High Dimensional Probability Summary

Chapter 1 - Probability Theory Preliminaries

① **Lemma (Integral Identity):** $X \in \mathbb{R}_{\geq 0}$ R.V. $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t)\,dt$

Pf: $x \in \mathbb{R}_+$ $\quad x = \int_0^x 1\,dt = \int_0^\infty \mathbb{1}_{\{t<x\}}(t)\,dt \xRightarrow{\text{take expectation},\ x \to X} \mathbb{E}[X] = \int_0^\infty \mathbb{E}[\mathbb{1}_{\{t<x\}}(t)]\,dt = \int_0^\infty \mathbb{P}(t < X)\,dt$ [integral exchange JU]

↳ **Extension:** $X \in \mathbb{R}$, $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X>t)\,dt - \int_{-\infty}^0 \mathbb{P}(X<t)\,dt$ **Pf:** if $x<0$ $x = -\int_x^0 1\,dt = \int_{-\infty}^0 \mathbb{1}_{\{x<t<0\}}\,dt$

let $t^p = u \Rightarrow pt^{p-1} = u'$

↳ **p-th moment:** $\mathbb{E}[|X|^p] = \int_0^\infty \mathbb{P}(|X|^p > u)\,du = \int_0^\infty \mathbb{P}(|X| > u^{\frac{1}{p}})\,du = \int_0^\infty p t^{p-1}\,\mathbb{P}(|X|>t)\,dt$

② **Prop (Jensen's Inequality):** $\Phi: I \to \mathbb{R}$ convex, $X \in \mathbb{R}$ R.V. Then $\Phi(\mathbb{E}[X]) \leq \mathbb{E}[\Phi(X)]$

interval $\subset \mathbb{R}$

Pf: Omitted

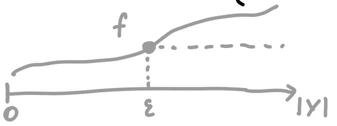
$L^\infty = \{X : \Omega \to \mathbb{R},\ X \text{ measurable},\ \|X\|_{L^\infty} < \infty\}$

③ **L^p spaces:** Take $(\Omega, \mathcal{F}, \mathbb{P})$ to be a prob. space

- $L^p := \{X : \Omega \to [-\infty, \infty]$ measurable and $\|X\|_{L^p} < \infty\}$
- $\|X\|_{L^p} = \left(\int |f|^p d\mathbb{P}\right)^{\frac{1}{p}} = \left(\int |f(x)|^p \mathbb{P}(dx)\right)^{\frac{1}{p}}$
- $\|X\|_{L^p}$ is "an ↑ function" in param $p \Rightarrow \|X\|_{L^p} \leq \|X\|_{L^q}$ for $0 \leq p \leq q \leq \infty$
- $\|X + Y\|_{L^p} \leq \|X\|_{L^p} + \|Y\|_{L^p}$
- $|\mathbb{E}[XY]| \leq \|X\|_{L^2} \|Y\|_{L^2}$

$\|X\|_{L^\infty} := \text{ess-sup}(|X|)$

Smallest # $\alpha \in \mathbb{R}$ s.t. $\{x \in \Omega : X(x) \geq \alpha\}$ has measure zero



④ **Prop (Markov's Inequality):** $Y \in \mathbb{R}$, $f : [0,\infty) \to [0,\infty)$ ↑. Then $\forall \varepsilon > 0$ $\mathbb{P}(|Y| > \varepsilon) \leq \frac{\mathbb{E}[f \circ |Y|]}{f(\varepsilon)}$

Pf: $f \circ |Y|$ is positive R.V. w/ $f(\varepsilon)\mathbb{1}_{\{|Y| \geq \varepsilon\}} \leq f \circ |Y| \Rightarrow f(\varepsilon)\mathbb{P}(|Y| \geq \varepsilon) \leq \mathbb{E}[f \circ |Y|]$

Note: CLT & WLLN give linear rather than exponential tail bounds...

Poisson limit Thm
$X_i^{(N)} \sim \text{Bernoulli}(p_i)$
$S_N = \sum_{i=1}^N X_i$, $p_i \to 0$
$\mathbb{E}[S_N] \to \lambda$
$\Downarrow$
$S_N \to \text{Poi}(\lambda)$

Chapter 2 - Concentration Inequalities of Independent Random variables

$Z \in \{\pm 1\}$ $\mathbb{P}(Z=1) = \mathbb{P}(Z=-1) = \frac{1}{2}$

⑤ **Thm 2.5 (Hoeffding's Inequality):** $X_1, \dots, X_N$ independent Symmetric Bernoulli R.V.s $a \in \mathbb{R}^n$, then $\forall t \geq 0$

$$\mathbb{P}\left(\sum_{i=1}^n a_i X_i \geq t\right) \leq \exp\left(-\frac{t^2}{2\|a\|_2^2}\right)$$

[Bad for small prob success p_i]

Pf: WLOG, set $\|a\|_2 = 1$, e.g. if $\|a\|_2 \neq 1$, put $\tilde{a}_i := \frac{a_i}{\|a\|_2} \Rightarrow \mathbb{P}\left(\sum_{i=1}^n a_i X_i \geq \|a\|_2 t\right) \leq e^{-\frac{t^2}{2}}$

Then markov $e^{\lambda\cdot}$, $\mathbb{E}[e^{\lambda a_i X_i}] = \frac{e^{\lambda a_i} - e^{-\lambda a_i}}{2} = \cosh(\lambda a_i) \leq e^{\lambda^2 a_i^2/2}$ $\left[\cosh(x) = 1 + \frac{x^2}{2!} + \dots\right]$

Get $\mathbb{P}\left(\sum_{i=1}^n a_i X_i \geq t\right) \leq \exp\left(-\lambda t + \frac{\lambda^2 \|a\|_2^2}{2}\right)$ & maximise over $\lambda \Rightarrow \lambda = t$ ■

⑥ **Thm 2.11 (Chernoff's Inequality):** $X_i \sim \text{Ber}(p_i)$ $\quad i = 1, \dots, N$ $\quad S_N = \sum X_i$ $\quad \mu := \mathbb{E}[S_N]$ $\quad \forall t > \mu$

$$\mathbb{P}(S_N \geq t) \leq e^{-\mu}\left(\frac{e\mu}{t}\right)^t$$

Pf: Markov w/ $e^{\lambda x}$ $\lambda > 0$, compute MGF, $1 + x \leq e^x$, optimise over $\lambda \Rightarrow \lambda^* = \log\left(\frac{t}{\mu}\right) \Rightarrow t > \mu$ ■

Note: can show result for the degree of random graphs.

bdd & exists

⑦ **Chernoff Bound:** Suppose $\exists b > 0$ s.t. the centered MGF $\Phi(\lambda) = \mathbb{E}[e^{\lambda(X - \overbrace{\mathbb{E}[X]}^{\mu})}] \in \mathbb{R}_+$ $\forall |\lambda| < b$

Apply markov to $Y := e^{\lambda(X - \mathbb{E}[X])} \Rightarrow \mathbb{P}(X - \mu \geq t) = \mathbb{P}(e^{\lambda(X - \mathbb{E}[X])} \geq e^{\lambda t}) \leq \frac{\Phi(\lambda)}{e^{\lambda t}}$

Then optimise: $\log \mathbb{P}(X - \mu \geq t) \leq \inf_{\lambda \in [0,b]} \{\log \Phi(\lambda) - \lambda t\}$

⑧ **Moments of Normal Distribution:** $X \sim N(0,1) \Rightarrow \|X\|_{L^p} \overset{p \to \infty}{=} O(\sqrt{p})$ clever trick w/ $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\,dt$

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$\|X\|_{L^p} = \left(\mathbb{E}[|X|^p]\right)^{\frac{1}{p}} = \left(\frac{2}{\sqrt{2\pi}} \int_0^\infty x^p e^{-\frac{1}{2}x^2}\,dx\right)^{\frac{1}{p}} = \left(\frac{2}{\sqrt{2\pi}} \int_0^\infty (\sqrt{2}\sqrt{w})^p e^{-w} \frac{\sqrt{2}}{\sqrt{w}} \frac{1}{2}\,dw\right)^{\frac{1}{p}}$

$w = \frac{1}{2}x^2$

$\Gamma(z) \leq z^z$

⑨ **Prop 2.17 (Sub Gaussian Properties):** $X \in \mathbb{R}$. $\exists C_i$ $i \in \{1, \dots, 5\}$ s.t.

(i) Tails of X: $\mathbb{P}(|X| \geq t) \leq 2\exp(-t^2/c_1^2)$ $\forall t \geq 0$

(ii) Moments: $\|X\|_{L^p} \leq C_2 \sqrt{p}$ $\forall p \geq 1$

(iii) MGF X^2: $\mathbb{E}[\exp(\lambda^2 X^2)] \leq \exp(C_3^2 \lambda^2)$ $\forall |\lambda| < \frac{1}{C_3}$

(iv) MGF bd: $\mathbb{E}[\exp(X^2/C_4^2)] \leq 2$

If $\mathbb{E}[X] = 0$, then (i)-(iv) $\Longleftrightarrow$ (v)

(v) MGF bd X: $\mathbb{E}[\exp(\lambda X)] \leq \exp(C_5^2 \lambda^2)$ $\forall \lambda \in \mathbb{R}$

Pf: Hefty... Go through carefully. Use sterling, Γ, power series, markov, optimise etc ■

⑩ **Def 2.18 (Subgaussian R.V. I):** $X \in \mathbb{R}$ R.V. satisfying one of (i)-(iv) is Sub gaussian.
Define s.g. norm $\|X\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}[\exp(X^2/t^2)] \leq 2\}$

Pf: Norm... clever trick, use convexity & take a line...

⑪ **Def 2.22 (Sub Gaussian R.V. II):** $X \in \mathbb{R}$ R.V. w/ $\mu = \mathbb{E}[X]$ is s.g. if $\exists \sigma > 0$ s.t.
$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\lambda^2\sigma^2/2}$ $\forall \lambda \in \mathbb{R}$ σ is the s.g. parameter. $X \sim S.G.(\sigma)$

⑫ **Prop 2.24 (Sums of independent S.G. R.V.s):** $X_1, \dots, X_N$ S.G. w/ params σ_i,
Then $S_N = \sum X_i$ S.G. w/ param $\sigma = \sqrt{\sum \sigma_i^2}$. If mean zero, more...

Pf: Go through.. mostly definition pushing. ■

⑬ **Prop 2.25 (Hoeffding bounds on sums of independent S.G. R.V.s):** $X_1, \dots, X_N$ independent S.G. w/ mean μ_i, param σ_i, then $\forall t \geq 0$

$$\mathbb{P}\left(\sum_{i=1}^{N}(X_i - \mu_i) \geq t\right) \leq \exp\left(\frac{-t^2}{2\sum_{i=1}^{N}\sigma_i^2}\right)$$

Pf: exponential Markov, def & optimisation ■

⑭ **Lemma 2.27 (Centering):** X S.G. Then $X - \mathbb{E}[X]$ S.G. too w/ $\|X - \mathbb{E}[X]\|_{\psi_2} \leq C\|X\|_{\psi_2}$

Pf: $\|\cdot\|_{\psi_2}$ norm $\Rightarrow$ $\|X - \mathbb{E}[X]\|_{\psi_2} \leq \|X\|_{\psi_2} + \|\overbrace{\mathbb{E}[X]}^{\text{number}}\|_{\psi_2}$.

Note: pick $t > \frac{|a|}{\log 2}$ $\forall a \in \mathbb{R}$ and $\mathbb{E}[\exp(a^2/t^2)] \leq 2 \Rightarrow \|a\|_{\psi_2} = \overbrace{\frac{|a|}{\log 2}}^{c|a|}$

So $\|\mathbb{E}[X]\|_{\psi_2} \leq c|\mathbb{E}[X]| \leq c\,\mathbb{E}[|X|] = c\|X\|_{L^1} \overset{L^p \text{ bd}}{\leq} C\underbrace{\|X\|_{\psi_2}}_{\text{may as well}}\sqrt{1}$ ■

⑮ **Prop 2.28 (Sub-Exponential R.V.s):** X $\mathbb{R}$-valued R.V. Following equiv for abs. const.

(i) Tails: very important $\mathbb{P}(|X| > t) \leq 2\exp(-t/C_1)$ (✱) $\forall t \geq 0$

(ii) Moments: stat 1:1 here $\|X\|_{L^p} = \mathbb{E}[|X|^p]^{\frac{1}{p}} \leq C_2 p$ $\forall p \geq 1$

(iii) MGF of $|X|$: $\mathbb{E}[\exp(\lambda \underline{|X|})] \leq \exp(C_3 \lambda)$ $\forall \lambda$ s.t. $0 \leq \lambda \leq \frac{1}{C_3}$

(iv) MGF of $|X|$ bdd at some pt: $\mathbb{E}[\exp(|X| C_4)] \leq 2$

If $\mathbb{E}[X] = 0$, then (i)-(iv) equiv to (v)

(v) MGF X: If $X \sim N(0,1)$ $\mathbb{E}[\exp(\lambda X)] \leq \exp(C_5^2 \lambda^2)$ $\forall \lambda$ s.t. $|\lambda| < \frac{1}{C_5}$

$\mathbb{E}[X] = 0$, $\mathbb{E}[|X|] \neq 0$!!!

Pf: Same as $\overset{\text{symmetry}}{\text{}}$ Subgaussian $\overset{\text{no symmetry}}{\text{}}$ so just copy...

