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

Gallant Tsao

July 18, 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

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

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

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

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:

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

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

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

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

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 2

Let $f_1, \dots, f_m$ 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 3

(a)

($\Longrightarrow$) 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).

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

($\Longleftarrow$) 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 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 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

$$\begin{aligned} &xx_1, \cdots, x_{n+1}) \\ &= \frac{1+x_{n+1}}{2}(x_1, \cdots, x_n, 1) + \frac{1-x_{n+1}}{2}(x_1, \cdots, x_n, -1). \end{aligned}$$

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 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 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 8

Exercise 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 10

(a)

Proving this statement is equivalent of proving

$$\mathbb{E}[|X|^p]^{q/p} \le \mathbb{E}[|X|^q].$$

The function $f(x) = x^{q/p}$ is convex because $q/p \ge 1$. Then by Jensen's Inequality,

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

(b)

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

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

Then we have that

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

2 Sums of Independent Random Variables

Exercise 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

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 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 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)

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

Exercise 5

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

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

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

From the above, if we subtract the Taylor series of $\cosh x$ from that of $\exp\left(x^2/2\right)$, 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 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 7

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

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

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

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

Combining the above gives

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

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,

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

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

Exercise 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 $p \in [0, 1]$,

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

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 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,

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

Then by Taylor's Theorem,

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

3 Concentration Without Independence

Exercise 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 2

(a)

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

(b)

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

(c)

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