Homework I
Theoretical Statistics/Machine Learning

Sagar Ghosh
Department of Statistics and Data Sciences
The University of Texas at Austin
Email: `sg63684@my.utexas.edu`

January 21, 2026

1 Practice with norms.

a ℓ_p norms on $\mathbb{R}^n$. Let $x = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^n$. We define for $p \geq 1$

$$\|x\|_{\ell_p} = \left[\sum_{i=1}^{n} |x_i|^p\right]^{1/p}, \quad \|x\|_{\ell_\infty} = \max_{1\leq i\leq n} |x|_i.$$

We also define the unit ball in these norms: $B^n_{\ell_p} = \{x \in \mathbb{R}^n : \|x\|_{\ell_p} \leq 1\}$.

a_1 **Comparison and Monotonicity.** Suppose $1 \leq p \leq q \leq \infty$. Let $y := x/\|x\|_{\ell_p}$. Then $\|y\|_{\ell_p} = 1$. As a result, $|y_i|^q \leq |y_i|^p \leq 1$. [Since all the components of y have absolute value ≤ 1.] Therefore, $\|x\|_{\ell_q}/\|x\|_{\ell_p} = \|y\|_{\ell_q} \leq \|y\|_{\ell_p} \leq 1$. This implies $\|x\|_{\ell_q} \leq \|x\|_{\ell_p}$. Therefore, the ℓ_p norms are (weakly) decreasing in p. This argument holds even $q = \infty$. We will use Jensen's inequality to prove the other side. For $p \leq q$, the function $f(x) = x^{q/p}$ is increasing for $x \geq 0$. Using Jensen, we have $f(\sum_i \lambda_i x_i) \leq \sum_i \lambda_i f(x_i)$. Choose $\lambda_i = 1/n$ and replacing x_i by $|x_i|^p$, we get

$$\left(\frac{1}{n}\sum_{i=1}^{n} |x_i|^p\right)^{q/p} \leq \frac{1}{n}\sum_{i=1}^{n} (|x_i|^p)^{q/p} = \frac{1}{n}\sum_{i=1}^{n} |x_i|^q,$$

$$\|x\|_{\ell_p} = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \leq n^{1/p-1/q}\left(\sum_{i=1}^{n} |x_i|^q\right)^{1/q} = n^{1/p-1/q}\,\|x\|_{\ell_q}.$$

If $q = \infty$, then $\|x\|_{\ell_\infty} = \max_{1\leq i\leq n} |x_i|$. Therefore, $\|x\|_{\ell_p} \leq \left[\sum_{i=1}^{n} \|x\|^p_{\ell_\infty}\right]^{1/p} = n^{1/p}\|x\|_{\ell_\infty}$. Additionally, $x \in B^n_{\ell_p} \implies \|x\|_{\ell_p} \leq 1$. Hence, $\|x\|_{\ell_q} \leq \|x\|_{\ell_p} \leq 1$. So, $x \in B^n_{\ell_q}$. Hence, $B^n_{\ell_p} \subseteq B^n_{\ell_q}$.

a_2 **When are the bounds tight ?** Consider $x = \{1, 0, 0, \ldots, 0\}$. Then, $\|x\|_{\ell_p} = \|x\|_{\ell_q} = 1$. On the other hand, consider $x = \{1, 1, \ldots, 1\}$. $\|x\|_{\ell_p} = n^{1/p}$ and $\|x\|_{\ell_q} = n^{1/q}$. Hence, $\|x\|_{\ell_p} = n^{1/p} = n^{1/p-1/q}n^{1/q} = n^{1/p-1/q}\|x\|_{\ell_q}$.

a_3 **The ℓ_∞ limit.** The bounds are already shown in part (a1). Now, the norm $\|\cdot\|_{\ell_p}$ is decreasing in p for $p \geq 1$. Also, $\|x\|_{\ell_p} \geq \|x\|_{\ell_\infty}$. Hence, for all $x \in \mathbb{R}^n$, the sequence $\{\|x\|_{\ell_p}\}_p$ is decreasing and is bounded below by $\|x\|_{\ell_\infty}$. By Bolzano-Weistrass, any monotonic bounded sequence must converge to its limit, and since the lower bound here is achievable, the limit $\lim_{p\to\infty} \|x\|_{\ell_p} = \|x\|_{\ell_\infty}$. This can be also shown by the sandwich property: as $p \to \infty$, $n^{1/p} \to 1$. Hence, $\|x\|_{\ell_\infty} \leq \lim_{p\to\infty} \|x\|_{\ell_p} \leq \lim_{p\to\infty} n^{1/p}\|x\|_{\ell_\infty}$. Hence, $\lim_{p\to\infty} \|x\|_{\ell_p} = \|x\|_{\ell_\infty}$. If $p \geq \log n$, $n^{1/p} \leq n^{1/\log n} = \exp^{\log_n n} = e$. Therefore, $\|x\|_{\ell_p} \leq e\|x\|_{\ell_\infty}$.