Note: If $X \sim \text{Exponential}(\lambda)$
$\mathbb{P}(X \geq t) = \int_t^\infty \lambda e^{-\lambda t}\, dt = \lambda e^{-\lambda t}$ $\boxed{\text{SAME DECAY AS } (✱)}$

⑯ **Def 2.29 (Sub-Exponential R.V. I):** $X \sim$ Sub-Exponential if it satisfies any of (i)-(iv) above
Set $\|X\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}[\exp(|X|/t)] \leq 2\}$

Pf: Norm prop in support class

⑰ Relationship between Sub-Gaussian & Sub-Exponential R.V.s

- **Lemma 2.31:** X sub gaussian $\iff$ X^2 sub-exponential and $\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$

easy

Pf: X subgaussian $\Rightarrow \mathbb{P}(|X|^2 \geq t) = \mathbb{P}(|X| \geq \sqrt{t}) \leq 2\exp(-\sqrt{t}^2/c_1^2) \Rightarrow X^2$ sub exponential ($t \geq 0$)

X^2 subexponential $\Rightarrow \mathbb{P}(|X| \geq t) = \mathbb{P}(|X|^2 \geq t^2) \leq 2\exp(-t^2/c_2^2) \Rightarrow X$ sub gaussian

Norms: $\|X^2\|_{\psi_1} = \inf\{c > 0 : \mathbb{E}[\exp(X^2/c)] \leq 2\}$

$\|X\|_{\psi_2} = \inf\{k > 0 : \mathbb{E}[\exp(X^2/k^2) \leq 2\}$ $\longrightarrow$ put $c = k^2 \Rightarrow \|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$

- **Lemma 2.32:** X, Y subgaussian $\Rightarrow$ XY sub exponential and $\|XY\|_{\psi_1} \leq \|X\|_{\psi_2} \|Y\|_{\psi_2}$

Pf: WLOG $\|X\|_{\psi_2} = \|Y\|_{\psi_2} = 1$. WTS $\mathbb{E}[\exp(X^2)] \leq 2$ & $\mathbb{E}[\exp(Y^2)] \leq 2 \Rightarrow \mathbb{E}[\exp(|XY|)] \leq 2 \Rightarrow \|XY\|_{\psi_1} \leq 1$

Tool: Young's inequality! $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$ $\forall a, b \in \mathbb{R}$

⑱ **Def 2.35 (sub-exponential R.V. II):** X w/ mean $\mu \in \mathbb{R}$ is sub-exponential if $\exists (\alpha, \nu) \geq 0$ s.t. $\mathbb{E}[\exp(\lambda(X - \mu))] \leq e^{\nu^2\lambda^2/2}$ $\forall |\lambda| < \frac{1}{\alpha}$

note: This is just centered w/ a lambda bound

Example 2.36 [Bbd R.V.]: $X \in \mathbb{R}$ R.V. mean 0, bbd on $[a, b]$, $X' \perp X$ copy, ε is rademacher, then $\mathbb{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/2}$

Pf: $\mathbb{E}_x[e^{\lambda x}] = \mathbb{E}_x[e^{\lambda(x - \mathbb{E}_{x'}[x'])}] \overset{[\text{Jensen}]}{\leq} \mathbb{E}_{x,x'}[e^{\lambda(x-x')}] = \mathbb{E}_{x,x'}[\mathbb{E}_\varepsilon[e^{\lambda\varepsilon(x-x')}]]$

$\mathbb{E}_\varepsilon[e^{\lambda\varepsilon a}] = \frac{1}{2}e^{\lambda a} + \frac{1}{2}e^{-\lambda a} = \frac{1}{2}\left(\sum_{k=0}^{\infty} \frac{(-\lambda a)^k}{k!} + \sum_{k=0}^{\infty} \frac{(\lambda a)^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{(\lambda a)^{2k}}{(2k)!}$

$\leq 1 + \sum_{k=1}^{\infty} \frac{(\lambda a)^{2k}}{2^k k!} = e^{\lambda^2 a^2/2}$ [def exp]

note: stressing so will just go into down if not 100% sure of its importance

Returning, $\mathbb{E}_x[e^{\lambda x}] \leq \mathbb{E}_{x,x'}[e^{\lambda^2(x-x')^2/2}] \leq e^{\lambda^2(b-a)^2/2}$ $\because |x - x'| \leq b - a$

⑲ **Proposition 2.38 (Sub-exponential tail-bound)** *Let X be a real-valued sub-exponential random variable with parameters (ν, α) and mean $\mu = \mathbb{E}[X]$. Then, for every $t \geq 0$,*

$$\mathbb{P}(X - \mu \geq t) \leq \begin{cases} e^{-\frac{t^2}{2\nu^2}} & \text{for } 0 \leq t < \nu^2/\alpha, \\ e^{-\frac{t}{2\alpha}} & \text{for } t \geq \nu^2/\alpha\,. \end{cases} \tag{2.12}$$

Two diff regions based on

Proof. Recall Definition 2.35 and obtain

$$\mathbb{P}(X - \mu \geq t) \leq e^{-\lambda t}\mathbb{E}[e^{\lambda(X-\mu)}] \leq \exp\left(-\lambda t + \frac{\lambda^2\nu^2}{2}\right) \quad \text{for } \lambda \in \underbrace{[0, \alpha^{-1})}_{\text{from def of sub exp.}}\,.$$

Define $g(\lambda, t) := -\lambda t + \lambda^2\nu^2/2$. We need to determine $g^*(t) = \inf_{\lambda \in [0,\alpha^{-1})}\{g(\lambda, t)\}$. Suppose that t is fixed, then $\partial_\lambda g(\lambda, t) = -t + \lambda\nu^2 = 0$ if and only if $\lambda = \lambda^* = \frac{t}{\nu^2}$. If $0 \leq t < \nu^2/\alpha$, then the infimum equals the unconstrained one and $g^*(t) = -t^2/2\nu^2$ for $t \in [0, \nu^2/\alpha)$. Suppose now that $t \geq \nu^2/\alpha$. As $g(\cdot, t)$ is monotonically decreasing on $[0, \lambda^*)$ (derivative is not positive), the constrained infimum is achieved on the boundary $\lambda^* = \alpha^{-1}$, and hence $g^*(t) = -t/2\alpha$. □

note: just using constraint that $|\lambda| < \frac{1}{|\alpha|}$ & minimizing exp. Markov.

⑳ **Def 2.39 (Bernstein condition):** $X \in \mathbb{R}$. R.V. mean $\mu \in \mathbb{R}$, var $\sigma^2 \in (0, \infty)$ satisfies the Bernstein condition w/ param $b > 0$ if $|\mathbb{E}[(X - \mu)^k]| \leq \frac{1}{2}k!\sigma^2 b^{k-2}$, $k = 2, 3, 4 \ldots$

Exercise 2.40 (a) Show that a bounded random variable X with $|X-\mu| \le b$ with variance $\sigma^2 > 0$ satisfies the *Bernstein condition* (2.13).

(b) Show that the bounded random variable X in (a) is sub-exponential and derive a bound on the centred moment generating function

$$\mathbb{E}[\exp(\lambda(X - \mathbb{E}[X]))] .$$

Clever use!

☕☕

Solution. (a) From our assumption we have $\mathbb{E}[(X-\mu)^2] = \mathbb{E}[|X-\mu|^2] = \sigma^2$ and ess sup$|X-\mu|^{k-2} \le b^{k-2} \le b^{k-2}\frac{1}{2}k!$ for $k \in \mathbb{N}, k \ge 2$. Using Hölder's inequality we obtain

$\mathbb{E}[XY] \le \mathbb{E}[|XY|] \le \|X\|_{L^p}\|X\|_{L^q}$ $\frac{1}{p}+\frac{1}{q}=1$

$$\mathbb{E}[|X-\mu|^{k-2}|X-\mu|^2] \le \mathbb{E}[|X-\mu|^2]\,\text{ess sup}|X-\mu|^{k-2} \le \sigma^2 b^{k-2}\frac{1}{2}k!$$

for all $k \in \mathbb{N}, k \ge 2$.

(b) By power series expansion we have (using the Bernstein bound from (a),

$$\mathbb{E}[\mathrm{e}^{\lambda(X-\mu)}] = 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty}\lambda^k\frac{\mathbb{E}[(X-\mu)^k]}{k!}$$
$$\le 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2}\sum_{k=3}^{\infty}(|\lambda|b)^{k-2},$$

and for $|\lambda| < 1/b$ we can sum the geometric series to obtain

$$\mathbb{E}[\mathrm{e}^{\lambda(X-\mu)}] \le 1 + \frac{\lambda^2\sigma^2/2}{1-b|\lambda|} \le \exp\left(\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}\right)$$

by using $1+t \le \mathrm{e}^t$. Thus X is sub-exponential as we obtain

$$\mathbb{E}[\mathrm{e}^{\lambda(X-\mu)}] \le \exp\left(\lambda^2(\sqrt{2}\sigma)^2/2\right)$$

for all $|\lambda| < 1/2b$. □

(General Hoeffding)

(21) Exercise 2.41: $X_1, \dots, X_N$ indpt. mean zero s.g. R.V. $a \in \mathbb{R}^n$, then $\forall t \ge 0$

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} a_i X_i\right| \ge t\right) \le 2\exp\left(-\frac{ct^2}{K^2\|a\|_2^2}\right) \qquad K = \max\{\|X_i\|_{\psi_2}\}$$

Chapter 3 - Random vectors in High Dimensions

(22) Ex 3.2 (Centering for sub. exp. R.V.s): $X \in \mathbb{R}$ s.e. then $\|X - \mathbb{E}[X]\|_{\psi_1} \le C\|X\|_{\psi_1}$

(23) Ex. 3.3 (Bernstein Inequality). $X_1, \dots, X_N$ independent mean zero s.e. $\mathbb{R}$- R.V. Then $\forall t \ge 0$

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} X_i\right| \ge t\right) \le 2\exp\left(-c\min\left\{\frac{t^2}{K^2}, \frac{t}{K}\right\}N\right)$$

sub exponential ⇒ 2 diff regimes, not so bad...

so some abs. cnst. $c > 0$, $K = \max_{1 \le i \le N}\{\|X_i\|_{\psi_1}\}$

Pf: same as (19) but trickier g to minimize

(24) Thm 3.1 (Concentration of the Norm): $X \in \mathbb{R}^n$ random vector, n indi s.g. coords. X_i with $\mathbb{E}[X_i^2] = 1$, then $\big\| \|X\|_2 - \sqrt{n} \big\|_{\psi_2} \le cK^2$ where $K = \max_{1 \le i \le n}\{\|X_i\|_{\psi_2}\}$, $c > 0$ abs. cnst.

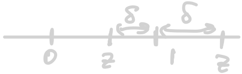
Pf: WLOG, assume $K \geq 1$. $\frac{1}{n}\|X\|_2^2 - 1 = \overbrace{\frac{1}{n}\sum_{i=1}^{n}\underbrace{(X_i^2 - 1)}_{\text{recentered sub exp.}}}^{(*)}$.

Note: X_i sub gaussian $\Rightarrow X_i^2 - 1$ sub exponential ⑰

Centering ㉒ $\Rightarrow \|X_i^2 - 1\|_{\psi_1} \leq C\|X_i^2\|_{\psi_1} \overset{⑰}{=} C\|X_i\|_{\psi_2}^2 \leq CK^2$ [∵ max]

Bernstein ㉓ $\Rightarrow \mathbb{P}\Big(\overbrace{\big|\frac{1}{n}\|X\|_2^2 - 1\big|}^{(*)} \geq u\Big) \leq 2\exp\left(-cn\min\left\{\frac{u^2}{C^2K^4}, \frac{u}{CK^2}\right\}\right) \leq 2\exp\left(-\frac{cn}{K^4}\min\{u^2, u\}\right)$ using $K^4 \geq K^2$ as $K \geq 1$

$\boxed{\text{Fact I}}$ For $z \geq 0$, if $|z-1| \geq \delta \Rightarrow |z^2 - 1| \geq \max\{\delta, \delta^2\}$



Pf: If $z > 1 \Rightarrow z+1 \geq z-1 \Rightarrow |z^2-1| = |z+1||z-1| \geq \delta^2$

If $0 \leq z \leq 1 \Rightarrow z+1 \geq 1 \Rightarrow |z^2-1| = |z+1||z-1| \geq \delta$

So set $u = \max\{\delta, \delta^2\}$ and $z = \frac{1}{\sqrt{n}}\|X\|_2$ in $\boxed{\text{Fact I}}$ to see that

$$\mathbb{P}\Big(\overbrace{\big|\tfrac{1}{\sqrt{n}}\|X\|_2 - 1\big| \geq \delta}^{\text{horrible norm, hard to ctrl}}\Big) \leq \mathbb{P}\Big(\overbrace{\underbrace{\big|\tfrac{1}{n}\|X\|_2^2 - 1\big|}_{\text{nice norm townee}}}^{\text{much easier to ctrl}} \geq \max\{\delta, \delta^2\}\Big) \leq 2\exp\left(-\frac{cn}{K^4}\delta^2\right) \boxed{\text{Fact II}}$$

$\boxed{\text{Fact II}}$ $v = \min\{u, u^2\} = \delta^2$ when $u = \max\{\delta, \delta^2\}$

Pf: $\delta \geq \delta^2 \Rightarrow \delta^2 \leq 1 \Rightarrow u = \delta \Rightarrow v = \delta^2$

