frac{|c|\|Ax\|_2}{\|x\|_2} = |c|\|A\|.$$

For the triangle inequality, we have

$$\|A + B\| = \max_{x \neq 0} \frac{\|(A + B)x\|_2}{\|x\|_2} \leq \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} + \frac{\|Bx\|_2}{\|x\|_2} \leq \|A\| + \|B\|.$$

(b)

We can write the SVDs respectively: $A = U\Sigma V^T$ and $A^T = V\Sigma^T U$. Since Σ and Σ^T have the same entries, we directly get that

$$\|A\| = \|A^T\|.$$

(c)

We'll first show this: $\|Ax\|_2 \leq \|A\| \|x\|_2$ for all x. If $x = 0$ then the inequality is trivial. If $x \neq 0$,

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \geq \frac{\|Ax\|_2}{\|x\|_2}.$$

With this fact, we have

$$\|AB\| = \max_{x \neq 0} \frac{\|ABx\|_2}{\|x\|_2} \leq \max_{x \neq 0} \frac{\|B\| \|Ax\|_2}{\|x\|_2} = \|A\| \|B\|.$$

(d)

Let A be the matrix and B be its submatrix, and let u be a unit vector. We have

$$\|B\| = \max_{\|u\|_2 = 1} \|Bu\|_2 = \max_{\|u\|_2 = 1} \| \begin{bmatrix} B \\ 0 \end{bmatrix} \begin{bmatrix} u \\ 0 \end{bmatrix} \|_2 \leq \max_{\|u\|_2 = 1} \|A \begin{bmatrix} u \\ 0 \end{bmatrix} \|_2 \leq \max_{\|u\|_2 = 1} \|Au\|_2 = \|Au\|.$$

Exercise 4.3

(a)

We can write the reduced svd of uv^T as

$$uv^T = u' \cdot \|u\|_2 \|v\|_2 \cdot v'^T$$

where u', v' are unit vectors, hence we get the equality

$$\|uv^T\| = \|u\|_2 \|v\|_2.$$

We also have that

$$\begin{aligned}
\|u\|_2 \|v\|_2 &= \sqrt{\sum_{i=1}^{m} u_i^2} \cdot \sqrt{\sum_{j=1}^{n} v_j^2} \\
&= \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (u_i v_j)^2} \\
&= \|uv^T\|_F.
\end{aligned}$$

(b)

We can write the matrix A as

$$A = \sum_{i=1}^{n} a_i e_i e_i^T$$

where e_i is the ith standard basis vector in $\mathbb{R}^n$ (which in this case are both the left and right singular vectors). To express A in its SVD, we'll have to look at a_i: if it is negative, we'll have to reverse the sign, and put another negative sign on one of the singular vectors. Therefore we get

$$\|A\| = \max_i |a_i|.$$

Exercise 4.4

(a)

The first inequality directly follows from Lemma 4.1.11:

$$\|A\| = \sigma_1 \leq \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} = \|A\|_F.$$

For the second inequality,

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \leq \sqrt{\sigma_1^2 + \cdots + \sigma_1^2} = \sqrt{r}\|A\|.$$

(b)

By Proposition 3.2.1 (b),

$$\begin{aligned}
\mathbb{E}\left[\|AZ\|_2^2\right] &= \operatorname{tr}(\mathbb{E}\left[AZZ^TA^T\right]) \\
&= \operatorname{tr}(A\mathbb{E}\left[ZZ^T\right]A^T) \\
&= \operatorname{tr}(AA^T) \quad \text{(By isotropy)} \\
&= \|A\|_F^2.
\end{aligned}$$

(c)

Let $g \sim N(0, I_n)$. We know that g is isotropic. Then by part (b),

$$\|BA\|_F^2 = \mathbb{E}\left[\|BAg\|_2^2\right] \leq \|B\|^2\mathbb{E}\left[\|Ag\|_2^2\right] = \|B\|^2\|A\|_F^2.$$

Taking square roots on both sides gives the result.