b **Norms of random Variables** $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and X is a real valued random variable. For

$1 \leq p < \infty$ we define

$$\|X\|_{L_p} = [\mathbb{E}|X|^p]^{1/p}, \quad \|X\|_{L_\infty} = \operatorname{ess\,sup}|X|.$$

b_1 **Monotonicity and Inclusions.** For $1 \leq p \leq q < \infty$, we have the function $f(t) = t^{q/p}$ is convex. So, applying Jensen's inequality on this function (by replacing t by $|X|^p$), we have

$$\|X\|_{L_p}^q = f(\mathbb{E}|\mathbb{X}|^{\prime}) \leq \mathbb{E}(f(|X|^p)) = \mathbb{E}|X|^q = \|X\|_{L_q}^q.$$

Eliminating the indices on both sides, we get the required inequality. For $q = \infty$,

$$\|X\|_{L_p} = [\mathbb{E}|X|^p]^{1/p} = \left[\int_\Omega |x|^p d\mathbb{P}\right]^{1/p} \leq \operatorname*{ess\,sup}_{x\in\Omega}|X(x)| = \operatorname{ess\,sup}|X|.$$

Therefore if $X \in L_q$, $\|X\|_{L_p} \leq \|X\|_{L_q} < \infty$. Hence, $X \in L_p$. So, $L_q \subseteq L_p$.

b_2 **No Dimension-free equivalence.** The pdf of a Pareto distribution with parameters $\alpha > 0$ and 1 is given as

$$f_X(x) = \begin{cases} \alpha x^{-\alpha-1}, x \geq 1 \\ 0, \text{otherwise.} \end{cases}$$

Now, for $X \sim \text{Pareto}(\alpha, 1)$, $\mathbb{E}|X|^p = \int_1^\infty \alpha t^p t^{-\alpha-1} dt = \left[\frac{\alpha t^{p-\alpha}}{p-\alpha}\right]_1^\infty$. And $\mathbb{E}|X|^q = \int_1^\infty \alpha t^q t^{-\alpha-1} dt = \left[\frac{\alpha t^{q-\alpha}}{q-\alpha}\right]_1^\infty$. If we choose $q > \alpha > p$ [for example, $\alpha = \frac{p+q}{2}$ works], then $\mathbb{E}|X|^q = \infty$ and $\mathbb{E}|X|^p < \infty$. Therefore, there can be no universal constant C such that $\|X\|_{L_q} \leq C\|X\|_{L_p}$ works for any general probability space.

b_3 **The L_∞ limit.** For $\epsilon > 0$, let $\Omega_\epsilon = \{w \in \Omega : |X(w)| > \|X\|_{L_\infty} - \epsilon\}$. Then

$$\mathbb{E}|X|^p = \int_\Omega |X|^p d\mathbb{P} \geq \int_{\Omega_\epsilon} |X|^p d\mathbb{P} \geq \int_{\Omega_\epsilon} (\|X\|_{L_\infty} - \epsilon)^p d\mathbb{P} = (\|X\|_{L_\infty} - \epsilon)^p \mathbb{P}(\Omega_\epsilon). \tag{1}$$

Therefore, $\|X\|_{L^p} \geq (\|X\|_{L_\infty} - \epsilon)\mathbb{P}(\Omega_\epsilon)^{1/p}$. Since $\mathbb{P}(\Omega_\epsilon) > 0$, $\lim_{p\to\infty} \mathbb{P}(\Omega_\epsilon)^{1/p} = 1$. taking the limit $p \to \infty$ on both sides of (1) gives, $\lim_{p\to\infty} \|X\|_{L_p} \geq \|X\|_{L_\infty} - \epsilon$. Also from part (b1) we know that for all $p \geq 1$, $\|X\|_{L_p} \leq \|X\|_{L_\infty}$. Therefore, as $\epsilon > 0$ is arbitrary, $\lim_{p\to\infty} \|X\|_{L_p} = \|X\|_{L_\infty}$