$\delta > 1 \Rightarrow \delta^2 \geq \delta \Rightarrow u = \delta^2 \Rightarrow v = \delta^2$

Set $t = \delta\sqrt{n} \Rightarrow \mathbb{P}\big(\big|\|X\|_2 - \sqrt{n}\big| \geq t\big) \leq 2\exp\left(-\frac{ct^2}{K^4}\right) \Rightarrow \big|\|X\|_2 - \sqrt{n}\big|$ s.g. ■

note: Ex Sheet 2 extended this to show that

(a) $-\sqrt{n} - CK^2 \leq \mathbb{E}[\|X\|_2] \leq \sqrt{n} + CK^2$

(b) $\mathrm{var}(\|X\|_2) \leq \tilde{C}K^4$

Proof$^{(A)}$: set $z = \|X\|_2 - \sqrt{n}$

$\|z\|_{\psi_2} \leq CK^2$. WTS $|\mathbb{E}[z]| \leq CK^2$

$|\mathbb{E}[z]| \leq \mathbb{E}[|z|] = \|z\|_{L^1}$

$\leq \tilde{C}\|z\|_{\psi_2}\sqrt{1}$

$\leq CK^2$

Proof$_{(B)}$: $\mathrm{var}(\|X\|_2) = \mathbb{E}[(\|X\|_2 - \mathbb{E}[\|X\|_2])^2]$

$\leq \mathbb{E}[(\|X\|_2 - \sqrt{n})^2]$

$= \big\|\|X\|_2 - \sqrt{n}\big\|_{L^2}^2 \leq \tilde{C}\big\|\|X\|_2 - \sqrt{n}\big\|_{\psi_2}^2\sqrt{2}^2$

㉕ Ex 3.5 (Small Ball Probabilities): $X \in \mathbb{R}^n$ w/ X_i having pdf $f_i : \mathbb{R} \to \mathbb{R}$ s.t. $|f_i(x)| \leq 1$ $\forall i$ $\forall x \in \mathbb{R}$. WTS $\forall \varepsilon > 0$ $\mathbb{P}(\|X\|_2 \leq \varepsilon\sqrt{n}) \leq (C\varepsilon)^n$ for some abs. cts. $C > 0$

Pf: $\mathbb{P}(\|X\|_2 \leq \varepsilon\sqrt{n}) = \mathbb{P}(\|X\|_2^2 \leq \varepsilon^2 n) = \mathbb{P}(-\|X\|_2^2 \geq -\varepsilon^2 n) \leq e^{\lambda\varepsilon^2 n}\mathbb{E}[\exp(-\lambda\|X\|_2^2)]$

$\mathbb{E}[\exp(-\lambda X_i^2)] \leq \int_{\mathbb{R}} e^{-\lambda x_i^2}|f_i(x)|dx \leq \int_{\mathbb{R}} e^{-\lambda x_i^2}dx = \sqrt{\frac{\pi}{\lambda}}$ $\quad = e^{\lambda\varepsilon^2 n}\prod_{i=1}^{n}\mathbb{E}[\exp(-\lambda X_i^2)]$

Plug in & optimise over λ. ■

㉖ Facts about geometry in high dimensions

ⓐ $B_R^{(n)} = \{x \in \mathbb{R}^n : \|x\|_2 \leq R\}$, $S_R^{(n-1)} = \{x \in \mathbb{R}^n : \|x\|_2 = R\}$

ⓑ $\mathrm{vol}(B_R^{(n)}) = \frac{\pi^{n/2}}{\frac{n}{2}\Gamma(\frac{n}{2})}R^n$, $\mathrm{Area}(S_R^{(n-1)}) = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})}R^{n-1}$

e.g. $n=3$, $R=1$

$\Gamma(\frac{3}{2}) = \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}$

$\mathrm{Area}(S^2) = 4$

$\mathrm{vol}(B^{(3)}) = \frac{4\pi}{3}$

ⓒ n-dim polar: $\mathrm{vol}(B^{(n)}) = \int_{S^{n-1}}\int_0^1 r^{n-1}dr\,\overbrace{d\sigma}^{\text{S.A. measure}} = \frac{1}{n}\int_{S^{n-1}}d\sigma = \frac{\mathrm{Area}(S^{n-1})}{n}$

$t = r^2 \quad \frac{dt}{dr} = 2r$

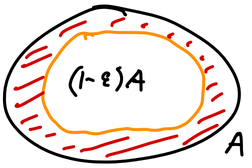
$dr = \frac{dt}{2\sqrt{t}}$

ⓓ Relation: $I(n) := \int_{\mathbb{R}}\cdots\int_{\mathbb{R}} e^{-(x_1^2 + \cdots + x_n^2)}dx_n \cdots dx_1 = (\sqrt{\pi})^n = \pi^{n/2}$

alternatively $\quad t = r^2 \quad \frac{dt}{dr} = 2r$

$I(n) = \int_{S^{n-1}}d\sigma\int_0^\infty e^{-r^2}r^{n-1}dr = \mathrm{Area}(S^{n-1})\int_0^\infty e^{-t}t^{\frac{n-1}{2}}\frac{dt}{2t^{\frac{1}{2}}} = \frac{1}{2}\mathrm{Area}(S^{n-1})\int_0^\infty e^{-t}t^{\frac{n}{2}-1}dt = \mathrm{Area}(S^{n-1})\frac{1}{2}\Gamma(\frac{n}{2})$

Hence, $\mathrm{Area}(S^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})}$



volume overwhelming lies here !!!

㉗ Skin of a high Dim orange!

$(1-\varepsilon)A := \{(1-\varepsilon)x : x \in A\}$, $\mathrm{vol}((1-\varepsilon)A) = (1-\varepsilon)^n \mathrm{vol}(A) \Rightarrow \frac{\mathrm{vol}((1-\varepsilon)A)}{\mathrm{vol}(A)} = (1-\varepsilon)^n \leq e^{-n\varepsilon}$

Random vectors on the unit ball are orthogonal w/ high probability!

㉘ **Prop 3.7:** Sample N pts $X^{(1)}, \dots, X^{(N)}$ unif. on unit ball B^n_n Then w/ prob $1 - O(\frac{1}{N})$

(a) $\|X^{(i)}\|_2 \geq 1 - \frac{2\log N}{n}$, $\forall i \in \{1, \dots, n\}$

(b) $\langle X^{(i)}, X^{(j)} \rangle \leq \frac{6 \log N}{\sqrt{n-1}}$ $\forall i \neq j$

$\mathbb{P}(\bigcup A_i) \leq \sum \mathbb{P}(A_i)$

(a)

Pf: $\mathbb{P}(\|X^{(i)}\|_2 < 1-\varepsilon) \leq e^{-n\varepsilon}$ from ㉗, pick $\varepsilon = \frac{2\log N}{n}$, then

$\leq \exp(-2\log N) = \frac{1}{N^2}$

Union bd: $\mathbb{P}(\exists i \in \{1, \dots, N\} : \|X^{(i)}\|_2 \leq 1 - \frac{2\log N}{n}) \leq N \cdot \frac{1}{N^2} = \frac{1}{N}$

(b) spt does...

㉙ **Covariance Matrices:** $X \in \mathbb{R}^n$ random vector. $XX^T = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} (X_1, \dots, X_n) = \begin{pmatrix} X_1^2 & X_1X_2 & \cdots & X_1X_n \\ X_2X_1 & & & \vdots \\ \vdots & & \ddots & \\ X_nX_1 & & \cdots & X_n^2 \end{pmatrix}$

$\mathbb{E}[X] = \mu \in \mathbb{R}^n$, then

- covariance-matrix: $\mathrm{cov}(X) = \mathbb{E}[(X-\mu)(X-\mu)^T] = \mathbb{E}[XX^T] - \mu\mu^T$
- 2nd moment matrix: $\Sigma = \Sigma(X) = \mathbb{E}[XX^T] = (\mathbb{E}[X_iX_j])_{ij}$

PCA: write $\Sigma(X) = \sum_{i=1}^n s_i u_i u_i^T$ where $u_i \in \mathbb{R}^n$ eigenvectors for s_i eigenvalues. order by size $s_1 \geq s_2 \geq \dots \geq s_n$, then forget about smaller eigenvalues

↳ dimension of data reduced!

identity operator or matrix on $\mathbb{R}^n$

㉚ **Def 3.11 (Isotropic random vector):** $X \in \mathbb{R}^n$ isotropic if $\Sigma(X) = \mathbb{E}[XX^T] = \mathbb{1}_n$

↳ unit variance, components x depend on each other

㉛ **Lemma 3.13 (Isotropy):** $X \in \mathbb{R}^n$ isotropic $\iff$ $\mathbb{E}[\langle X, x \rangle^2] = \|x\|_2^2$ $\forall x \in \mathbb{R}^n$

IFF $\Sigma = \mathbb{1}_n$

Pf: $\mathbb{E}[\langle X, x\rangle^2] = \sum_{i,j=1}^n \mathbb{E}[x_i X_i x_j X_j] = \sum_{i,j=1}^n x_i \mathbb{E}[X_iX_j] x_j = \langle x, \Sigma x \rangle \overset{\downarrow}{=} \|x\|_2^2$

note: Enough to show $\mathbb{E}[\langle X, e_i\rangle^2] = 1$ $\forall$ basis vectors $e_1, \dots, e_n$

↳ 1d marginals all have unit variance $\Rightarrow$ isotropic dist evenly extends in all directions.

㉜ **Lemma 3.14:** $X \in \mathbb{R}^n$ is isotropic. Then $\mathbb{E}[\|X\|_2^2] = n$, if X, Y also indi $\Rightarrow \mathbb{E}[\langle X, Y\rangle] = n$

Pf: (a) $\mathbb{E}[\|X\|_2^2] = \mathbb{E}[\overset{1\times 1}{X^TX}] = \mathbb{E}[\mathrm{trace}(X^TX)] \overset{\text{[cyclic property of trace]}}{=} \mathbb{E}[\mathrm{trace}(XX^T)] = \mathrm{trace}(\overbrace{\mathbb{E}[XX^T]}^{\mathbb{1}_n}) = n$

(b) Fix realisation of Y, $\mathbb{E}[\langle X, Y\rangle^2] = \mathbb{E}_Y[\overbrace{\mathbb{E}_X[\langle X, Y\rangle^2 \mid Y]}^{\mathbb{E}_X[\langle Y, Y\rangle \mid Y] = \|Y\|_2^2}]$ [Law of total expectation]

Put $x = Y$ $= \mathbb{E}_Y[\|Y\|_2^2] = n$ [part (a)]

note: X, Y isotropic, indi $\Rightarrow \|X\|_2 \sim \sqrt{n}$, $\|Y\|_2 \sim \sqrt{n}$, $(X, Y) \sim \sqrt{n}$ $\Rightarrow (\frac{X}{\|X\|_2}, \frac{Y}{\|Y\|_2}) \sim \frac{1}{\sqrt{n}}$

so in high dim, X, Y almost always orthogonal.

㉝ **Def 3.15 (Spherical Dist.):** X R. vector is spherically distributed if $X \sim \mathrm{Unif}(\sqrt{n}\, S^{n-1})$

note: $X \sim \mathrm{Unif}(\sqrt{n}\, S^{n-1})$ is isotropic, but $X_1^2 + \dots + X_n^2 = n \Rightarrow$ not independent.

uniform so puts [illegible]

$\cos^2 x - \sin^2 x = 1$

Pf: $n = 2$ $X = \sqrt{n}\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ $\mathbb{E}[\langle X, e_1\rangle^2] = \mathbb{E}[n\cos^2\theta] = \frac{n}{2\pi}\int_0^{2\pi} \cos^2\theta = 1$

Fact: Any random $X \in \mathbb{R}^n$ w/ indi, mean zero coords X_i w/ unit var is isotropic in $\mathbb{R}^n$

$Y \in \mathbb{R}^n$ is $Y \sim N(0, \mathbb{1}_n)$ if $Y_i \sim N(0,1)$

$f_Y(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2}$

e.g. Symmetrize bernoulli $X \sim \mathrm{Unif}(\{-1, 1\}^n)$

$\mathbb{E}[\langle X, x\rangle^2] = \mathbb{E}[\sum_{i=1}^n X_i^2 x_i^2] + \cancel{\mathbb{E}[2 \sum_{1 \leq i < j \leq n} X_i x_i X_j x_j]}^{\text{indi mean zero}} = \sum_{i=1}^n x_i^2 = \|x\|_2^2$

recall: If $Y \sim \mathcal{N}(\mu, C)$ $\mu \in \mathbb{R}^n$, variance-covariance matrix $C \in \mathbb{R}^{n \times n}$, then

$$\varphi_x(t) = \mathbb{E}[e^{it^TY}] = \mathbb{E}[e^{i<t,\mu> - \frac{1}{2}<t,Ct>}]$$

MGF doesn't always converge, CF solves this problem

(34) **Prop 3.18 (Standard Normal Dist. is rotation Invariant):** $Y \sim \mathcal{N}(0, \mathbb{I}_n)$, $\underbrace{U \in O(n)}_{\text{orthogonal}}$ $UU^T = U^TU = \mathbb{I}_n$

|| Then $UY \sim \mathcal{N}(0, \mathbb{I}_n)$ (a rotation doesn't change the distribution)

