Review: Mid Exam

SDS 391P.6, Spring 2026

Pratik Patil

These notes are a work in progress and are provided as-is for instructional purposes only. They are not (yet) at the level of a scholarly document. In particular, the notes draw from various sources and do not (yet) have sufficient references to the original sources. Additionally, almost surely the notes have errors and they are only probably approximately correct. The notes will be updated regularly as the course progresses. Last updated: 2026-03-10.

1 Overview

This note is meant as a (non-exhaustive) review of the material covered up to Lecture 6 or Homework 3. We collect (some of) the main facts that may be useful on the midterm.

A recurring theme in the course so far has been:

$$\text{structure of the random quantity} + \text{the right inequality/tool} \implies \text{variance / tail / moment control.}$$

For the midterm, the most important skills are:

- recognizing which norm / variance / tail / MGF identity is relevant;
- moving between tails, moments, and exponential moments;
- knowing when to use variance tensorization, Efron–Stein–Steele, Poincaré, or Chernoff;
- carrying out standard calculations cleanly.

2 Norms and matrix norms

2.1 Vector norms

For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $1 \leqslant p < \infty$,

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

If $1 \leqslant p \leqslant q \leqslant \infty$, then

$$\|x\|_q \leqslant \|x\|_p \leqslant n^{\frac{1}{p} - \frac{1}{q}} \|x\|_q.$$

Two standard extremizers:

- equality on the left is attained by one-sparse vectors;
- equality on the right is attained by flat vectors such as $(1, \ldots, 1)$.

Also,

$$\|x\|_\infty \leqslant \|x\|_p \leqslant n^{1/p} \|x\|_\infty, \qquad \|x\|_p \to \|x\|_\infty \quad \text{as } p \to \infty.$$

2.2 Hölder and dual norms

If $p, p' \in [1, \infty]$ satisfy $\frac{1}{p} + \frac{1}{p'} = 1$, then

$$|\langle x, y\rangle| \leqslant \|x\|_p \|y\|_{p'}.$$

In particular, for $p = 2$ this becomes Cauchy–Schwarz:

$$|\langle x, y\rangle| \leqslant \|x\|_2 \|y\|_2.$$

The dual of ℓ_p is $\ell_{p'}$:

$$\|x\|_p = \sup_{\|y\|_{p'} \leqslant 1} \langle x, y\rangle.$$

2.3 Matrix norms

For $A \in \mathbb{R}^{m\times n}$, the induced operator norm is

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

Equivalent forms:

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

Composition bound:

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

The most important special cases are:

$$\|A\| := \|A\|_{2\to 2}, \qquad \|A\|_F := \Big(\sum_{i,j} A_{ij}^2\Big)^{1/2}.$$

Useful identities:

$$\|A\| = \sigma_1(A), \qquad \|A\|_F^2 = \sum_{i=1}^{\operatorname{rank}(A)} \sigma_i(A)^2 = \operatorname{tr}(A^\top A).$$

Also,

$$\|A\| \leqslant \|A\|_F \leqslant \sqrt{\operatorname{rank}(A)} \, \|A\|.$$

Rank-one case: if $A = uv^\top$, then

$$\|A\| = \|A\|_F = \|u\|_2 \|v\|_2.$$

If $D = \operatorname{diag}(d_1, \ldots, d_n)$ is diagonal, then

$$\|D\| = \max_i |d_i|, \qquad \|D\|_F = \Big(\sum_i d_i^2\Big)^{1/2}.$$

3 Probability identities

3.1 Variance and covariance

For a real-valued random variable X,

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2] - (\mathbb{E}X)^2.$$

For random variables X, Y,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}[XY] - \mathbb{E}X\,\mathbb{E}Y.$$

Variance of a linear combination:

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y).$$

If $X_1, \ldots, X_n$ are independent (or just pairwise uncorrelated), then

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

3.2 Variance identities

For any square-integrable random variable Z,

$$\mathrm{Var}(Z) = \min_{a\in\mathbb{R}} \mathbb{E}[(Z - a)^2],$$

with unique minimizer $a = \mathbb{E}Z$.