2 Practice with Matrix Norms.

a **The Induced Operator Norm.** We define the induced operator norm as

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

a_1 **Equivalent Definitions.** At first we will show that the norm is attained on the unit sphere w.r.t. to the ℓ_p norm. Since the unit ball $\{\|x\|_p \le 1\}$ is compact and $x \to Ax$ is a continuous map, the sup is attained at some point. Let's call the point x_0. Suppose x_0 lies strictly inside the unit ball, i.e. $\|x_0\|_p < 1$. Then $y = x/\|x_0\|_p$ has ℓ_p norm 1 and $\|Ay\|_q = (1/\|x_0\|_p)\|Ax\|_q > \|Ax\|_q$, i.e., $\exists y \in B^n_{\ell_p}$ such that $\|Ay\|_q > \|Ax_0\|_q$, this is clearly not feasible since x_0 is the point of sup. Therefore, x_0 has to lie on the boundary of $B^n_{\ell_p}$, i.e. $\|x_0\|_p = 1$. Hence, $\|A\|_{p\to q} = \sup_{\|x\|_p=1} \|Ax\|_q$. The other equivalence is just a restatement of the positive homogeneity, i.e. $\sup_x \|Ax\|_q/\|x\|_p = \sup_x \|A(x/\|x\|_p)\|_q = \sup_{\|x\|_p=1} \|Ax\|_q$, since $\|x/\|x\|_p\|_p = 1$.

a_2 **Submultiplicativity.** Take $x \in \mathbb{R}^k$. Then

$$\|AB\|_{p\to r} = \sup_{\|x\|_p=1} \|ABx\|_r = \sup_{\|x\|_p=1} \|A(Bx)\|_r \le \|A\|_{p\to q} \sup_{\|x\|_p=1} \|Bx\|_r \le \|A\|_{p\to q}\|B\|_{q\to r}.$$

a_3 **Duality form and transpose relationship.** The dual norm identity on the ℓ_q norm gives us $\|u\|_q = \sup_{\|y\|_{q'}} |y^\top u| = \sup_{\|y\|_{q'}} y^\top u$ [we can remove the absolute sign since flipping $y \to -y$ does not change the sup]. Therefore,

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

Hence,

$$\begin{aligned}
\|A\|_{p\to q} &= \sup_{\|x\|_p=1} \sup_{\|y\|_{q'}=1} y^\top Ax \\
&= \sup_{\|x\|_p=1} \sup_{\|y\|_{q'}=1} x^\top A^\top y \\
&= \sup_{\|y\|_{q'}=1} \sup_{\|x\|_p=1} x^\top A^\top y \quad \text{(the suprema may be interchanged by compactness)} \\
&= \sup_{\|y\|_{q'}=1} \|A^\top y\|_{p'} \quad \text{(definition of the dual norm)} \\
&= \|A^\top\|_{q'\to p'} \quad \text{(definition of the operator norm).}
\end{aligned}$$

When $p = q = 2$, $p' = q' = 2$ [p, q are Hölder conjugate of themselves], hence $\|A^\top\|_{2\to 2} = \|A\|_{2\to 2}$.

b **Spectral versus Frobenius norms.** Here we write $\|\cdot\| = \|\cdot\|_{2\to 2}$ for the spectral norm and the Frobenius norm $\|A\|_F = \left(\sum_{i,j} A_{i,j}^2\right)^{1/2}$.