Pf: $\overset{Z \in \mathbb{R}^{n\times 1}}{Z} = UY$, $\|Z\|_2^2 = Z^TZ = Y^TU^TUY = Y^TY = \|Y\|_2^2$ and $|\det(U)| = |\det(U^T)| = 1$

so $\forall J \in \mathbb{C}^n$, write $Z = Uy$ $\forall y \in \mathbb{R}^n$ & characteristic function

$$\underbrace{\mathbb{E}_Z[e^{<J,Z>}]}_{\varphi_Z(t) = \mathbb{E}[e^{it^TZ}]} = \int_{\mathbb{R}^n} e^{<J,Z>} \cdot \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\|Z\|_2^2} \prod_{i=1}^n dz_i$$

$\mathbb{E}[g(X)] = \int g(x) f_X(x) dx$

$$= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}<z,z> + <J,z>\right) \prod_{i=1}^n dz_i$$

$z = Ux \Rightarrow <J, Ux> = <U^TJ, x>$

$\Rightarrow <Ux, Ux> = x^TU^TUx = x^Tx = <x,x>$

$$= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}<y,y> + <U^TJ, y>\right) |\det U| \prod_{i=1}^n dy_i$$

$$= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{i=1}^n e^{-\frac{1}{2}y_i^2 + (U^TJ)_i y_i} \prod_{i=1}^n dy_i$$

complete □ & Fubini ∵ independence

$$= \frac{1}{(2\pi)^{n/2}} \prod_{i=1}^n \int_{\mathbb{R}} \exp\left(-\frac{1}{2}(y_i - (U^TJ)_i)^2 + \frac{1}{2}(U^TJ)_i^2\right) \prod_{i=1}^n dy_i$$

$$= \frac{1}{(2\pi)^{n/2}} \prod_{i=1}^n e^{\frac{1}{2}(U^TJ)_i^2} \overbrace{\int_{\mathbb{R}} e^{-\frac{1}{2}(y_i - (U^TJ)_i)^2} \prod_{i=1}^n dy_i}^{\sqrt{2\pi}}$$

$$= \exp\left(\frac{1}{2}<U^TJ, U^TJ>\right)$$

$$= e^{\frac{1}{2}<J,J>}$$ [MGF of standard normal]

$= \mathbb{E}[e^{<J,Y>}] \Longrightarrow Z$ & Y have same CF $\Rightarrow Z \sim UY$ ■

(35) **Def (General Normal Distribution):** $\Sigma \in \mathbb{R}^{n\times n}$ symmetric, positive definite.

$X \in \mathbb{R}^n$ w/ $\mu = \mathbb{E}[X]$, then

$X = \mu + \Sigma^{\frac{1}{2}}Z$ ∵ $X = \mu + \sigma Z$

$$X \sim \mathcal{N}(\mu, \Sigma) \Longleftrightarrow Z = \Sigma^{-\frac{1}{2}}(X - \mu) \sim N(0, \mathbb{I}) \Longleftrightarrow f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\langle x-\mu, \Sigma^{-1}(x-\mu)\rangle\right)$$

(36) **Def 3.19 (Subgaussian random vectors):** $X \in \mathbb{R}^n$ is sub-gaussian

|| if the 1d marginals $<X,x>$ are all s.g. $\forall x \in \mathbb{R}^n$, moreover...

$$\|X\|_{\psi_2} = \sup_{x \in S^{n-1}} \{\|<X,x>\|_{\psi_2}\}$$

$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 1d parallel

(37) **Prop 3.20:** If $X \in \mathbb{R}^n$ has independent mean zero sub-gauss. coords $X_i \Rightarrow X$ sub g.

| and $\|X\|_{\psi_2} \leq C \max_{1\leq i\leq n}\{\|X\|_{\psi_2}\}$ for some abs. const. $C > 0$

Pf: Pick a direction $x \in S^{n-1}$ & compute

$\sum x_i^2 = 1$

$$\|<X,x>\|_{\psi_2}^2 = \|\sum_{i=1}^n X_i x_i\|_{\psi_2}^2 \underset{(12)}{\leq} C\sum_{i=1}^n x_i^2 \|X_i\|_{\psi_2}^2 \leq C \max\{\|X_i\|_{\psi_2}^2\}$$

■

(38) **Thm 3.21 (Uniform Distribution on the Sphere):** $X \in \mathbb{R}^n$ w/ $X \sim \text{Unif}(\sqrt{n}\, S^{n-1})$,
|| Then X is subgaussian and $\|X\|_{\psi_2} \le C$ for some a.c. $C > 0$.

Pf: For ease, work w/ unit sphere so $Z = \frac{X}{\sqrt{n}} \sim \text{Unif}(S^{n-1})$

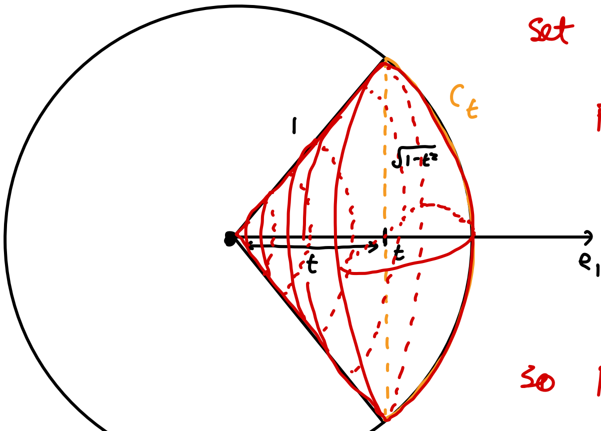
Uniformly chosen ⇓ just pick one direction

WTS $\|Z\|_{\psi_2} \le \frac{C}{\sqrt{n}} \iff \|\langle Z, x\rangle\|_{\psi_2} \le C/\sqrt{n}$ $\forall$ unit vectors $x \in S^{n-1}$

Show $\mathbb{P}(|Z_1| \ge t) \le 2\exp(-ct^2 n)$ WTS this ∵ can pick $x = e_1$

Upper tail: $\frac{\text{length}(C_t)}{\text{length}(S^{n-1})} = \frac{\text{vol}(K_t)}{\text{vol}(B^n)}$ $\quad C_t = \{z \in S^{n-1} : z_1 \ge t\}$ [spherical cap]

$$\mathbb{P}(Z_1 > t) = \mathbb{P}(Z \in C_t) = \frac{\text{vol}(K_t)}{\text{vol}(B^{(n)})}$$



only for $\sqrt{1-t^2} \ge t \iff 1-t^2 \ge t^2 \iff \frac{1}{\sqrt{2}} \ge t \ge 0$

$$\text{Set } O' = (t, 0, \dots, 0) \Rightarrow \underbrace{K_t \subset B(O', \sqrt{1-t^2})}$$

length scale factor

$$\mathbb{P}(Z_1 \ge t) \le \frac{\text{vol}(B^{(n)}(O', \sqrt{1-t^2}))}{\text{vol}(B^{(n)}(0,1))} = \left(\sqrt{1-t^2}\right)^n \le e^{-\frac{t^2 n}{2}}$$

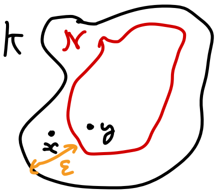
[valid for $0 \le t \le \frac{1}{\sqrt{2}}$], for $\frac{1}{\sqrt{2}} \le t \le 1$,

$$\mathbb{P}(Z_1 \ge t) \le \mathbb{P}(Z_1 \ge \tfrac{1}{\sqrt{2}}) \le e^{-n/4} \le e^{-\frac{t^2 n}{4}}$$

so $\mathbb{P}(Z_1 \ge t) \le e^{-\frac{t^2 n}{4}}$. Symmetry ⇒ same for $-Z_1$

$$\therefore \mathbb{P}(|Z_1| \ge t) = \mathbb{P}(\{Z_1 \ge t\} \cup \{-Z_1 \le -t\}) \le 2e^{-\frac{t^2 n}{4}}$$ ■

Chapter 4 - Random Matrices



(39) **Def 4.1:** Let (T, d) be a metric space, $K \subset T$, $\varepsilon > 0$

(a) ε-net: a subset $N \subset K$ is an ε-net of K if every pt in K is within a distance ε of some pt of N. $\forall x \in K$ $\exists y \in N$ s.t. $d(x,y) \le \varepsilon$

note: 3 other concepts but will skip... not the point of the module...

(40) **Def 4.2:** $A, B \subset \mathbb{R}^n$, $\quad A + B = \{a + b : a \in A, b \in B\}$ [set addition]

(41) **Prop 4.5 (Covering Numbers of the Euclidean Ball):** Let $K \subset \mathbb{R}^n$, $\varepsilon > 0$, $|K| = \text{vol}(K)$,
$B^{(n)} = \{x \in \mathbb{R}^n : \|x\|_2 \le 1\}$ (closed). Then

smallest possible cardinality of an ε-net of K · largest possible cardinality of an ε-separated subset of K

(a)
$$\frac{|K|}{|\varepsilon B^n|} \le \underbrace{N(K, \|\cdot\|_2, \varepsilon)}_{\text{covering number}} \le \underbrace{P(K, \|\cdot\|_2, \varepsilon)}_{\text{packing number}} \le \frac{|(K + \frac{\varepsilon}{2} B^{(n)})|}{|\frac{\varepsilon}{2} B^{(n)}|}$$

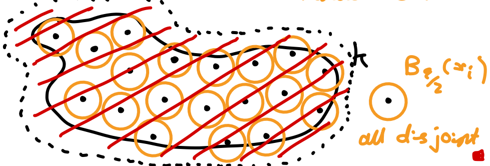
(b)
$$\left(\frac{1}{\varepsilon}\right)^n \le N(B^{(n)}, \|\cdot\|_2, \varepsilon) \le \left(\frac{2}{\varepsilon} + 1\right)^n$$

Pf: Let $N = N(K, \|\cdot\|_2, \varepsilon)$, cover K w/ N balls of radius ε $\Rightarrow |K| \le N \overbrace{|\varepsilon B^{(n)}|}^{B_\varepsilon^{(n)}(0) \text{ vol of ball radius } \varepsilon}$ [lower bd]

For upper bd, let $N = P(B^{(n)}, \|\cdot\|_2, \varepsilon)$, construct N closed disjoint balls $B_{\frac{\varepsilon}{2}}(x_i)$ w/ centers $x_i \in K$, balls will certainly fit into $\underbrace{K + \frac{\varepsilon}{2} B^{(n)}}_{\text{extended set}}$
So we see $N |\frac{\varepsilon}{2} B^{(n)}| \le |K + \frac{\varepsilon}{2} B^{(n)}|$

(b) apply w/ K = unit ball itself

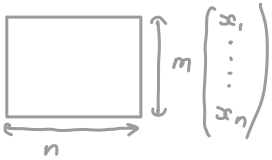
note: middle inequality from skipped...



(42) **Def 4.8 (Hamming cube):** $\mathcal{H} = \{0,1\}^n$, $d_{\mathcal{H}}(x,y) = \#\{i \in [n] : x(i) \neq y(i)\}$, $x, y \in \{0,1\}^n$

note: ex. to show metric?

(43) **Def 4.10:** $A \in \mathbb{R}^{m \times n}$ represents a linear map from $\mathbb{R}^n \longrightarrow \mathbb{R}^m$



(a) Operator norm of A

$$\|A\| = \max_{x \in \mathbb{R}^n \setminus \{0\}} \left\{ \frac{\|Ax\|_2}{\|x\|_2} \right\} = \max_{x \in S^{n-1}} \{\|Ax\|_2\} = \max_{x \in S^{n-1}, y \in S^{m-1}} \{\langle Ax, y \rangle\}$$

← eigenvalues

(b) Singular values $s_i = s_i(A)$ are the $\sqrt{\lambda_i}$ of AA^T and A^TA, $s_i(A) = \sqrt{\lambda_i(AA^T)}$
order them $s_1 \geq s_2 \geq \dots \geq s_n \geq 0$, if A symmetric $s_i = |\lambda_i|$

(c) If $r = \text{rank}(A)$, SVD decomp: $A = \sum_{i=1}^{r} s_i u_i v_i^T$

s_i = singular values
u_i = left singular vectors $\in \mathbb{R}^m$
v_i = right sing. vectors $\in \mathbb{R}^n$

note: $\|A\| = s_1(A)$ [largest singular value]

(44) **Lemma 4.12 (Operator Norm on a net):** Let $\varepsilon \in [0,1)$, $A \in \mathbb{R}^{m \times n}$, then for any ε-net $\mathcal{N}$ of S^{n-1}, we have

$$\sup_{x \in \mathcal{N}} \{\|Ax\|_2\} \leq \|A\| \leq \frac{1}{1-\varepsilon} \sup_{x \in \mathcal{N}} \{\|Ax\|_2\}$$

ε-net is a subset...

exercise below too...