If Z' is an independent copy of Z, then

$$\mathrm{Var}(Z) = \frac{1}{2}\,\mathbb{E}[(Z - Z')^2].$$

For a random vector $X \in \mathbb{R}^d$,

$$\mathrm{Cov}(X) = \mathbb{E}[(X - \mathbb{E}X)(X - \mathbb{E}X)^\top] = \mathbb{E}[XX^\top] - \mathbb{E}X(\mathbb{E}X)^\top.$$

Also, for every $v \in \mathbb{R}^d$,

$$\mathrm{Var}(v^\top X) = v^\top\,\mathrm{Cov}(X)\,v.$$

And

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

3.3 Integrated tail identities

For a nonnegative random variable X,

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

More generally, for $p > 0$,

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

Equivalently, for a general real-valued random variable,

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

These are often the fastest way to move from tails to moments.

3.4 Markov, Chebyshev, Cantelli, Paley–Zygmund

Markov: if $Y \geqslant 0$, then for $t > 0$,

$$\mathbb{P}\{Y \geqslant t\} \leqslant \frac{\mathbb{E}Y}{t}.$$

Chebyshev:

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

Cantelli (one-sided Chebyshev):

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

3.5 Mean versus median

If M_X is a median of X, then

$$|M_X - \mathbb{E}X| \leqslant \sqrt{\mathrm{Var}(X)}.$$

A quick way to prove this is via Cantelli's inequality.

4 Variance bounding methods

4.1 Tensorization of variance

If $Z = f(X_1, \ldots, X_n)$ with independent $X_1, \ldots, X_n$, and

$$X^{(i)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n), \qquad \mathbb{E}^{(i)}[\cdot] = \mathbb{E}[\cdot \mid X^{(i)}],$$

then

$$\mathrm{Var}(Z) \leqslant \sum_{i=1}^n \mathbb{E}\big[\mathrm{Var}(Z \mid X^{(i)})\big].$$

This is the starting point for bounded differences and Efron–Stein–Steele.

4.2 Bounded differences variance bound

If changing only coordinate i can change f by at most c_i, then

$$\mathrm{Var}(f(X_1, \ldots, X_n)) \leqslant \frac{1}{4} \sum_{i=1}^n c_i^2.$$

This is often the easiest variance bound for nonlinear functions.

4.3 Efron–Stein–Steele

If X_i' is an independent copy of X_i and

$$Z^{(i)} := f(X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n),$$

then

$$\mathrm{Var}(Z) \leqslant \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}[(Z - Z^{(i)})^2].$$

There is also the one-sided form:

$$\mathrm{Var}(Z) \leqslant \sum_{i=1}^{n} \mathbb{E}[(Z - Z^{(i)})_+^2].$$

This one-sided version is especially useful for suprema.

4.4 Gaussian and convex Poincaré

Gaussian Poincaré: if $g \sim \mathcal{N}(0, I_n)$ and f is smooth enough, then

$$\mathrm{Var}(f(g)) \leqslant \mathbb{E}\|\nabla f(g)\|_2^2.$$

Convex Poincaré (bounded coordinates): if $X_1, \dots, X_n \in [0, 1]$ are independent and f is separately convex, then

$$\mathrm{Var}(f(X)) \leqslant \mathbb{E}\|\nabla f(X)\|_2^2.$$

A standard use: if f is L-Lipschitz, then

$$\|\nabla f(x)\|_2 \leqslant L \quad \text{(a.e.)},$$

so

$$\mathrm{Var}(f(X)) \leqslant L^2.$$

5 Exponential moment methods: the key formulas

5.1 MGF and CGF

For a random variable X, define

$$m_X(\lambda) = \mathbb{E}e^{\lambda X}, \qquad \kappa_X(\lambda) = \log \mathbb{E}e^{\lambda X}.$$

The centered log-MGF (CGF) is

$$\psi_X(\lambda) = \log \mathbb{E}e^{\lambda(X - \mathbb{E}X)}.$$

For independent sums,

$$\psi_{\sum_i X_i}(\lambda) = \sum_i \psi_{X_i}(\lambda).$$

This additivity is the reason MGFs are so useful for sums.

5.2 Chernoff bound

For every $\lambda \geqslant 0$,

$$\mathbb{P}\{X - \mathbb{E}X \geqslant t\} \leqslant \exp\big(-\lambda t + \psi_X(\lambda)\big).$$

Optimizing over $\lambda \geqslant 0$ gives the best Chernoff bound. To get the lower tail, apply the same argument to $-X$.