b_1 **Rank-one and diagonal matrices.** Let $u \in \mathbb{R}^m, v \in \mathbb{R}^n$. Using the dual norm porperty we write,

$$\|uv^\top\| = \|uv^\top\|_{2\to 2} = \sup_{\|y\|_2=1} \sup_{\|x\|_2=1} y^\top uv^\top x \geq \frac{u^\top}{\|u\|_2} uv^\top \frac{v}{\|v\|_2} = \|u\|_2\|v\|_2,$$

where the inequality follows from taking a specific choice of x, y. On the other hand, $\|uv^\top\|_{2\to 2} \leq \|u\|_{2\to 2}\|v^\top\|_{2\to 2} = \|u\|_2\|v\|_2$ [By parts (a2) and (a3)]. Hence, $\|uv^\top\| = \|u\|_2\|v\|_2$. Also,

$$\|u\|_2\|v\|_2 = \left(\sum_i u_i^2\right)^{1/2} \left(\sum_j v_j^2\right)^{1/2} = \left(\left(\sum_i u_i^2\right)\left(\sum_j v_j^2\right)\right)^{1/2} = \left(\sum_{i,j} (u_i v_j)^2\right)^{1/2} = \|uv^\top\|_F.$$

$D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$. Then writing $x = \{x_1, x_2, \ldots, x_n\}$

$$\|D\| = \sup_{\|x\|_2=1} \|Dx\|_2 = \sup_{\|x\|_2=1} \left(\sum_{i=1}^n (d_i x_i)^2\right)^{1/2} \leq \max_{1\leq i\leq n} |d_i| \sup_{\|x\|_2=1} \|x\|_2 = \max_{1\leq i\leq n} |d_i| = \|D\|_\infty.$$

The equality is attained for $x = e_k$ where $|d_k| = \max_{1\leq i\leq n} |d_i|$, e_k being the standard unit vector in the $k-$th direction. Hence, $\|D\| = \max_{1\leq i\leq n} |d_i| = \|D\|_\infty$. It is easy to see that $\|D\|_F = \left(\sum_{i=1}^n d_i^2\right)^{1/2} = \|D\|_2$.

b_2 **General Relationship.** At first we note that the largest singular value of A, $\sigma_1 = \|A\|$. This can be seen by writing the SVD of $A = U\Sigma V^\top$ and by noting that the spectral norm is invariant under pre/post multiplication by unitary matrices. Hence, $\sigma_1 = \max_{1\leq i\leq r} \sigma_i = \|\Sigma\| = \|U\Sigma V^\top\| = \|A\|$. Therefore,

$$\|A\| = \sigma_1 \leq \left(\sum_{i=1}^r \sigma_i^2\right)^{1/2} = \|A\|_F \leq \left(\sum_{i=1}^r \max_{1\leq i\leq r} \sigma_i^2\right)^{1/2} = \sqrt{r}\sigma_1 = \sqrt{r}\|A\|.$$

b_3 **Frobenius submultiplicativity.** Using the hint we write

$$\|AB\|_F^2 = \operatorname{tr}(B^\top A^\top AB) \leq \operatorname{tr}(B^\top \|A\|^2 IB) = \|A\|^2 \operatorname{tr}(B^\top B) = \|A\|^2\|B\|_F^2.$$

Simlarly, using the cyclic property of tr we get

$$\|AB\|_F^2 = \operatorname{tr}(B^\top A^\top AB) = \operatorname{tr}(A^\top ABB^\top) \leq \operatorname{tr}(A^\top A\|B^\top\|^2 I) = \|B^\top\|^2 \operatorname{tr}(A^\top A) = \|B|^2\|A\|_F^2.$$

Computing the square root we get

$$\|AB\|_F \leq \|A\|\|B\|_F \quad \text{and} \quad \|AB\|_F \leq \|A\|_F\|B\|.$$

From part (b2) we know that $\|A\| \leq \|A\|_F$, hence $\|AB\|_F \leq \|A\|\|B\|_F \leq \|A\|_F\|B\|_F$ i.e. the Frobenius norm is sub-multiplicative.

b_4 **Orthonogal Invariance.** We know that for an orthogonal matrix U, $\|Ux\| = \|x\|$, this follows from the fact that the coloumns of U have ℓ_2 norm 1. For orthogonal matrices $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$, we have

$$\|UAV\| = \sup_{\|x\|_2=1} \|UAVx\|_2 = \sup_{\|x\|_2=1} \|UA(Vx)\| = \sup_{\|y\|_2=1} \|UAy\| = \sup_{\|y\|=1} \|Ay\| = \|A\|.$$

By using $\|A\|_F^2 = \mathrm{tr}(A^\top A)$, we get

$$\|UAV\|_F^2 = \mathrm{tr}(V^\top A^\top U^\top UAV) = \mathrm{tr}(V^\top A^\top AV) = \mathrm{tr}(A^\top AVV^\top) = \mathrm{tr}(A^\top A) = \|A\|_F^2,$$

where the third equality follows from the cyclic property of the tr, the other equalities follows by noting that $U^\top U = I_m, VV^\top = I_n$. Hence, $\|UAV\|_F = \|A\|_F$.