Pf: Lower bd: $\underbrace{\mathcal{N} \subset S^{n-1}} \Rightarrow \sup_{x \in \mathcal{N}} \{\|Ax\|_2\} \leq \sup_{x \in S^{n-1}} \{\|Ax\|_2\} = \|A\|$ [trivial]

Upper bd: Pick $x \in S^{n-1}$ s.t. $\|A\| = \|Ax\|_2$ [Note cts image compact set compact so will attain supremum (S^{n-1} compact)]
choose $x_0 \in \mathcal{N}$ s.t. $\|x - x_0\|_2 \leq \varepsilon$
Then $\|Ax - Ax_0\|_2 \leq \|A\| \|x - x_0\|_2 \leq \varepsilon \|A\|$
$\|Ax_0\|_2 = \|Ax - (Ax - Ax_0)\|_2 \geq \|Ax\|_2 - \|Ax - Ax_0\|_2 \geq (1-\varepsilon)\|A\|$ $\Rightarrow \|A\| \leq \frac{\|Ax_0\|_2}{1-\varepsilon}$

(45) **Thm 5.15 (Norm of S.G. Random matrices):** $A \in \mathbb{R}^{m \times n}$ random matrix w/ independent mean zero subgaussian random entries A_{ij}, $i \in [m]$, $j \in [n]$, then $\forall t > 0$
$\|A\| \leq CK(\sqrt{m} + \sqrt{n} + t)$ w/ prob. at least $1 - 2e^{-t^2}$ where $K = \max_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \{\|A_{ij}\|_{\psi_2}\}$

{ε-net argument}

$x \in \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$ $y \in \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

Pf: $\boxed{\text{STEP 1}}$: For $\varepsilon = \frac{1}{4}$, can find an ε-net $\mathcal{N} \subset S^{n-1}$ and ε-net $\mathcal{M} \subset S^{m-1}$ w/
$|\mathcal{N}| \leq (\frac{2}{\varepsilon} + 1)^n = 9^n$, $|\mathcal{M}| \leq (\frac{2}{1/4} + 1)^m = 9^m$ [Prop (41)] cardinality of ε-net bounded

S^{n-1} uncountable
ε-net ⇒ finite # of points
[concentration]

Ex that improves (44) to $\|A\| \leq \frac{1}{\underbrace{1 - 2\varepsilon}_{=2}} \sup_{x \in \mathcal{N}, y \in \mathcal{M}} \{\langle Ax, y \rangle\}$

$\frac{1}{1 - 2 \cdot \frac{1}{4}} = \frac{1}{1 - \frac{1}{2}} = 2$

$\boxed{\text{STEP 2}}$: Pick $x \in \mathcal{N}$, $y \in \mathcal{M}$

Δ inequality

Single entries S.G. w/ norm bd by K

just take max

$= \sum_{i=1}^m x_i^2 \sum_{j=1}^n y_j^2$

$$\|\langle Ax, y \rangle\|_{\psi_2}^2 = \left\| \sum_{i=1}^m \sum_{j=1}^n A_{ij} x_i y_j \right\|_{\psi_2}^2 \leq C \sum_{i=1}^m \sum_{j=1}^n \|A_{ij} x_i y_j\|_{\psi_2}^2 \leq CK^2 \underbrace{\sum_{i=1}^m \sum_{j=1}^n y_j^2 x_i^2}_{=1} = CK^2$$

↖ abs cnst. $C > 0$

X sub-gaussian means
$\mathbb{P}(|X| \geq u) \leq 2e^{-\frac{t^2}{\|X\|_{\psi_2}^2}}$

concentration bd for specific choice

So, $\forall u \geq 0$ $\mathbb{P}(\langle Ax, y \rangle \geq u) \leq 2\exp\left(-\frac{cu^2}{K^2}\right)$ $c > 0$ some abs. const.

$\boxed{\text{STEP 3}}$: Unfix choice of x & y and union bd $\mathbb{P}(\bigcup A_i) \leq \sum \mathbb{P}(A_i)$
(union bd)

ε-net so able to do finite sum!

$$\mathbb{P}\left(\max_{x \in \mathcal{N}, y \in \mathcal{M}} \{\langle Ax, y \rangle\} \geq u\right) \leq \sum_{x \in \mathcal{N}, y \in \mathcal{M}} \mathbb{P}(\langle Ax, y \rangle \geq u) \leq \underbrace{9^{n+m}}_{\#\text{elements}} \cdot 2e^{-\frac{cu^2}{K^2}}$$

Set $u = CK(\sqrt{n} + \sqrt{m} + t)$
$\Rightarrow u^2 \geq C^2K^2(n + m + t^2)$, adjust $C > 0$ s.t. $\frac{cu^2}{K^2} \geq 3(n+m) + t^2$

Then $\mathbb{P}\left(\max_{x \in \mathcal{N}, y \in \mathcal{M}} \{\langle Ax, y \rangle \geq u\}\right) \leq \underbrace{9^{n+m} \cdot 2\exp(-3(n+m) + t^2)}_{\left(\frac{9}{e^3}\right)^{n+m} \leq 1} \leq 2e^{-t^2}$

So $\mathbb{P}(\|A\| \geq u) \leq 2e^{-t^2}$ ∎

---

independence made calc. rel. easy, now relax...

Chapter 5 - Concentration of Measure - General Case

Random variables NOT necessarily independent...

Part 1: Entropy Methods

㊻ **Def 5.1 (Entropy):** X $\mathbb{R}$-valued, $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ convex, then $H_\varphi(X) = \mathbb{E}[\varphi(X)] - \varphi(\mathbb{E}[X])$ is the entropy of X for φ.

note: Jensen: $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)] \Rightarrow H_\varphi(X) \geq 0$

(a) $\varphi(x) = x^2$, $H_\varphi(X) = Var(X)$

(b) $\varphi(x) = -\log u$, $Z = e^{\lambda X}$, $H_\varphi(Z) = \log \mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}]$

(c) $\varphi(u) = u \log u$, $u > 0$ $\quad \varphi(0) := 0$ $\quad H(Z) = H_\varphi(Z) = \mathbb{E}[Z \log Z] - \mathbb{E}[Z] \log \mathbb{E}[Z]$

㊼ **Def 5.2 (relative / shannon):** Ω finite sample space, $\mathcal{M}_1(\Omega)$ set of prob. measures

(a) The relative entropy wrt $q \in \mathcal{M}_1(\Omega)$ is

$$H(\cdot|q) : \mathcal{M}_1(\Omega) \longrightarrow [0, \infty];\ p \longmapsto H(p|q) = \begin{cases} \sum_{w \in \Omega} p(w) \log \frac{p(w)}{q(w)} & \text{if } q(w) \Rightarrow p(w) = 0 \\ +\infty & \text{o/w} \end{cases}$$

(b) The shannon entropy of $X \in \mathbb{R}$ w/ pdf $p \in \mathcal{M}_1(\Omega)$ is

$$\mathcal{H}(X) = \mathcal{H}(p) = -\sum_{x \in \Omega} p(w) \log p(w)$$

㊽ **Entropy & MGF:** $X \in \mathbb{R}$ R.V. $Z := e^{\lambda X}$, $\lambda > 0$, then $H(e^{\lambda X}) = \lambda M'_X(\lambda) - M_X(\lambda) \log M_X(\lambda)$

$\varphi(u) = u \log u$

Pf: $H(e^{\lambda X}) = H_\varphi(e^{\lambda X}) = \mathbb{E}[\varphi(e^{\lambda X})] - \varphi(\mathbb{E}[e^{\lambda X}])$ $\qquad \varphi(u) = u \log u$

$$= \mathbb{E}[\lambda X e^{\lambda X}] - \mathbb{E}[e^{\lambda X}] \log \mathbb{E}[e^{\lambda X}]$$

$$= \lambda M'_X(\lambda) - M_X(\lambda) \log M_X(\lambda)$$

note: $X \sim N(0, \sigma^2)$, then $M_X(\lambda) = e^{\frac{1}{2}\lambda^2\sigma^2}$, $M'_X(\lambda) = \lambda\sigma^2 M_X(\lambda) \Rightarrow H(e^{\lambda X}) = \ldots = \frac{1}{2}\lambda^2\sigma^2 M_X(\lambda)$

㊾ **Prop 5.10 (Herbst Arguement):** $X \in \mathbb{R}$ R.V. & suppose for $\sigma > 0$, $\overbrace{H(e^{\lambda X}) \leq \frac{1}{2}\sigma^2\lambda^2 M_X(\lambda)}^{\text{bound on the entropy}}$ for $\lambda \in I$ w/ $I = [0, \infty)$ or $\mathbb{R}$, then $\underbrace{\log \mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] \leq \frac{1}{2}\lambda^2\sigma^2}_{\text{bound on centered moment generating f.}}$ $\forall \lambda \in I$

note: $I = \mathbb{R}$, then bd $\Rightarrow$ $X - \mathbb{E}[X]$ is subgaussian w/ param σ

$I = [0, \infty)$, then bd $\Rightarrow$ $\mathbb{P}(X \geq \mathbb{E}[X] + t) \leq e^{-t^2/2\sigma^2}$ $\quad t \geq 0$ $\quad$, $I = \mathbb{R} \rightarrow 2$ tail

Pf: $I = [0, \infty)$, assumption $\Longrightarrow$ $\overbrace{\lambda M'_X(\lambda) - M_X(\lambda) \log M_X(\lambda) \leq \frac{1}{2}\sigma^2\lambda^2 M_X(\lambda)}^{(*)}$ $\quad \lambda \geq 0$

recall $\mathbb{E}[X^k] = \frac{\partial^k}{\partial \lambda^k} M_X(\lambda)\Big|_{\lambda=0}$ $\therefore M_X(0) = 1$

Define $G(\lambda) := \frac{1}{\lambda} \log M_X(\lambda)$ for $\lambda \neq 0$, $\overbrace{G(0) = \lim_{\lambda \to 0} G(\lambda)}^{\text{extend to 0 via L'Hopital's rule!}} = \frac{d}{d\lambda}\Big|_{\lambda=0} \log M_X(\lambda) = \frac{M'_X(\lambda)}{M_X(\lambda)}\Big|_{\lambda=0} = \mathbb{E}[X]$

MGF exist $\Rightarrow$ $G'(\lambda) = \frac{1}{\lambda}\frac{M'_X(\lambda)}{M_X(\lambda)} - \frac{1}{\lambda^2} \log M_X(\lambda)$ so $(*)$ becomes $G'(\lambda) \leq \frac{1}{2}\sigma^2$ $\ \forall \lambda \in [0, \infty)$

Integrate inequality: $\forall \lambda_0 > 0$ $\quad G(\lambda) - G(\lambda_0) \leq \frac{1}{2}\sigma^2(\lambda - \lambda_0)$, let $\lambda_0 \downarrow 0$ & $G(0)$ yields

$$G(\lambda) - \mathbb{E}[X] = \frac{1}{\lambda}\left(\log M_X(\lambda) - \log e^{\lambda \mathbb{E}[X]}\right) \leq \frac{1}{2}\sigma^2\lambda$$ ■

㊿ **Def 5.15:** Concentration of functions of many R.V.s

(a) $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ separately convex if $\forall k \in \{1, \ldots, n\}$ $f_k : \mathbb{R} \longrightarrow \mathbb{R}$ is convex $\forall \begin{pmatrix} x_1 \\ \vdots \\ x_{k-1} \\ x_{k+1} \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^{n-1}$ $\quad y_k \longmapsto f_k(y_k) = f\begin{pmatrix} x_1 \\ \vdots \\ x_{k-1} \\ y_k \\ x_{k+1} \\ \vdots \\ x_n \end{pmatrix}$

(b) $f : X \longrightarrow Y$ Lipschitz cts if $\forall x \in X$ $\exists U \subset X$ s.t. $f|_U$ globally lipschitz cts

(c) $f : X \longrightarrow Y$ L-Lipschitz cts if $\exists L \in \mathbb{R}$ s.t. $d_Y(f(x), f(y)) \leq L\, d_X(x, y)$ $\forall x, y \in X$

Smallest such $L := \|f\|_{Lip}$

(51) **Thm 5.16 (Tail bds for Lipschitz functions):** $X \in \mathbb{R}^n$, indi coords w/ $x_i \in [a, b]$, $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ separately

convex, f L-lipschitz cts wrt $\|\cdot\|_2$, Then $\forall t \geq 0$

$$\mathbb{P}(f(X) \geq \mathbb{E}[f(X)] + t) \leq \exp\left(-\frac{t^2}{4L^2(b-a)^2}\right)$$

Pf: long, dense, stripped... need to know

ⓐ $\|\nabla f(X)\|_2^2 = \sum_{i=1}^{n}\left(\frac{\partial f(X)}{\partial x_k}\right)^2 \leq L^2$ a.s.

ⓑ Tower property; use conditional expectations, fix some data & consider the outcomes on unfixed data...

note: can apply thm by showing f lipschitz & separably cts... e.g. $f: \mathbb{R}^{n\times d} \longrightarrow \mathbb{R}$, $M \longmapsto \|M\|$

$\big| \|M\| - \|M'\| \big| \leq \|M - M'\| \leq \|M - M'\|_F$ ← frobenius is $\|\cdot\|_2$ on $\mathbb{R}^{n\times d}$

operator

Idea: extend $X \sim Unif(\sqrt{n}S^{n-1})$ then $f: \mathbb{R}^n \to \mathbb{R}$ & $f(X)$ sub gaussian, DEPENDANCE!

