Homework 1

SDS 391P.6, Spring 2026

Pratik Patil

Due: Feb 2 (Monday)

This homework is (also) a work in progress and is provided as-is for instructional purposes. The problems are drawn from various sources and do not (yet) have sufficient references to the original material. Additionally, errors may be present. so some cauution is advised! The document will be updated if corrections are necessary. Last updated: 2026-02-02.

0 Guidelines

- Please start early. If you have questions about the statements, notation, or possible typos, email us as soon as possible. When emailing about the course, please begin the subject line with [SDS 391P.6].

- Please begin your answer to each *main* question on a separate page. If you use any code, include it in an appendix. Submit a single combined PDF to Canvas. (If you encounter any submission issues, please let us know.)

- The problems and motivations draw on multiple sources (past course material, textbooks, and other standard references). If you use any external resources that materially guide your solution (beyond routine lookups), please cite them in your write-up.

- These questions are designed to build intuition and technique. You are welcome to go beyond what is explicitly asked. If you introduce additional assumptions (while keeping the spirit of the problem), state them clearly. If you discover something interesting along the way, feel free to include it as a brief remark; we may share especially instructive observations with the class.

- Parts labeled [Bonus] are optional: they are not required for full credit. They are intended as extra practice. You may skip them without penalty. If you attempt them, please label your solutions clearly with [Bonus].

- Many parts include hints intended to help you get started. You are not required to follow the suggested route, and you are encouraged to try alternative approaches when appropriate.

- We will grade primarily for correctness and clear reasoning. Do not over-optimize for minor presentation details. The spirit of the homework is for you to learn something new!

1 Practice with ℓ_p and L_p norms

This exercise is meant to build comfort with manipulating ℓ_p norms on $\mathbb{R}^n$ and L_p norms of random variables. We also highlight the key difference between finite-dimensional (dimension-dependent) norm equivalence and the infinite-dimensional/probability-space setting.

(a) ℓ_p **norms on** $\mathbb{R}^n$. Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. For $1 \leqslant p < \infty$, define

$$\|x\|_{\ell_p} := \Big(\sum_{i=1}^n |x_i|^p\Big)^{1/p}, \qquad \|x\|_{\ell_\infty} := \max_{1 \leqslant i \leqslant n} |x_i|.$$

Also, define the unit ball $B_{\ell_p}^n := \{x \in \mathbb{R}^n : \|x\|_{\ell_p} \leqslant 1\}$.

(a1) (Comparison and monotonicity) Prove that if $1 \leqslant p \leqslant q \leqslant \infty$, then

$$\|x\|_{\ell_q} \leqslant \|x\|_{\ell_p} \leqslant n^{\frac{1}{p} - \frac{1}{q}} \|x\|_{\ell_q}.$$

Deduce that $\|x\|_{\ell_p}$ is (weakly) decreasing in p and that the unit balls are nested: $B_{\ell_p}^n \subseteq B_{\ell_q}^n$ for $p \leqslant q$.
(*Hint:* For $q < \infty$, normalize x so that $\|x\|_{\ell_q} = 1$ and apply Hölder's inequality to $\sum_i |x_i|^p = \sum_i |x_i|^p \cdot 1$. For $q = \infty$, use $|x_i| \leqslant \|x\|_{\ell_\infty}$.)

(a2) (When are the bounds tight?) Give examples of nonzero vectors $x, y \in \mathbb{R}^n$ such that

$$\|x\|_{\ell_p} = \|x\|_{\ell_q} \qquad \text{and} \qquad \|y\|_{\ell_p} = n^{\frac{1}{p} - \frac{1}{q}} \|y\|_{\ell_q}.$$

(*Hint:* think "one-sparse" versus "flat".)