(d)

In the simpler case where A is diagonal, we can actually write the matrix BA out explicitly:

$$BA = \begin{bmatrix} a_{11}b_{11} & a_{22}b_{12} & \cdots & a_{mm}b_{1m} \\ a_{11}b_{21} & a_{22}b_{22} & \cdots & a_{mm}b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{11}b_{k1} & a_{22}b_{k2} & \cdots & a_{mm}b_{km} \end{bmatrix}.$$

Then we can directly get that

$$\begin{aligned}
\|BA\|_F^2 &= \sum_{i=1}^m a_{ii}^2\|B_{:i}\|_2^2 \\
&\leq \sum_{i=1}^m a_{ii}^2 \cdot \max_{i=1,\ldots,m}\|B_{:i}\|_2^2 \\
&= \|B\|_{1\to2}^2\|A\|_F^2.
\end{aligned}$$

Taking square roots on both sides gives the result.

Exercise 4.5

From Lemma 4.1.11,

$$\|A\|_F^2 = \sigma_1^2 + \cdots + \sigma_r^2.$$

There are two cases:
Case 1: $k \leq r$**.** We have

$$k\sigma_k^2 \leq \sum_{i=1}^{k} \sigma_1^2 \leq \|A\|_F^2.$$

Taking square roots on both sides gives the result.
Case 2: $K > r$**.** For $k > r$, $\sigma_k = 0$ hence the inequality holds trivially.

Exercise 4.6

Exercise 4.7

5 Concentration Without Independence

Exercise 5.1

(a)

Let $\varepsilon > 0$ be given, and $f : X \to Y$ with X, Y being metric spaces. Choose $p, q \in X$ such that $d_X(p, q) < \varepsilon/L$ where L is the Lipschitz constant of the function f. Then

$$d_Y(f(p), f(q)) \leq L d_X(p, q) < L \cdot \frac{\varepsilon}{L} = \varepsilon$$

so that f is uniformly continuous.

(b)

(Assume that we have bounded gradient or else the statement would be false) By the mean value theorem, for $x, y \in \mathbb{R}^n$,

$$\begin{aligned} |f(y) - f(x)| &= |\nabla f(\lambda x + (1-\lambda)y)^T (y - x)| \quad \text{(Mean value theorem)} \\ &\leq \|\nabla f(\lambda x + (1-\lambda)y)\|_2 \|y - x\|_2 \quad \text{(Cauchy-Schwartz inequality)} \\ &\leq \sup_{x \in \mathbb{R}^n} \|\nabla f(x)\|_2 \|y - x\|_2. \end{aligned}$$

(c)

Define

$$f(x) = \begin{cases} x \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases}.$$

We can check that f is indeed continuous. The squeeze theorem gives

$$0 \leq \left| x \sin \frac{1}{x} \right| \leq |x|,$$

meaning that the limit of $f(x)$ as $x \to 0$ is indeed 0. Since f is continuous on the compact interval $[-1, 1]$, by the Heine-Cantor theorem, f is uniformly continuous. However, if we look at the derivative of f,

$$f'(x) = \begin{cases} \sin \frac{1}{x} - \frac{\cos(1/x)}{x} & \text{if } x \neq 0, \\ \text{undefined} & \text{if } x = 0. \end{cases}$$

As we take $x \to 0$, the derivative becomes unbounded, hence it cannot be the case that f is Lipschitz.

(d)

Let $f(x) = |x|$. Then clearly $f(x)$ is not differentiable at $x = 0$. However, $f(x)$ is 1-Lipschitz as the supremum of the absolute value of the derivative in the interval $[-1, 1]$ is 1.

Exercise 5.2

(a)

$$\begin{aligned}|f(y) - f(x)| &= |\langle y, \theta\rangle - \langle x, \theta\rangle| \\ &= |\langle y - x, \theta\rangle| \\ &\le \|\theta\|_2 \|y - x\|_2. \quad \text{(Cauchy-Schwartz inequality)}\end{aligned}$$

(b)

$$\begin{aligned}|f(y) - f(x)| &= |Ay - Ax| \\ &= |A(y - x)| \\ &\le \|A\| \|y - x\|_2.\end{aligned}$$

(c)

$$\begin{aligned}|f(y) - f(x)| &= |\|y\| - \|x\|| \\ &\le \|y - x\| \quad \text{(Reverse triangle inequality)} \\ &\le L\|y - x\|_2.\end{aligned}$$

6 Quadratic Forms, Symmetrization and Contraction

7 Random Processes

8 Chaining

9 Deviations of Random Matrices on Sets