3 Practice with variance and Covariance Identities.

$Z \in \mathbb{R}^d$ is a random variable with mean $\mathbb{E}Z = m \in \mathbb{R}^d$. The expression $\mathbb{E}[\|Z - \mathbb{E}Z\|_2^2]$ is sometimes called the total variance.

a **Variance Identities.**

a_1 **Variance via pythagorean decomposition.** We assume $\mathbb{E}Z^2 < \infty$ and $\mathbb{E}Z = m$. Then

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

More generally, for any $a \in \mathbb{R}^d$,

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

Hence, $\min_{a \in \mathbb{R}^d} \mathbb{E}[\|Z - a\|_2^2] = \mathbb{E}[\|Z - m\|_2^2]$.

a_2 **Variance via symmetrization.** We assume Z' is an independent copy of Z, i.e. $Z' \perp Z$ and $Z' \overset{\mathrm{d}}{=} Z$. Then

$$
\begin{aligned}
\mathbb{E}[\|Z - Z'\|_2^2] &= \mathbb{E}[\|(Z - m) - (Z' - m)\|_2^2] \\
&= \mathbb{E}[\|Z - m\|_2^2] + \mathbb{E}[\|Z' - m\|_2^2] - 2\mathrm{Corr}(Z - m, Z' - m) \\
&= 2\mathbb{E}[\|Z - m\|_2^2] \quad [\text{since } Z' \overset{\mathrm{d}}{=} Z].
\end{aligned}
$$

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

a_3 **Variance of Independent Sums.** $Z_1, Z_2, \dots Z_k$ are independent random vectors with $\mathbb{E}Z_i = 0$ and

$\mathbb{E}\|Z_i\|_2^2 < \infty$. So, we have

$$\begin{aligned}
\mathbb{E}\left\|\sum_{i=1}^k Z_i\right\|_2^2 &= \mathbb{E}\left\|\sum_{i=1}^k Z_i - \mathbb{E}\sum_{i=1}^k Z_i\right\|_2^2 \\
&= \mathrm{Var}\left(\sum_{i=1}^k Z_i\right) \\
&= \sum_{i=1}^k \mathrm{Var}(Z_i) + 2\sum_{1\le i<j\le k} \mathrm{Cov}(Z_i, Z_j) \\
&= \sum_{i=1}^k \mathrm{Var}(Z_i) \quad \text{[Since } Z_i \text{ s are all independent]} \\
&= \sum_{i=1}^k \left[\mathbb{E}\|Z_i\|_2^2 - \|\mathbb{E}Z_i\|_2^2\right] \\
&= \sum_{i=1}^k \left[\mathbb{E}\|Z_i\|_2^2\right] \quad \text{[Since } \mathbb{E}Z_i = 0 \text{ for all } i]
\end{aligned}$$

b **Covariance Identities.** For a random vector $Z \in \mathbb{R}^d$ with $m = \mathbb{E}Z$ we define the covariance matrix as

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

b_1

$$\begin{aligned}
Cov(Z) &= \mathbb{E}[(Z-m)(Z-m)^t] \\
&= \mathbb{E}[ZZ^t - mZ^t - Z^T m + mm^t] \\
&= \mathbb{E}[ZZ^t] - m[\mathbb{E}Z]^t - [\mathbb{E}Z]^t m + mm^t \quad \text{[Since } \mathbb{E} \text{ and } (\cdot)^t \text{ commute]} \\
&= \mathbb{E}[ZZ^t] - mm^t.
\end{aligned}$$

b_2 For any vector $v \in \mathbb{R}^d$, $\mathbb{E}[v^t Z] = v^t\mathbb{E}[Z] = v^t m$. Hence,

$$\begin{aligned}
\mathrm{Var}(v^t Z) &= \mathbb{E}[(v^t Z - v^t m)(v^t Z - v^t m)] \\
&= \mathbb{E}[(v^t Z - v^t m)(v^t Z - v^t m)^t] \quad \text{[Since } v^t Z - v^t m \text{ is a scalar]} \\
&= \mathbb{E}[v^t (Z-m)(Z-m)^t v] \\
&= v^t \mathbb{E}[(Z-m)(Z-m)^t] v \\
&= v^t \mathrm{Cov}(Z) v.
\end{aligned}$$