(a3) [Bonus] (The ℓ_∞ limit) Show that $\|x\|_{\ell_p} \to \|x\|_{\ell_\infty}$ as $p \to \infty$. Also, show the quantitative bound

$$\|x\|_{\ell_\infty} \leqslant \|x\|_{\ell_p} \leqslant n^{1/p} \|x\|_{\ell_\infty}.$$

Conclude that if $p \geqslant \log n$ then $\|x\|_{\ell_p} \leqslant e\,\|x\|_{\ell_\infty}$. (Here, log denotes the natural logarithm.)
(*Hint:* Note that $n^{1/p} \to 1$ as $p \to \infty$.)

(b) L_p **norms of random variables.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a real-valued random variable. For $1 \leqslant p < \infty$, define

$$\|X\|_{L_p} := (\mathbb{E}|X|^p)^{1/p}, \qquad \|X\|_{L_\infty} := \operatorname{ess\,sup} |X|.$$

Write $L_p := \{X : \|X\|_{L_p} < \infty\}$.

(b1) (Monotonicity and inclusions) Prove that, for $1 \leqslant p \leqslant q \leqslant \infty$,

$$\|X\|_{L_p} \leqslant \|X\|_{L_q}.$$

Deduce the set inclusion $L_q \subseteq L_p$ for $q \geqslant p$.
(*Hint:* For $q < \infty$, use Jensen's inequality on the convex function $t \mapsto t^{q/p}$ applied to $|X|^p$. For $q = \infty$, use $|X| \leqslant \|X\|_{L_\infty}$ a.s. from the definition of essential supremum.)

(b2) (No dimension-free equivalence) Explain why there cannot exist a universal constant C such that $\|X\|_{L_q} \leqslant C\|X\|_{L_p}$ holds for all random variables on general probability spaces whenever $p < q$. Give an explicit example of X such that $X \in L_p$ but $X \notin L_q$.
(*Hint:* take a Pareto-type tail $\mathbb{P}(|X| > t) \asymp t^{-\alpha}$ for large t and choose $\alpha \in (p, q]$ so that $\mathbb{E}|X|^p < \infty$ but $\mathbb{E}|X|^q = \infty$.)

(b3) [Bonus] (The L_∞ limit) Assume $\|X\|_{L_\infty} < \infty$. Show that $\|X\|_{L_p} \to \|X\|_{L_\infty}$ as $p \to \infty$.
(*Hint:* Show for any $\varepsilon > 0$, $\mathbb{P}(|X| > \|X\|_{L_\infty} - \varepsilon) > 0$ by definition of essential supremum, and lower-bound $\mathbb{E}|X|^p$ on this event.)

2 Practice with matrix norms

In this exercise, you will practice with operator norms (including spectral norm) and the Frobenius norm, and prove a few inequalities that appear repeatedly in statistics and machine learning. Throughout, let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times k}$ be conformable matrices, and let $1 \leqslant p, q, r \leqslant \infty$. Recall $\|x\|_p$ denotes the ℓ_p norm on Euclidean spaces. For any $s \in [1, \infty]$, let s' denote the conjugate exponent: $\frac{1}{s} + \frac{1}{s'} = 1$ (with the conventions $1/\infty = 0$ and $1/0 = \infty$).

(a) **The $p \to q$ operator norm.** Define the induced (operator) norm

$$\|A\|_{p \to q} := \sup_{\|x\|_p \leqslant 1} \|Ax\|_q.$$

(a1) (Equivalent definitions) Show that

$$\|A\|_{p \to q} = \sup_{\|x\|_p = 1} \|Ax\|_q = \sup_{x \neq 0} \frac{\|Ax\|_q}{\|x\|_p}.$$

(*Hint:* use homogeneity: $\|A(\alpha x)\|_q = |\alpha| \|Ax\|_q$ and rescale x.)

(a2) (Submultiplicativity) Prove the composition bound

$$\|AB\|_{p \to r} \leqslant \|A\|_{q \to r} \|B\|_{p \to q}.$$

Side note: One can interpret $\|A\|_{q \to r}$ and $\|B\|_{p \to q}$ as the Lipschitz constants of the linear maps $u \mapsto Au$ and $x \mapsto Bx$ between the corresponding normed spaces. So the Lipschitz constant of the composition map $z \mapsto ABz$ is at most the product of Lipschitz constants of the individual maps.