Part 2 - Concentration via Isoperimetric Inequalities

(52) Thm 5.24 (Isoperimetric Inequality on $\mathbb{R}^n$): Among all subsets $A \subset \mathbb{R}^n$ w/ given volume, Euclidian balls have minimal surface area. Moreover, $\forall \varepsilon > 0$ euclidian balls minimise the volume of the ε-neighbourhood of A. $A_\varepsilon = \{x \in \mathbb{R}^n : \exists y \in A \ \|x-y\|_2 \leq \varepsilon\} = A + \varepsilon B^{(n)}$

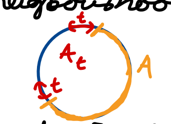
Pf: omitted (just analysis)

Spherical cap $c(a,\varepsilon) = \{x \in S^{n-1} : \|x - a\|_2 \leq \varepsilon\}$

(53) Thm 5.25 (Isoperimetric inequality on sphere): Let $\varepsilon > 0$, then among all $A \subset S^{n-1}$ w/ given area $\sigma_{n-1}(A)$, the spherical cap minimises the area of the ε-neighbourhood $\sigma_{n-1}(A_\varepsilon)$

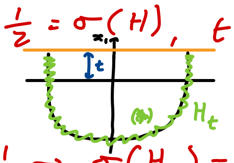
Pf: no pf (note $\|\cdot\|_2$, NOT spherical metric from geometry)

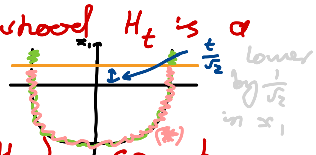
$\sigma(A) = \frac{\sigma_{n-1}(A)}{\sigma_{n-1}(S^{n-1})}$



(54) Lemma 5.26 (Blow-Up lemma): For any $A \subset \sqrt{n}S^{n-1}$, let σ denote the normalised area on the sphere. If $\sigma(A) \geq \frac{1}{2}$, then $\sigma(A_t) \geq 1 - 2e^{-ct^2}$ for some a.c. $c > 0$

Pf: Hemisphere $H = \{x \in \sqrt{n}S^{n-1} \cdot x_1 < 0\}$, $\sigma(A) \geq \frac{1}{2} = \sigma(H)$, t-neighbourhood H_t is a spherical cap, so (53) $\Rightarrow \sigma(A_t) \geq \sigma(H_t)$





Copy (38), σ is uniform $\mathbb{P}$ measure on $\sqrt{n}S^{n-1} \Rightarrow \sigma(H_t) = \mathbb{P}(X \in H_t)$, so by thm, $X \sim Unif(\sqrt{n}S^{n-1})$ is sub-gaussian. Note that $\overset{(*)}{H_t} \supset \underbrace{\{x \in \sqrt{n}S^{n-1} : x_1 \leq \frac{t}{\sqrt{2}}\}}_{(*)}$

(36) all 1d marginals are sub gaussian

Hence $\sigma(A_t) \geq \sigma(H_t) \geq \mathbb{P}(X_1 \leq \frac{t}{\sqrt{2}}) = 1 - \mathbb{P}(X_1 > \frac{t}{\sqrt{2}}) \geq 1 - 2e^{-ct^2}$ ∎

This extends (38) to non-linear functions!

(55) Thm 5.22 (Concentration of Lipschitz functions on the sphere): Let $f: \sqrt{n}S^{n-1} \longrightarrow \mathbb{R}$ be a lipschitz function and $X \sim Unif(\sqrt{n}S^{n-1})$. Then $\|f(X) - \mathbb{E}[X]\|_{\psi_2} \leq C\|f\|_{Lip}$ for some abs. const. $C > 0$

ie X is subgaussian...

Note: This implies $\forall t \geq 0$ $\mathbb{P}(|f(X) - \mathbb{E}[X]| \geq t) \leq 2\exp\left(-\frac{ct^2}{\|f\|_{Lip}^2}\right)$ for some a.c. $c > 0$

Pf: WLOG, $\|f\|_{Lip} = 1$, define median: $\mathbb{P}(f(X) \leq M) \geq \frac{1}{2}$ and $\mathbb{P}(f(X) \geq M) \geq \frac{1}{2}$.

$A = \{x \in \sqrt{n}S^{n-1} : f(x) \leq M\}$ has $\mathbb{P}(X \in A) \geq \frac{1}{2}$. $\overset{\text{Blow up}}{\overset{(54)}{\Rightarrow}} \mathbb{P}(X \in A_t) \geq 1 - 2e^{-ct^2}$ [$c > 0$ a.c.]

Claim: $\forall t \geq 0$ $\mathbb{P}(X \in A_t) \leq \mathbb{P}(f(X) \leq M + t)$ [$\Rightarrow \mathbb{P}(f(X) \leq M + t) \geq 1 - 2e^{-ct^2}$]

Pf: $X \in A_t \Rightarrow \|X - y\|_2 \leq t$ for some $y \in A$. Def A $\Rightarrow f(y) \leq M$, $\|f\|_{Lip} = 1$

So $f(X) - f(y) \leq \|f(X) - f(y)\|_2 \leq \|X - y\|_2$ [f lipschitz]

$\Leftrightarrow f(X) \leq f(y) + \|X - y\|_2 \leq M + t \Rightarrow$ claim

Repeat w/ $-f \Rightarrow \mathbb{P}(f(X) \geq M - t) \geq 1 - 2e^{-\tilde{c}t^2}$, combining the two estimates yields $\mathbb{P}(|f(X) - M| \leq t) = 1 - 2\exp(-\hat{c}t^2) \Rightarrow \|f(X) - M\|_{\psi_2} \leq C$

For centering, WTS $\|f(X)\|$ S.G.

Replace $M \to \mathbb{E}[X]$: $\big| \|f(X)\|_{\psi_2} - \|M\|_{\psi_2} \big| \overset{\text{reverse }\triangle}{\leq} \|f(X) - M\|_{\psi_2} \leq C$,

$\|M\|_{\psi_2} \leq \tilde{C} \Rightarrow -\tilde{C} + \|f(X)\|_{\psi_2} \leq C \Rightarrow \underbrace{\|f(X)\|_{\psi_2} \leq C + \tilde{C}}_{\text{subgaussian!}} \xrightarrow[\text{centering}]{(14)} \|f(X) - \mathbb{E}[f(X)]\|_{\psi_2} \leq C$ ∎

(56) **Thm 5.30 (Matrix Bernstein Inequality):** Let $X_1, \dots, X_N$ be independent mean zero random $n \times n$ symmetric matrices s.t. $\|X_i\| \le k$ a.s. $\forall i \in [N]$, then $\forall t \ge 0$

$\{n \neq N\}$ ↗ $\mathbb{R}^{n \times n}$ ↖ N R.V.

$$\mathbb{P}\left(\Big\|\sum_{i=1}^{N} X_i \Big\| \ge t \right) \le 2n \exp\left(- \frac{t^2/2}{\sigma^2 + kt/3} \right) \qquad \sigma^2 = \underbrace{\Big\| \sum_{i=1}^{N} \mathbb{E}[X_i^2] \Big\|}_{\text{norm of the matrix variance of the sum}}$$

Pf: Whole estimate "too hard", but lots of named results needed... Think $e^x = \sum_{i=1}^{n} \frac{x^k}{k!}$

can calc expectation bd, but just computation

-values ↙ -vectors ↙ given

ⓐ **Def 5.31:** $X \in \mathbb{R}^{n \times n}$ sym. λ_i, u_i, $f: \mathbb{R} \longrightarrow \mathbb{R}$, function of a matrix $\underbrace{f(X) := \sum_{i=1}^{n} f(\lambda_i) u_i u_i^T}$.
If $X \in \mathbb{R}^{n \times n}$ positive semidefinite, write $X \succcurlyeq 0$, so if $\lambda_i(M) \ge 0 \ \forall i \Rightarrow M \succcurlyeq 0$
For $Y \in \mathbb{R}^{n \times n}$, $X \succcurlyeq Y \iff Y \preccurlyeq X \iff X - Y \succcurlyeq 0$

ⓑ **Golden - Thompson Inequality:** $A, B \in \mathbb{R}^{n \times n}$ symmetric, then $\mathrm{Trace}(e^{A+B}) \le \mathrm{Trace}(e^{A}) \, \mathrm{Trace}(e^{B})$

ⓒ **Lieb's Inequality:** $H \in \mathbb{R}^{n \times n}_{sym}$ $f(X) := \mathrm{Trace}(\exp(H + \log X)) \Rightarrow f$ concave on $\mathbb{R}^{n \times n}_{pos. def}$.

ⓓ **Bd on MGF:** $X \in \mathbb{R}^{n \times n}_{sym}$ mean zero, random, $\|X\| < k$ a.s. Then

$$\mathbb{E}[e^{\lambda X}] \le \exp\left(g(\lambda) \mathbb{E}[X^2] \right) \quad \text{where} \quad g(\lambda) = \frac{\lambda^2/2}{1 - |\lambda| k/3} \quad \text{for } |\lambda| < 3/k$$

Pf: For $z \in \mathbb{C}$,

$$e^z = 1 + z + \sum_{k \ge 2} \frac{z^k}{k!} \le 1 + z + \frac{z^2}{2} \overbrace{\sum_{k \ge 3} \frac{z^{k-2}}{3^{k-2}}}^{\frac{1}{1 - \frac{|z|}{3}}} = 1 + z + \frac{1}{1 - \frac{|z|}{3}} \times \frac{z^2}{2} \qquad \because k! \le 2 \times 3^{k-2}$$

Put $z = \lambda x$, $|x| < k$, $\overbrace{|z| < 3}^{\text{geom. series}} \Rightarrow |\lambda| k < 3 \Rightarrow |\lambda| < \frac{3}{k} \Rightarrow e^{\lambda x} \le 1 + \lambda x + g(\lambda) x^2$

used in applications
↳ Foundations of Data Science

Convert from scalar to matrix $\Rightarrow e^{\lambda X} \preccurlyeq \mathbb{1}_n + \lambda X + g(\lambda) X^2$, take exp & $1 + x \le e^x$

(57) **Thm 5.38 (Johnson-Lindenstrauss Lemma):** Let $\mathcal{X} = \overbrace{\{x_1, \dots, x_N\}}^{N \text{ data pts}}$, $X_i \in \mathbb{R}^n$, pick $\varepsilon > 0$
$G_{n,m} = \{$all m dimensional vector subspaces of $\mathbb{R}^n\}$ (aiming for $m << n$), then $\underbrace{E \sim \mathrm{Unif}(G_{n,m})}_{\text{sample on smaller space}}$ sampled with orthogonal projection P into E, then w/ high probability $1 - 2e^{-c\varepsilon^2 m}$, $Q := \sqrt{\frac{n}{m}} P$ is an approximate $\underbrace{\text{isometry}}_{\text{norm of the mostl same.}}$ of the set $\mathcal{X}$, that is

smaller subspace for pairwise search

$$(1 - \varepsilon) \|x - y\|_2 \le \|Qx - Qy\|_2 \le (1 + \varepsilon) \|x - y\|_2$$

Pf: Not covered. Basically you can efficiently represent a large dataset in a smaller one! See Foundations of Data Science p. 25.

Chapter 6 - Basic Tools in High Dimensional Probability

① Decoupling

(58) **Def 6.1 (Chaos):** $X_1, \dots, X_n$ indi $\mathbb{R}$ R.V. $a_{ij} \in \mathbb{R}$, $i, j \in [n]$. The random quadratic form below is called chaos
1

$$\sum_{i,j=1}^{n} a_{ij} X_i X_j = X^T A X = \langle X, AX \rangle \qquad x \in \mathbb{R}^n, \quad A = (a_{ij})$$

Note: For ease $\mathbb{E}[X_i] = 0$, $\mathrm{Var}(X_i) = 1$, then $\mathbb{E}[\langle X, AX \rangle] = \sum_{i,j=1}^{n} a_{ij} \mathbb{E}[X_i X_j] = \sum_{i=1}^{n} a_{ii} = \mathrm{Trace}(A)$
To overcome independence, use {decoupling technique}, take X' as an identical independent copy & condition on X'

(59) **Lemma 6.2:** $Y, Z \in \mathbb{R}^n$, $Y \perp Z$ s.t. $\mathbb{E}[Y] = \mathbb{E}[Z] = 0$. Then $\forall F: \mathbb{R}^n \to \mathbb{R}$ convex, $\overbrace{\mathbb{E}}^{\text{exp. w/ } Y}[F(Y)] \le \mathbb{E}[F(Y + Z)]$ ↙ joint $\mathbb{E}$ wrt Y and Z
1

Pf: Fix $y \in \mathbb{R}^n$, $F(y) = F(y + \overbrace{\mathbb{E}_Z[Z]}^{= 0}) \le \overbrace{\mathbb{E}_Z[F(y + Z)]}^{F \text{ is convex}}$, then pick $y = Y$, take $\mathbb{E}_Y$

$$\mathbb{E}_Y[F(Y)] \le \mathbb{E}_Y[F(Y + \mathbb{E}_Z[Z])] = \mathbb{E}_Y[F(\mathbb{E}_Z[Y + Z])] \le \mathbb{E}_Y \otimes \mathbb{E}_Z[F(Y + Z)] \qquad \blacksquare$$

Note, as Y, Z independent, $\mathbb{E}_{Y,Z} = \mathbb{E}_Y \otimes \mathbb{E}_Z$ is just a product of integrals...

(60) **Thm 6.3 (Decoupling):** $A \in \mathbb{R}^{n \times n}$ $\overbrace{\text{diagonal}}^{a_{ii}=0 \ \forall i \in [n]}$ free, $X \in \mathbb{R}^n$ R. vector w/ independent mean
1 zero coordinates X_i, X' independent copy of X. Then for every convex $F: \mathbb{R} \longrightarrow \mathbb{R}$,

$$\mathbb{E}[\underbrace{F(<X, AX>)}_{\text{convex image of chaos}}] \leq \mathbb{E}[\underbrace{F(4<X, AX'>)]}_{\text{expectation of the mixed chaos}}$$

replace X w/ an identical copy & pay a factor of 4!