b_3

$$
\begin{aligned}
\operatorname{tr}(\operatorname{Cov}(Z)) &= \operatorname{tr}\left[\mathbb{E}[(Z-m)(Z-m)^t]\right] \\
&= \mathbb{E}\left[\operatorname{tr}((Z-m)(Z-m)^t)\right] \quad \text{[Since tr and } \mathbb{E} \text{ can commute]} \\
&= \mathbb{E}\left[\|Z-m\|_2^2\right] \quad \text{[Since } \operatorname{tr}[uu^t] = \|u\|_2^2\text{]} \\
&= \mathbb{E}\left[\|Z-\mathbb{E}Z\|_2^2\right].
\end{aligned}
$$

4 Practice with Moving between Moments and Tails.

a **From Tails to Moments.** $X \geq 0$ is a non-negative random variable.

a_1 **Moment Identity using Tail Integral.** Fox $x \geq 0$. Then, $\mathbf{1}_{t \leq x} = 1$ for $t \leq x$ and is 0 otherwise.

Therefore, $x = \int_0^\infty \mathbf{1}_{x>t} dt$. Then

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x\mathbb{P}(X = x)dx \\
&= \int_0^\infty \int_0^\infty \mathbb{1}_{x>t}\mathbb{P}(X = x)dtdx \\
&= \int_0^\infty \left[\int_0^\infty \mathbb{1}_{x>t}\mathbb{P}(X = x)dx\right] dt \quad \text{[interchanging integrals by Tonelli's Theorem, since both are integrable]} \\
&= \int_0^\infty \left[\int_t^\infty \mathbb{P}(X = x)dx\right] dt \\
&= \int_0^\infty \mathbb{P}[X > t]dt.
\end{aligned}$$

More generally, we can write $x^p = \int_0^\infty pt^{p-1}\nVdash_{x>t}dt$. Therefore,

$$\begin{aligned}
\mathbb{E}[X^p] &= \int_0^\infty x^p\mathbb{P}[X = x]dx \\
&= \int_0^\infty \int_0^\infty pt^{p-1}\mathbb{1}_{x>t}\mathbb{P}[X = x]dtdx \\
&= \int_0^\infty pt^{p-1}\left[\int_0^\infty \mathbb{1}_{x>t}\mathbb{P}[X = x]dx\right] dt \quad \text{[Similarly, by applying Tonelli's Theorem]} \\
&= \int_0^\infty pt^{p-1}\left[\int_t^\infty \mathbb{P}[X = x]dx\right] dt \\
&= \int_0^\infty pt^{p-1}\mathbb{P}[X > t]dt.
\end{aligned}$$

a_2 **Moment growth bounds from exponential tail behavior.** We assume $\exists c, C, \alpha > 0$ such that for all $t > 0$

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

Then we have by Part (a1)

$$\begin{aligned}
\mathbb{E}[X^p] &= \int_0^\infty pt^{p-1}\mathbb{P}[X>t]\,dt \\
&\le \int_0^\infty pt^{p-1}Ce^{-ct^\alpha}\,dt \\
&= \frac{pC}{\alpha}c^{-p/\alpha}\int_0^\infty z^{\frac{p}{\alpha}-1}e^{-z}\,dz. \\
&= \frac{C}{\alpha}c^{-p/\alpha}p\Gamma(p/\alpha) \\
&= \frac{C}{\alpha}c^{-p/\alpha}\Gamma(p/\alpha+1) \\
&\le \frac{C}{\alpha}c^{-p/\alpha}\left(C_0(p/\alpha)\right)^{p/\alpha}.
\end{aligned}$$

Hence,

$$\|X\|_{L_p} = \mathbb{E}[X^p]^{1/p} \le \left[\frac{C}{c}\right]^{1/p}\left[\frac{C_0}{\alpha}\right]^{1/\alpha}p^{1/\alpha} \le \left[\frac{C}{c}\right]\left[\frac{C_0}{\alpha}\right]^{1/\alpha}p^{1/\alpha} \le C'p^{1/\alpha}$$

where $C' = \left[\frac{C}{c}\right]\left[\frac{C_0}{\alpha}\right]^{1/\alpha}$. The last inequality follows from by choosing $C > c$ and hence $\left[\frac{C}{c}\right]^{1/p} \le \left[\frac{C}{c}\right]$ for $p \ge 1$.