(*Hint:* start from $\|ABx\|_r \leqslant \|A\|_{q \to r} \|Bx\|_q$ and then take $\sup_{\|x\|_p \leqslant 1}$.)

(a3) [Bonus] (Duality form and transpose relationship) Show the bilinear representation

$$\|A\|_{p \to q} = \sup_{\|x\|_p = 1, \ \|y\|_{q'} = 1} |y^\top A x| = \sup_{\|x\|_p = 1, \ \|y\|_{q'} = 1} y^\top A x.$$

Deduce the transpose identity

$$\|A^\top\|_{q' \to p'} = \|A\|_{p \to q}.$$

In particular, conclude that for the $2 \to 2$ operator norm (spectral norm),

$$\|A^\top\|_{2 \to 2} = \|A\|_{2 \to 2}.$$

(*Hint:* use the dual norm identity for ℓ_q norm: $\|u\|_q = \sup_{\|y\|_{q'} = 1} y^\top u$ for vectors $u \in \mathbb{R}^m$, and apply it to $u = Ax$. The absolute value can be dropped as one can flip $y \mapsto -y$. For the transpose relationship, rewrite $y^\top A x = x^\top A^\top y$)

(b) **Spectral versus Frobenius norms.** In this part, write $\|A\|$ for the spectral norm $\|A\|_{2 \to 2}$. Recall the Frobenius norm $\|A\|_F = \left(\sum_{i,j} A_{ij}^2 \right)^{1/2}$.

(b1) (Rank-one and diagonal matrices) Let $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$. Show that

$$\|uv^\top\| = \|uv^\top\|_F = \|u\|_2 \|v\|_2.$$

Let $D \in \mathbb{R}^{n \times n}$ be a diagonal matrix with entries $(d_1, \ldots, d_n)$. Write $d = (d_1, \ldots, d_n)^\top \in \mathbb{R}^n$. Show that

$$\|D\| = \max_{1 \leqslant i \leqslant n} |d_i| = \|d\|_\infty \quad \text{and} \quad \|D\|_F = \|d\|_2.$$

(b2) (General relationship) If $\mathrm{rank}(A) = r$, show that

$$\|A\| \leqslant \|A\|_F \leqslant \sqrt{r}\,\|A\|.$$

Show that both inequalities can be tight (achieved) for any admissible (m, n, r).
(*Hint:* rank-one versus equal singular values.)

(b3) (Frobenius submultiplicativity) Show that for conformable matrices A, B,

$$\|AB\|_F \leqslant \|A\|\,\|B\|_F \qquad \text{and} \qquad \|AB\|_F \leqslant \|A\|_F\,\|B\|.$$

Conclude (by combining an inequality from (b2)) that $\|\cdot\|_F$ is submultiplicative, i.e., $\|AB\|_F \leqslant \|A\|_F\|B\|_F$.
(*Hint:* write $\|AB\|_F^2 = \mathrm{tr}(B^\top A^\top AB)$ and use $A^\top A \preceq \|A\|^2 I$.)

(b4) [Bonus] (Orthogonal invariance) Let $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ be orthogonal. Show that

$$\|UAV\| = \|A\| \qquad \text{and} \qquad \|UAV\|_F = \|A\|_F.$$

(*Hint:* use $\|Uz\|_2 = \|z\|_2$ for all z and apply the definitions; for Frobenius you may also use $\|A\|_F^2 = \mathrm{tr}(A^\top A)$ and the cyclic property of trace.)

3 Practice with variance and covariance identities

This problem provides practice using a few core variance/covariance identities for random vectors. Unless specified otherwise, assume all random variables/vectors below have finite second moments. Throughout, for $u, v \in \mathbb{R}^d$ we write $\langle u, v \rangle := u^\top v$ and $\|u\|_2^2 = \langle u, u \rangle$.