5.3 Hoeffding's lemma and Hoeffding's inequality

If $a \leqslant X \leqslant b$ almost surely, then

$$\psi_X(\lambda) \leqslant \frac{\lambda^2(b-a)^2}{8} \qquad \text{for } \lambda \in \mathbb{R}.$$

So bounded random variables are sub-Gaussian.

If $X_1, \ldots, X_n$ are independent and $a_i \leqslant X_i \leqslant b_i$, then

$$\mathbb{P}\left\{\left|\sum_{i=1}^n X_i - \mathbb{E}\sum_{i=1}^n X_i\right| \geqslant t\right\} \leqslant 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Equivalently, with

$$v = \frac{1}{4}\sum_{i=1}^n (b_i - a_i)^2,$$

one can write

$$\mathbb{P}\{|S_n - \mathbb{E}S_n| \geqslant t\} \leqslant 2e^{-t^2/(2v)}.$$

6 Sub-Gaussian and sub-exponential behaviors

6.1 Sub-Gaussian

A centered random variable Z is sub-Gaussian if

$$\log \mathbb{E}e^{\lambda Z} \leqslant \frac{\sigma^2\lambda^2}{2} \qquad \text{for all } \lambda \in \mathbb{R}.$$

Equivalent viewpoints (up to constants):

- Gaussian tail:
$$\mathbb{P}\{|Z| \geqslant t\} \leqslant 2e^{-ct^2/\sigma^2}.$$
- Moment growth:
$$\|Z\|_{L^p} \lesssim \sigma\sqrt{p} \qquad \text{for } p \geqslant 1.$$

The Orlicz norm is

$$\|Z\|_{\psi_2} = \inf\Big\{K > 0 : \ \mathbb{E}e^{Z^2/K^2} \leqslant 2\Big\}.$$

Up to constants,

$$\|Z\|_{\psi_2} \asymp \sup_{p \geqslant 1} \frac{\|Z\|_{L^p}}{\sqrt{p}}.$$

6.2 Sub-exponential

A centered random variable Z is sub-exponential if

$$\|Z\|_{\psi_1} = \inf\left\{K > 0 : \ \mathbb{E}e^{|Z|/K} \leqslant 2\right\} < \infty.$$

Equivalent viewpoints (up to constants):

- Exponential tail:
$$\mathbb{P}\{|Z| \geqslant t\} \leqslant 2e^{-ct/K};$$
- Moment growth:
$$\|Z\|_{L^p} \lesssim Kp.$$

A key difference from the sub-Gaussian case: the CGF control is only local:

$$\log \mathbb{E}e^{\lambda Z} \leqslant C\lambda^2 K^2 \qquad \text{for } |\lambda| \leqslant c/K.$$

6.3 Closure properties

If X is sub-Gaussian, then X^2 is sub-exponential and

$$\|X^2\|_{\psi_1} \lesssim \|X\|_{\psi_2}^2.$$

If X, Y are sub-Gaussian, then XY is sub-exponential and

$$\|XY\|_{\psi_1} \lesssim \|X\|_{\psi_2}\|Y\|_{\psi_2}.$$

7 Bernstein inequality and the two-regime picture

If $X_1, \ldots, X_n$ are independent, mean-zero sub-exponential random variables with

$$K_i = \|X_i\|_{\psi_1}, \qquad V = \sum_{i=1}^n K_i^2, \qquad B = \max_i K_i,$$

then

$$\mathbb{P}\left\{\left|\sum_{i=1}^n X_i\right| \geqslant t\right\} \leqslant 2\exp\left[-c\min\left(\frac{t^2}{V}, \frac{t}{B}\right)\right].$$

This is the most important tail bound for sums of sub-exponential random variables. You should remember the *two regimes*:

$$\text{Gaussian regime: } e^{-ct^2/V}, \qquad \text{Exponential regime: } e^{-ct/B}.$$

This same two-term structure also appears in L^p bounds:

$$\left\|\sum_{i=1}^n X_i\right\|_{L^p} \lesssim \sqrt{p}\sqrt{V} + p\,B.$$