- For $\alpha = 2$, X behaves like sub-Gaussian, since it's tail $\mathbb{P}[X>t] \le Ce^{-\alpha t^2}$ and it's moment grows in the order of $\sqrt{p}$, since $\|X\|_{L_p} \le C'\sqrt{p}$, [like a random normal variable where the moment grows in $\mathcal{O}\sqrt{p})$].
- For $\alpha = 1$, X behaves like sub-Exponential, since it's tail $\mathbb{P}[X>t] \le Ce^{-\alpha t}$ and it's moment grows in the order of p, since $\|X\|_{L_p} \le C'p$, [like a random exponential variable where the moment grows in $\mathcal{O}(p)$].

b **Classical tail bounds from second moments.**

b_1 **Upper Bound: Cantelli's Inequality.** Fix $t>0$. For any $a>0$,

$$\begin{aligned}
\mathbb{P}[Y-\mathbb{E}Y>t] = \mathbb{P}[Y-\mathbb{E}Y+a > Y-\mathbb{E}Y+a] &\le \mathbb{P}[(Y-\mathbb{E}Y+a)^2 \ge (t+a)^2] \\
&\le \frac{\mathbb{E}[(Y-\mathbb{E}Y+a)^2]}{(t+a)^2} \quad \text{[Using Chebyshev's Inequality]} \\
&= \frac{\mathrm{Var}(Y)+a^2}{(t+a)^2}.
\end{aligned}$$

The LHS on the last expression is independent of a. So, we can minimize the expression $\frac{\mathrm{Var}(Y)+a^2}{(t+a)^2}$ w.r.t. a and still the inequality holds. Let $\mathrm{Var}(Y) = \sigma^2$. Also, let $f(a) = \frac{\sigma^2+a^2}{(t+a)^2}$. It is easy to see that f is minimized at $a = \frac{\sigma^2}{t}$. And $f(\sigma^2/t) = \frac{\sigma^2}{\sigma^2+t^2}$. Hence,

$$\mathbb{P}[Y - \mathbb{E}Y \geq t] \leq \frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y) + t^2}.$$

b_2 **Comparison with Chebyshev's inequality.** Using Chebyshev's inequality, we get

$$\mathbb{P}[Y - \mathbb{E}Y \geq t] \leq \mathbb{P}[|Y - \mathbb{E}Y| \geq t] \leq \frac{\mathrm{Var}(Y)}{t^2}.$$

Since $\frac{\mathrm{Var}(Y)}{t^2} > \frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y)+t^2}$, Cantelli's inequality is strictly tighter than Chebyshev's inequality. In fact proving Cantelli's inequality, we have used Chebyshev's inequality. So, Cantelli's inequality is philosophically expected to be stronger. Applying Cantelli's inequality to $-Y$, we get

$$\mathbb{P}[-Y + \mathbb{E}Y \geq t] \leq \frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y) + t^2} \implies \mathbb{P}[Y - \mathbb{E}Y \leq -t] \leq \frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y) + t^2}.$$

Combining Cabtelli's inequality for Y and $-Y$ we get

$$\mathbb{P}[|Y - \mathbb{E}Y| \geq t] \leq \frac{2\mathrm{Var}(Y)}{\mathrm{Var}(Y) + t^2}. \tag{2}$$

For $t^2 > \mathrm{Var}(Y)$, $\frac{2\mathrm{Var}(Y)}{\mathrm{Var}(Y)+t^2} > \frac{\mathrm{Var}(Y)}{t^2}$, i.e. Chebyshev's inequality is stronger than 2, while for $t^2 < \mathrm{Var}(Y)$, $\frac{2\mathrm{Var}(Y)}{\mathrm{Var}(Y)+t^2} < \frac{\mathrm{Var}(Y)}{t^2}$, i.e. 2 is stronger than Chebyshev's inequality.