(a) **Variance identities.** Let $Z \in \mathbb{R}^d$ be a random vector with mean $m := \mathbb{E}[Z] \in \mathbb{R}^d$ (defined coordinatewise). The quantity $\mathbb{E}\|Z - \mathbb{E}Z\|_2^2$ is sometimes called the "total variance".

(a1) (Variance via Pythagorean decomposition) Assume $\mathbb{E}\|Z\|_2^2 < \infty$ and let $m = \mathbb{E}Z$. Show that

$$\mathbb{E}\|Z - m\|_2^2 \;=\; \mathbb{E}\|Z\|_2^2 - \|m\|_2^2.$$

More generally, show that for every $a \in \mathbb{R}^d$,

$$\mathbb{E}\|Z - a\|_2^2 \;=\; \mathbb{E}\|Z - m\|_2^2 + \|a - m\|_2^2,$$

and conclude that

$$\mathbb{E}\|Z - m\|_2^2 \;=\; \min_{a \in \mathbb{R}^d} \mathbb{E}\|Z - a\|_2^2, \qquad \text{with unique minimizer } a^\star = \mathbb{E}[Z].$$

Side note: This is an instance of a projection identity in L_2 and provides a geometric reason for why $\mathbb{E}[Z]$ is the best constant estimator under squared loss.
(*Hint:* write $Z - a = (Z - m) + (m - a)$ and expand $\|Z - a\|_2^2$.)

(a2) (Variance via symmetrization) Let Z' be an independent copy of Z (i.e., $Z' \perp\!\!\!\perp Z$ and $Z' \stackrel{d}{=} Z$). Show that

$$\mathbb{E}\|Z - \mathbb{E}Z\|_2^2 \;=\; \frac{1}{2}\,\mathbb{E}\|Z - Z'\|_2^2.$$

Side note: This way of expressing variance as an average squared distance is a standard trick. Centering can often be replaced by two i.i.d. copies and a trick of this form by the name "symmetrization".
(*Hint:* write $Z - Z'$ as $(Z - m) - (Z' - m)$ and expand $\|Z - Z'\|_2^2$.)

(a3) [Bonus] (Variance for independent sums) Let $Z_1, \ldots, Z_k \in \mathbb{R}^d$ be independent random vectors with $\mathbb{E}[Z_j] = 0$ and $\mathbb{E}\|Z_j\|_2^2 < \infty$. Show that

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

Side note: Independence is not necessary for this part, and pairwise uncorrelatedness is sufficient.

(b) **Covariance identities.** For a random vector $Z \in \mathbb{R}^d$ with $m = \mathbb{E}[Z]$, define the covariance matrix

$$\mathrm{Cov}(Z) := \mathbb{E}\left[(Z - m)(Z - m)^\top\right] \in \mathbb{R}^{d \times d}.$$

Show that:

(b1) $\mathrm{Cov}(Z) = \mathbb{E}[ZZ^\top] - mm^\top$.

(b2) For every $v \in \mathbb{R}^d$, $\mathrm{Var}(v^\top Z) = v^\top \mathrm{Cov}(Z) v$.

(b3) $\mathrm{tr}(\mathrm{Cov}(Z)) = \mathbb{E}\|Z - \mathbb{E}Z\|_2^2$.

(*Hint:* use $\mathrm{tr}(uu^\top) = \|u\|_2^2$ and linearity of trace/expectation.)

4 Practice with moving between moments and tails

A recurring theme in high-dimensional probability is that moments (e.g. $\mathbb{E}|X|^p$) and tails (e.g. $\mathbb{P}\{|X| > t\}$) control each other. This exercise asks you to practice moving in both directions. Throughout, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Assume all random variables below are real-valued and measurable, and that any expectations that appear are finite.

(a) **From tails to moments.** Let $X \geqslant 0$ be a nonnegative random variable.