Pf: Idea: Study partial chaos $\sum_{i,j \in I \times I^c} a_{ij} X_i X_j$ w/ $I \subset \{1, \dots, n\}$ a random subset.

(Note that $\because$ $a_{ii} = 0$, no square terms so can safely take $i,j \in I \times I^c$

Let $\delta_1, \dots, \delta_n$ be n indi Bernoulli R.V. w/ $\mathbb{P}(\delta_i = 0) = \mathbb{P}(\delta_i = 1) = \frac{1}{2}$, $I := \{i : \delta_i = 1\}$

note: $\mathbb{E}[\delta_i(1-\delta_j)] = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Then $<X, AX> = \sum_{i \neq j} a_{ij} X_i X_j = \overset{\text{just sneak this in!}}{4 \mathbb{E}_\delta}[\sum_{i \neq j} \delta_i (1-\delta_j) a_{ij} X_i X_j] = 4 \mathbb{E}_I [\sum_{i,j \in I \times I^c} a_{ij} X_i X_j]$

$$\mathbb{E}_X[F(<X, AX>)] \leq \mathbb{E}_I \mathbb{E}_X [F(4 \sum_{(i,j) \in I \times I^c} a_{ij} X_i X_j)]$$ [apply F, Jensen wrt I, Fubini]

Fix an I in above. $(X_i)_{i \in I}$ independent $\overset{}{\underset{(*)}{\text{of}}}$ $(X_j)_{j \in I^c}$ $\Rightarrow$ dist of sum same if $X_j \xrightarrow{j \in I^c} X_j'$

Hence $$\mathbb{E}_X[F(<X, AX>)] \leq \overbrace{\mathbb{E}_X[F(4 \underbrace{\sum_{i,j \in I \times I^c} a_{ij} X_i X_j'}_{Y + Z_1 + Z_2})]}^{}$$

WTS $(*) \leq$ $\mathbb{E}[F(4 \sum_{i,j \in [n] \times [n]} a_{ij} X_i X_j')]$ $[n] = \{1, \dots, n\}$

$Y := \sum_{i,j \in I \times I^c} a_{ij} X_i X_j'$,

$Z_1 := \sum_{i,j \in I \times I} a_{ij} X_i X_j'$

$Z_2 := \sum_{i,j \in I^c \times [n]} a_{ij} X_i X_j'$

condition on all R.V. BAR $(X_j')_{j \in I}$ and $(X_i)_{i \in I^c}$ $\Downarrow$ denote $\tilde{\mathbb{E}}$ Y fixed

$$\tilde{\mathbb{E}}[Z_1] = \sum_{i,j \in I \times I} a_{ij} X_i \overbrace{\mathbb{E}[X_j']}^{=0} = 0$$

$$\tilde{\mathbb{E}}[Z_2] = \sum_{i,j \in I^c \times [n]} a_{ij} X_j' \overbrace{\mathbb{E}[X_i]}^{=0} = 0$$

now apply (59)

keep Y fixed ; expectation wrt I

$$F(4Y) \leq \tilde{\mathbb{E}}[F(4Y + \underbrace{4Z_1 + 4Z_2}_{\text{sneak in mean zero random variables}})] \Longrightarrow \mathbb{E}[F(4Y)] \leq \mathbb{E}[F(\overbrace{4Y + 4Z_1 + 4Z_2}^{4<X, AX'>})]$$ ∎

(61) **Thm 6.4 (Hanson-Wright Inequality):** $X \in \mathbb{R}^n$ random vector w/ independent mean zero subgaussian coordinates. $A \in \mathbb{R}^{n \times n}$. Then $\forall t \geq 0$

$K = \max_{1 \leq i \leq n} \{\|X_i\|_{\psi_2}\}$

$$\mathbb{P}(|<X, AX> - \mathbb{E}[<X, AX>]| \geq t) \leq 2\exp\left(-c \min\left\{\left(\frac{t^2}{K^4 \|A\|_F^2}\right), \left(\frac{t^2}{K^2 \|A\|}\right)\right\}\right)$$

Note proof needs: (stripped proof... maybe revisit...)

ⓐ **Lemma 6.5 (MGF Gaussian chaos):** $X, X' \sim N(0, \mathbb{I}_n)$, $X \perp X'$, $A \in \mathbb{R}^{n \times n}$, Then

$$\mathbb{E}[\exp(\lambda <X, AX'>)] \leq \exp(C \lambda^2 \|A\|_F) \quad \forall |\lambda| \leq \frac{c}{\|A\|} \quad [\text{some a.c. } C > 0]$$

$A = \sum_{i=1}^n s_i \begin{pmatrix} u_{i_1} \\ \vdots \\ u_{i_n} \end{pmatrix} (v_{i_1} \cdots v_{i_n})$

$= \sum_{i=1}^n s_i \begin{pmatrix} u_{i_1}v_{i_1} & u_{i_1}v_{i_2} & \cdots & u_{i_1}v_{i_n} \\ \vdots & & \ddots & \\ u_{i_n}v_{i_1} & \cdots & & u_{i_n}v_{i_n} \end{pmatrix}$

$\Rightarrow A_{km} = \sum_{i=1}^n s_i u_{i_k} v_{i_m}$

$A_{ij} = \sum_{k=1}^n s_k u_{k_i} v_{k_j}$

$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \cdot \begin{pmatrix} \sum_{j=1}^n A_{1j} X_j' \\ \vdots \\ \sum_{j=1}^n A_{nj} X_j' \end{pmatrix} = \sum_{i,j=1}^n A_{ij} X_i X_j'$

$= \sum_{i,j=1}^n \sum_{k=1}^n s_k u_{k_i} v_{k_i} X_i X_j'$

$= \sum_{k=1}^n s_k <u_k, X_k> <u$

Pf: Using SVD of $A = \overset{(43)}{\sum_{i=1}^n} s_i u_i v_i^T \Longrightarrow <X, AX'> = \sum_{i=1}^n s_i \underbrace{<u_i, X>}_{= Y_i \sim N(0, \mathbb{1}_n)} \underbrace{<v_i, X'>}_{= Y_i' \sim N(0, \mathbb{1}_n)}$

$Y = \begin{pmatrix} <u_1, X> \\ \vdots \\ <u_n, X> \end{pmatrix}$

Independence: $\mathbb{E}[\exp(\lambda <X, AX'>)] = \prod_{i=1}^n \mathbb{E}[\exp(\lambda s_i Y_i Y_i')]$

$X \sim N(0,1) \Rightarrow \mathbb{E}[e^{\lambda X}] = e^{\lambda^2/2}$

For each $i \in [n]$, $\underbrace{\mathbb{E}[\exp(\lambda s_i Y_i Y_i') | Y]}_{\text{conditional } \mathbb{E}_{Y'} \ [Y \text{ fixed}]} = \mathbb{E}[\exp(\lambda^2 s_i^2 Y_i^2 / 2)] \leq \exp(C \lambda^2 s_i^2)$

Y_i sub-gauss, Y_i^2 sub-exp $\xrightarrow{(v)}$ $\lambda^2 s_i^2 \leq C$

Together: $\mathbb{E}[\exp(\lambda <X, AX'>)] \leq \exp(C\lambda^2 \underbrace{\sum_{i=1}^n s_i^2}_{\|A\|_F^2})$ w/ $\lambda \leq \frac{c}{\underbrace{\max_{1 \leq i \leq n} \{s_i^2\}}_{\|A\|}}$ ∎

ⓑ **Lemma 6.6 (comparision):** $X \perp X' \in \mathbb{R}^n$ mean zero subgaussian random vectors w/ $\|X\|_{\psi_2}, \|X'\|_{\psi_2} \leq K$. $Y, Y' \sim N(0, \mathbb{1}_n)$, $Y \perp Y'$. $A \in \mathbb{R}^{n \times n}$. Then

$$\mathbb{E}[\exp(\lambda <X, AX'>)] \leq \mathbb{E}[\exp(C \lambda K^2 <Y, AY'>)]$$ $\lambda \in \mathbb{R}$

Pf: condition on $X' \Rightarrow <X, AX'>$ subgaussian $\Rightarrow \mathbb{E}_X[\exp(\lambda <X, AX'>)] \leq \exp(C \lambda^2 K^2 \|AX'\|_2^2)$

$\mathbb{E}_Y[\exp(\mu <Y, AX'>)] = \exp(\mu^2 \|AX'\|_2^2 / 2)$ $\mu \in \mathbb{R}$, choose $\mu = \sqrt{2C\lambda} K$

$\mathbb{E}_X[\exp(\lambda <X, AX'>)] \leq \mathbb{E}_Y[\exp(\mu <Y, AX'>)] = \exp(C\lambda^2 K^2 \lambda \|AX'\|_2^2)$, $\mathbb{E}_{X'}$ & repeat for X' ∎

Pf Hanson-Wright: WLOG, $K=1$, show 1 sided tail estimate. $p := \mathbb{P}(\langle X, AX\rangle - \mathbb{E}[\langle X, AX\rangle] \geq t)$

$$\langle X, AX\rangle - \mathbb{E}[\langle X, AX\rangle] = \underbrace{\sum_{i=1}^{n} a_{ii}(X_i^2 - \mathbb{E}[X_i^2])}_{(*)} + \underbrace{\sum_{i,j:i\neq j} a_{ij} X_i X_j}_{(**)} \quad [\because \mathbb{E}[a_{ij} X_i X_j] = 0]$$

So! Just estimate

$$p \leq \underbrace{\mathbb{P}((*) \geq t/2)}_{p_1} + \underbrace{\mathbb{P}((**) \geq t/2)}_{p_2}$$

$\boxed{\text{Diagonal Sum}}$: X_i sub gauß $\Rightarrow$ X_i^2 sub exp. $\Rightarrow$ $X_i^2 - \mathbb{E}[X_i^2]$ mean zero sub exponential.

Centering (17), $\|X_i^2 - \mathbb{E}[X_i^2]\|_{\psi_1} \leq C\|X_i^2\|_{\psi_1} \leq C\|X_i\|_{\psi_2}^2 \leq C$ [X_i sub-gaussian]

Bernstein (23): $p_1 \leq \exp\left(-C \min\left\{\frac{t^2}{\sum a_{ii}^2}, \frac{t}{\max\{\|a_{ii}\|\}}\right\}\right)$

$\boxed{\text{Off Diagonal sum}}$: Set $S = \sum_{i\neq j} a_{ij} X_i X_j \overset{\text{Markov}}{\underset{\lambda>0}{\Longrightarrow}}$ $\mathbb{E}[e^{\lambda S}] \leq \mathbb{E}[\exp(4\lambda \langle X, AX'\rangle)]$ [decoupling (60)] X_i are sub gaussian

$\Downarrow$ $\mathbb{P}(X \geq t/2) \leq e^{-\frac{t\lambda}{2}} \mathbb{E}[e^{\lambda S}] \leq \mathbb{E}[\exp(4\lambda \langle Y, AY'\rangle)]$ [comparison (61b)] $Y \sim N(0, \mathbb{1}_n)$

$\leq \exp(c\lambda^2 \|A\|_F^2)$ [Gaußian Chaos (61a)]

$$p_2 \leq e^{-\frac{\lambda}{2}t}\, \mathbb{E}[e^{\lambda S}] \overset{\text{Markov}}{\leq} \exp\left(-\frac{\lambda t}{2} + c\lambda^2 \|A\|_F^2 t^2\right) \quad \forall |\lambda| < \frac{c}{\|A\|}$$

Optimize over $0 \leq \lambda \leq \frac{c}{\|A\|}$ for $t \leq \underset{\text{quad}}{\cdots}$, then for $t \geq \cdots$, linear ■

② concentration for vectors of the form BX, B fixed, X isotropic

(62) **Thm 6.8 (Concentration for random vectors):** $B \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^n$ random vector w/ independent mean zero subgaussian coordinates. $\mathbb{E}[X_i^2] = 1$. Then

$$\big\| \|BX\|_2 - \|B\|_F \big\|_{\psi_2} \leq CK^2 \|B\| \qquad K = \max_{1\leq i\leq n}\{\|X_i\|_{\psi_2}\}$$

Pf: will skip & often of exam b.

③ Symmetrisation

(63) **Def 6.10 (Symmetric R.V.s):** $X \in \mathbb{R}$ symmetric if X & $-X$ have same distribution.

| ↳ If $X \in \mathbb{R}$, ε sym. Bernoulli, $X \perp \varepsilon$, then

- ⓐ εX and $\varepsilon|X|$ sym & $\varepsilon X \sim \varepsilon|X|$ law of total expectation
- ⓑ X sym $\Rightarrow$ $\varepsilon X \sim \varepsilon|X|$ compute $\mathbb{P}(\varepsilon X \leq c) = \cdots$
- ⓒ X' ind. copy of X $\Rightarrow$ $X - X'$ symmetric R.V.
  - ↳ use that $(X, X') \sim (X', X) \Rightarrow f(x, x') = x - x'$ applied to $f_{X,X'}$ and $f_{X',X}$ the same.