b_3 **Lower Bound: Paley-Zygmund Inequality.** We decompose the expression of $\mathbb{E}X$ in the following way:

$$\begin{aligned}
\mathbb{E}X &= \mathbb{E}[X\mathbb{1}_{X<\theta\mathbb{E}X}] + \mathbb{E}[X\mathbb{1}_{X\geq\theta\mathbb{E}X}] \\
&\leq \mathbb{E}[\theta\mathbb{E}X] + (\mathbb{E}[X^2])^{1/2}\left(\mathbb{P}(X \geq \theta\mathbb{E}X)\right)^{1/2}.
\end{aligned}$$

The first inequality follows since $X\mathbb{1}_{X<\theta\mathbb{E}X} \leq \theta\mathbb{E}X$ and the second inequality follows from Cuchy-Schwarz inequality. By interchanging sides, we get

$$\begin{aligned}
\mathbb{P}(X \geq \theta\mathbb{E}X) &\geq \left[(1-\theta)\frac{\mathbb{E}X}{\mathbb{E}[X^2]^{1/2}}\right]^2 \\
&= (1-\theta)^2\frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]} \\
&= (1-\theta)^2\frac{(\mathbb{E}X)^2}{\mathrm{Var}(X) + (\mathbb{E}X)^2} \\
&= (1-\theta)^2\frac{1}{1+\frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2}} \\
&\geq \frac{(1-\theta)^2}{1+c} \quad [\text{If } \mathrm{Var}(X) \leq c(\mathbb{E}X)^2].
\end{aligned}$$

5 Mean vs Median

a **Mean-Median Closeness.** Cantellii's inequality says that for $\lambda > 0$ $\mathbb{P}(|X - \mathbb{E}X| > \lambda) \leq \frac{\text{var(X)}}{\text{var}(X)+\lambda^2}$ and for $\lambda < 0$, $\mathbb{P}(|X - \mathbb{E}X| < -\lambda) \leq \frac{\text{var(X)}}{\text{var}(X)+\lambda^2}$. Using the definition of Median and for $M_Z \geq \mathbb{E}Z$, we have

$$\frac{1}{2} \leq \mathbb{P}(Z - \mathbb{E}Z \geq M_Z - \mathbb{E}Z) \leq \frac{\text{var(Z)}}{\text{var}(Z) + (M_Z - \mathbb{E}Z)^2}.$$

Similarly for $M_Z \leq \mathbb{E}Z$, we have

$$\frac{1}{2} \leq \mathbb{P}(Z - \mathbb{E}Z \leq M_Z - \mathbb{E}Z) \leq \frac{\text{var(Z)}}{\text{var}(Z) + (M_Z - \mathbb{E}Z)^2}.$$

In either of the cases,

$$(M_Z - \mathbb{E}Z)^2 \leq \text{var}(Z)$$

$$|M_Z - \mathbb{E}Z| \leq \sqrt{\text{var}(Z)}.$$

b **Mean-Median closeness using variance proxy.** X is a real valued random variable with median M_X. Assume $\exists a, b > 0$ such that $\forall t > 0$

$$\mathbb{P}(|X - M_X| > t) \leq ae^{-t^2/b}.$$

The function $f(x) = |x|$ is convex and hence using Jensen's inequality on $|X - M_X|$, we get $|\mathbb{E}X - M_X| = |\mathbb{E}(X - M_X)| \leq \mathbb{E}|X - M_X|$. Now,

$$\begin{aligned}
\mathbb{E}|X - M_X| &= \int_0^\infty \mathbb{P}(|X - M_X| > t)dt \\
&\leq \int_0^\infty ae^{-t^2/b}dt \\
&= \frac{a\sqrt{b}}{\sqrt{2}}\Gamma\left(\frac{1}{2}\right) = a\sqrt{b\pi/2}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}|X - M_X| &\leq \sqrt{\mathbb{E}|X - M_X|^2} \quad [\text{as var}(|X - M_X|) \geq 0] \\
&= \sqrt{\int_0^\infty 2t\mathbb{P}(|X - M_X| > t)dt} \\
&\leq \sqrt{\int_0^\infty 2tae^{-t^2/b}dt} \\
&= \sqrt{ab\int_0^\infty e^{-z}dz} \\
&= \sqrt{ab}.
\end{aligned}$$

Combining these two cases, we get

$$|\mathbb{E}X - M_X| \leq \mathbb{E}|X - M_X| \leq \min\left\{\sqrt{ab}, a\sqrt{b\pi/2}\right\}.$$