(a1) (Moment identities using tail integrals) Show that

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}\{X > t\} \, dt.$$

More generally, show that for any $p > 0$,

$$\mathbb{E}[X^p] = \int_0^\infty p \, t^{p-1} \, \mathbb{P}\{X > t\} \, dt.$$

(*Hint:* Start from the pointwise identity $x = \int_0^\infty \mathbf{1}\{x > t\} \, dt$ and apply Tonelli's theorem. For the pth-moment identity, use the identity $x^p = \int_0^\infty p \, t^{p-1} \mathbf{1}\{x > t\} \, dt$.)

(a2) [Bonus] (Moment growth bounds from exponential tail behavior) Assume there exist constants $c, C > 0$ and $\alpha > 0$ such that for all $t \geqslant 0$,

$$\mathbb{P}\{X > t\} \leqslant C \exp(-ct^\alpha).$$

Use part (a1) to show that $X \in L_p$ for all $p < \infty$ and that there exists a constant $C' = C'(c, C, \alpha)$ such that

$$\|X\|_{L_p} \leqslant C' p^{1/\alpha} \qquad \text{for all } p \geqslant 1.$$

Finally, interpret the cases $\alpha = 2$ and $\alpha = 1$ in terms of the heuristic moment growth for sub-Gaussian and sub-exponential random variables.

(*Hints:* Plug the tail bound into the identity $\mathbb{E}[X^p] = \int_0^\infty p\, t^{p-1} \mathbb{P}\{X > t\}\, dt$. Then compare the resulting integral to a Gamma-function integral by the change of variables $u = ct^\alpha$. You may use (without proof) that $\Gamma(z + 1) \leqslant (C_0 z)^z$ for all $z \geqslant 1$ and some absolute constant C_0.)

(b) **Classical tail bounds from second moments.** Let X and Y be real-valued random variables.

(b1) (Upper bound) Fix $t > 0$. Show the one-sided *Cantelli's inequality*

$$\mathbb{P}\{Y - \mathbb{E}Y \geqslant t\} \;\leqslant\; \frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y) + t^2}.$$

(*Hint:* For any $a > 0$, apply Markov's inequality to the nonnegative random variable $(Y - \mathbb{E}Y + a)^2$. Then optimize the resulting upper bound over $a > 0$.)

(b2) (Comparison with Chebyshev's inequality) Use Chebyshev's inequality to give a "naive" one-sided bound on $\mathbb{P}\{Y - \mathbb{E}Y \geqslant t\}$ and compare it to Cantelli's inequality from (b1). Now apply Cantelli to Y and to $-Y$ and deduce a two-sided tail bound

$$\mathbb{P}\{|Y - \mathbb{E}Y| \geqslant t\} \leqslant \frac{2\,\mathrm{Var}(Y)}{\mathrm{Var}(Y) + t^2}.$$

Compare this to the standard two-sided Chebyshev inequality.

(*Hint:* Cantelli's inequality is always at least as strong as naive one-sided Chebyshev bound. For two-sided bounds, Chebyshev's bound is tighter for large t.)

(b3) [Bonus] (Lower bound) Let $X \geqslant 0$ with $\mathbb{E}[X^2] < \infty$ and $\mathbb{E}[X] > 0$. Fix $\theta \in (0, 1)$. Prove the *Paley–Zygmund inequality*

$$\mathbb{P}\{X \geqslant \theta\, \mathbb{E}X\} \;\geqslant\; (1 - \theta)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]} \;=\; (1 - \theta)^2 \frac{(\mathbb{E}X)^2}{\mathrm{Var}(X) + (\mathbb{E}X)^2} \;=\; \frac{(1 - \theta)^2}{1 + \mathrm{Var}(X)/(\mathbb{E}X)^2}.$$

In particular, if $\mathrm{Var}(X) \leqslant c\,(\mathbb{E}X)^2$ for some $c > 0$, then

$$\mathbb{P}\{X \geqslant \theta\, \mathbb{E}X\} \;\geqslant\; \frac{(1 - \theta)^2}{1 + c}.$$

Markov's inequality says a nonnegative X is unlikely to be *much larger* than its mean. Paley–Zygmund inequality gives a complementary result: unless X has very large second moment relative to $(\mathbb{E}X)^2$, it cannot be *much smaller* than its mean with very high probability. Equivalently, if X is not dominated by extremely rare "spikes," then X must be of order $\mathbb{E}X$ with at least constant probability.

(*Hint:* write $\mathbb{E}[X] = \mathbb{E}[X\mathbf{1}\{X < \theta\mathbb{E}X\}] + \mathbb{E}[X\mathbf{1}\{X \geqslant \theta\mathbb{E}X\}]$ and bound the first term by $\theta\mathbb{E}X$. For the second term, use Cauchy–Schwarz: $\mathbb{E}[X\mathbf{1}_E] \leqslant (\mathbb{E}[X^2])^{1/2}\,\mathbb{P}(E)^{1/2}$.)

5 Mean versus median

The mean and the median are both "typical values," but they optimize different losses: $\mathbb{E}X$ minimizes $a \mapsto \mathbb{E}(X - a)^2$ (squared loss), while any median minimizes $a \mapsto \mathbb{E}|X - a|$ (absolute loss). This exercise quantifies when these two notions agree up to the scale of the standard deviation. Throughout, assume all expectations that appear are finite.

(a) **Mean–median closeness (using variance).** Let Z be a real-valued random variable with $\mathbb{E}[Z^2] < \infty$, and let M_Z be a median of Z, meaning

$$\mathbb{P}(Z \geqslant M_Z) \geqslant \tfrac{1}{2} \qquad \text{and} \qquad \mathbb{P}(Z \leqslant M_Z) \geqslant \tfrac{1}{2}.$$

Show that

$$|M_Z - \mathbb{E}Z| \;\leqslant\; \sqrt{\operatorname{Var}(Z)}.$$

(*Hint:* Apply the one-sided Cantelli's inequality to $Y = Z - \mathbb{E}Z$. If $M_Z \geqslant \mathbb{E}Z$, bound $\mathbb{P}\{Z - \mathbb{E}Z \geqslant M_Z - \mathbb{E}Z\}$ from below using the median property and from above using Cantelli. Repeat similarly if $M_Z \leqslant \mathbb{E}Z$ (or apply Cantelli to $-Y$).)

(b) [Bonus] **Mean–median closeness (using variance proxy).** Let X be a real-valued random variable with median M_X. Assume there exist constants $a > 0$ and $b > 0$ such that for all $t > 0$,

$$\mathbb{P}\big(|X - M_X| > t\big) \;\leqslant\; a\, e^{-t^2/b}.$$

Show that

$$|M_X - \mathbb{E}X| \;\leqslant\; \mathbb{E}|X - M_X| \;\leqslant\; \min\Big\{\sqrt{ab},\; a\sqrt{b\pi/2}\Big\}.$$

(*Hint:* (i) Use Jensen's inequality: $|\mathbb{E}X - M_X| = |\mathbb{E}(X - M_X)| \leqslant \mathbb{E}|X - M_X|$. (ii) Let $Y = |X - M_X|$ and use the tail integral identity $\mathbb{E}Y = \int_0^\infty \mathbb{P}(Y > t)\,dt$ together with the assumed tail bound to prove $\mathbb{E}|X - M_X| \leqslant a\sqrt{b\pi/2}$. (iii) Also compute $\mathbb{E}Y^2 = \int_0^\infty 2t\,\mathbb{P}(Y > t)\,dt \leqslant ab$ and conclude $\mathbb{E}Y \leqslant \sqrt{\mathbb{E}Y^2} \leqslant \sqrt{ab}$.)

Source material

Parts of this homework were inspired by exercises from Vershynin (2018); Boucheron et al. (2013), in addition to the author's accumulated experience working on related topics.

References

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities - A Nonasymptotic Theory of Independence.* Oxford University Press.

Vershynin, R. (2018). *High-dimensional Probability: An Introduction with Applications in Data Science.* Cambridge University Press.