$\mathbb{P}(\varepsilon X \geq a) = \mathbb{P}(X \geq a \mid \varepsilon = 1)\mathbb{P}(\varepsilon = 1) + \mathbb{P}(-X \geq a \mid \varepsilon = -1)\mathbb{P}(\varepsilon = -1)$
$= \frac{1}{2}(\mathbb{P}(X \geq a) + \mathbb{P}(-X \geq a))$

$\mathbb{P}(\varepsilon|X| \geq a) = \mathbb{P}(|X| \geq a \mid \varepsilon = 1)\mathbb{P}(\varepsilon = 1) + \mathbb{P}(-|X| \geq a \mid \varepsilon = -1)\mathbb{P}(\varepsilon = -1)$
$= \underbrace{\frac{1}{2}(\mathbb{P}(|X| \geq a) + \mathbb{P}(-|X| \geq a))}_{=(*)}$

If $X \geq 0$, $(*) = \frac{1}{2}(\mathbb{P}(X \geq a) + \mathbb{P}(-X \geq a))$
If $X < 0$, $(*) = \frac{1}{2}(\mathbb{P}(-X \geq a) + \mathbb{P}(X \geq a))$
so in either case $(*) = \mathbb{P}(\varepsilon X \geq a)$
$\Rightarrow$ $\varepsilon X \sim \varepsilon|X|$

(64) **Lemma 6.12 (Symmetrisation):** Let $X_1, \ldots, X_N$ be independent mean zero random vectors

|| in a normed space $(E, \|\cdot\|)$. $\varepsilon_1, \ldots, \varepsilon_N$ independent sym. Bernoulli. Show that

$$\frac{1}{2}\mathbb{E}\left[\left\|\sum_{i=1}^{N} \varepsilon_i X_i\right\|\right] \leq \overbrace{\mathbb{E}\left[\left\|\sum_{i=1}^{N} X_i\right\|\right]}^{=p} \leq 2\mathbb{E}\left[\left\|\sum_{i=1}^{N} \varepsilon_i X_i\right\|\right]$$

Pf: $\boxed{\text{Upper Bound}}$: Let $X_i' \sim X_i$ iid copy. $\mathbb{E}\left[\sum_{i=1}^{N} X_i'\right] = 0$ so as $\|\cdot\|$ convex, $\overbrace{\mathbb{E}[F(Y)] \leq \mathbb{E}[F(Y+Z)]}^{F = \|\cdot\| \text{ (59)}}$ yields

$$p \leq \mathbb{E}\left[\left\|\sum_{i=1}^{N} X_i - \sum_{i=1}^{N} X_i'\right\|\right] = \mathbb{E}\left[\left\|\sum_{i=1}^{N}(X_i - X_i')\right\|\right]$$

Note $X_i - X_i'$ symmetric by (63c) $\Rightarrow$ $X_i - X_i' \sim \varepsilon_i(X_i - X_i')$, & use Δ ineq.

(*) ↷

$\mathbb{E}\left[\left\|\sum \varepsilon_i(X_i - X_i')\right\|\right] = \mathbb{E}\left[\mathbb{E}\left[\left\|\varepsilon_1(X_1 - X_1') + \sum_{i=2}^{N} \varepsilon_i(X_i - X_i')\right\|\right] \Big| \varepsilon_1\right]$ $(*) \sim (**)$

$= \mathbb{E}\left[\frac{1}{2}\overbrace{\left\|(X_1 - X_1') + \sum_{i=2}^{N} \varepsilon_i(X_i - X_i')\right\|}^{(*)} + \frac{1}{2}\overbrace{\left\|-(X_1 - X_1') + \sum_{i=1}^{N} \varepsilon_i(X_i - X_i')\right\|}^{(**)}\right]$

$= \mathbb{E}\left[\left\|(X_1 - X_1') + \sum_{i=1}^{N} \varepsilon_i(X_i - X_i')\right\|\right]$

$\boxed{\text{Lower Bound}}$: $\mathbb{E}\left[\left\|\sum_{i=1}^{N} \varepsilon_i X_i\right\|\right] \leq \mathbb{E}\left[\left\|\sum_{i=1}^{N} \varepsilon_i(X_i - X_i')\right\|\right] \overset{\text{condition on } \varepsilon_i (*)}{=} \mathbb{E}\left[\left\|\sum_{i=1}^{N} X_i - X_i'\right\|\right] \leq \overbrace{\mathbb{E}\left[\left\|\sum_{i=1}^{N} X_i\right\|\right] + \mathbb{E}\left[\left\|\sum_{i=1}^{N} X_i'\right\|\right]}^{2\times -}$

Chapter 7 - Random Processes

(65) **Def 7.1:** A random process over a indexing set T, is a collection of random variables $X = \{X_t\}_{t\in T}$

E.g: $T = \mathbb{N}$, $X_n = \sum_{i=1}^{n} Z_i$ $Z_i \in \mathbb{R}$ iid, then X is the discrete time random walk.

(66) **Def 7.3:** $(X_t)_{t\in T}$ random process with $\mathbb{E}[X_t] = 0$ $\forall t \in T$, then

(i) The covariance function of the process $\Sigma(t,s) = \operatorname{Cov}(X_t, X_s)$ $t, s \in T$

(ii) The increments of the process are defined as $d(t,s) = \|X_t - X_s\|_{L^2} = (\mathbb{E}[(X_s - X_t)^2])^{\frac{1}{2}}$ $t, s \in T$

E.g. Random walks: $\overset{n\geq m}{d(n,m)^2} = \|X_n - X_m\|_{L^2}^2 = \left\|\sum_{i=m+1}^{n} Z_i\right\|_{L^2}^2 = \mathbb{E}\left[\left(\sum_{i=m+1}^{n} Z_i\right)^2\right] = \sum_{i,j=m+1}^{n} \mathbb{E}[Z_i Z_j] = \sum_{i=m+1}^{n} \overset{\mathbb{E}[Z_i^2]=1}{\mathbb{E}[Z_i^2]} = n - m$

(67) Def 7.5 (Gaussian Process): (a) A random process is a Gaussian process if $\forall$ finite $T_0 \subset T$, the random vector $(X_t)_{t \in T_0}$ has a normal distribution. Equiv, $(X_t)_{t\in T}$ is gaussian process if every finite linear combination $\sum_{t \in T_0} a_t X_t$ $a_t \in \mathbb{R}$ is normally distributed. Also (b), $T \subset \mathbb{R}^n$, $Y \sim N(0, \mathbb{1}_n)$, $X_t := \langle Y, t \rangle$ $t \in T \subset \mathbb{R}^n$, then $(X_t)_{t \in T}$ is the canonical gaussian process in $\mathbb{R}^n$.

(68) Gaussian Interpolation: T finite, $X = (X_t)_{t\in T}$, $Y = (Y_t)_{t \in Y}$ gaussian random vectors. Gaussian interpolation $Z(u) := \sqrt{u} X + \sqrt{1-u}\, Y$ $u \in [0,1]$

ⓐ Covariance matrices interpolate linearly: $\Sigma(Z(u)) = u\,\Sigma(X) + (1-u)\,\Sigma(Y)$ $t \in [0,1]$

Pf: $(Z(u))_{ij} = \text{cov}(Z(u)_i, Z(u)_j) = \text{cov}(\sqrt{u}X_i + \sqrt{1-u}\,Y_i, \sqrt{u}X_j + \sqrt{1-u}\,Y_j)$

$= u\,\text{cov}(X_i, X_j) + (1-u)\,\text{cov}(Y_i, Y_j)$ $\underset{\text{[cross terms go]}}{\overset{\text{independence}}{\downarrow}}$

ⓑ Take $f: \mathbb{R}^n \to \mathbb{R}$, how does $\mathbb{E}[f(Z(u))]$ vary w/ $u \in [0,1]$?

Example: $f(x) = \mathbb{1}\{\max_{1\le i \le n}\{x_i\} \le u\}$ $x \in \mathbb{R}^n$

$\mathbb{E}[f(Z(u))] = \mathbb{P}(\max_{i=1,\dots,n}\{Z_i(u)\} \le u)$ $\uparrow$ w/ u $\Rightarrow \mathbb{E}[f(Z(1))] \ge \mathbb{E}[f(Z(0))]$

$\Rightarrow \mathbb{P}(\max_{1\le i\le n}\{X_i\} \le u) \ge \mathbb{P}(\max_{1 \le i \le n}\{Y_i\} \le u)$

ⓒ Lem 7.9 (Gaussian Integration by Parts (1d)): $X \sim N(0,1)$, $f: \mathbb{R} \to \mathbb{R}$ differentiable, then

$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$\mathbb{E}[f'(X)] = \mathbb{E}[X f(X)]$$

$(uv)' = u'v + uv'$

$\int u'v = [uv] - \int uv'$

$u' = xe^{-\frac{1}{2}x^2}$ $u = -e^{-\frac{1}{2}x^2}$

$v = f(x)$

Pf: $\mathbb{E}[Xf(X)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x f(x) e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}}\left(\left[-f(x)e^{-\frac{1}{2}x^2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f'(x) e^{-\frac{1}{2}x^2} dx \right) = \mathbb{E}[f'(X)]$

ⓓ Lem 7.10 (Gaussian IBP n-dim): $X \sim N(0, \Sigma)$, $X \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n \times n}_{\text{sym}}$ $\Sigma \succcurlyeq 0$, $f: \mathbb{R}^n \to \mathbb{R}$ diff.

$$\mathbb{E}[Xf(X)] = \Sigma\, \mathbb{E}[\nabla f(X)]$$

Pf: nasty. Go componentwise, normalise to $X = \Sigma^{\frac{1}{2}} Z$ $Z \sim N(0, \mathbb{I}_n)$ and apply 1d conditioning on all R.V. bar the component you consider.

ⓔ Lem 7.11 (Gaussian Interpolation): $X \sim N(0, \Sigma^X)$, $Y \sim N(0, \Sigma^Y)$ in $\mathbb{R}^n$. for $f: \mathbb{R}^n \to \mathbb{R}$ twice diff,

Nasty h/w Q

$$\frac{d}{du}\mathbb{E}[f(Z(u))] = \frac{1}{2}\sum_{i,j=1}^{n} (\Sigma^X_{ij} - \Sigma^Y_{ij})\, \mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u))\right]$$

$\mathbb{E}\left[\nabla f(Z(u)) \cdot \frac{dZ}{du}\right]$

Pf: chain rule: $\frac{d}{du}\mathbb{E}[f(Z(u))] = \sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial f}{\partial x_i}(Z(u)) \cdot \frac{dZ_i}{du}\right] = \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial f}{\partial x_i}(Z(u))\left(\overset{①}{\frac{X_i}{\sqrt{u}}} - \overset{②}{\frac{Y_i}{\sqrt{1-u}}}\right)\right]$

① $= \sum_{i=1}^{n} \frac{1}{\sqrt{u}} \mathbb{E}\Big[X_i \overbrace{\frac{\partial f}{\partial x_i}(Z(u))}^{:= g_i(X) = \frac{\partial f}{\partial x_i}(\sqrt{u}X + \sqrt{1-u}\,Y)}\Big]$, by (68d) $\mathbb{E}[X_i g_i(X)] = \sum_{j=1}^{n} \Sigma^X_{ij}\, \mathbb{E}\left[\frac{\partial g_i}{\partial x_j}\right]$

condition on Y

$(*) \Rightarrow = \sum_{i,j=1}^{n} \Sigma^X_{ij}\, \mathbb{E}\left[\frac{\partial^2}{\partial x_i \partial x_j}(Z(u))\right]$, take wrt X $\qquad = \underbrace{\sum_{j=1}^{n} \Sigma^X_{ij}\, \mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\sqrt{u}X + \sqrt{1-u}\,Y)\sqrt{u}\right]}_{(*)}$

② $= -\sum_{i=1}^{n} \frac{1}{\sqrt{1-u}} \mathbb{E}\Big[Y_i \overbrace{\frac{\partial f}{\partial x_i}(Z(u))}^{= g_i(X)}\Big]$, by (68d) $\mathbb{E}[Y_i g_i(X)] = \sum_{j=1}^{n} \Sigma^Y_{ij}\, \mathbb{E}\left[\frac{\partial g_i}{\partial x_j}\right] = \sum_{j=1}^{n} \Sigma^Y_{ij}\, \mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u)) \cdot \sqrt{1-u}\right]$

so ① + ② $\Rightarrow \frac{d}{du}\mathbb{E}[f(Z(u))] = \sum_{i,j=1}^{n} (\Sigma^X_{ij} - \Sigma^Y_{ij})\, \mathbb{E}\left[\frac{\partial^2}{\partial x_i \partial x_j}(Z(u))\right]$

(69) Lemma 7.12 (Slepian's Inequality – Functional Form): $X, Y \in \mathbb{R}^n$ two mean zero gaussian random vectors. Assume $\forall i,j$ (i) $\mathbb{E}[X_i^2] = \mathbb{E}[Y_i^2]$, (ii) $\mathbb{E}[(X_i - X_j)^2] \le \mathbb{E}[(Y_i - Y_j)^2]$, $f: \mathbb{R}^n \to \mathbb{R}$ twice diff s.t. $\frac{\partial^2 f}{\partial x_i \partial x_j} \ge 0$ $\forall i,j$. Then $\mathbb{E}[f(X)] \ge \mathbb{E}[f(Y)]$

Pf: Here, $\Sigma^X_{ii} = \Sigma^Y_{ii}$ and $-\mathbb{E}[X_iX_j] \le -\mathbb{E}[Y_iY_j] \Rightarrow \Sigma^Y_{ij} \le \Sigma^X_{ij}$, assuming $X \perp Y$, then

(68e) $\Rightarrow \frac{d}{du}\mathbb{E}[f(Z(u))] \ge 0$ so $\uparrow$ in u $\Rightarrow \underbrace{\mathbb{E}[f(Z(1))] \ge \mathbb{E}[f(Z(0))]}_{= \mathbb{E}[f(X)] \ge \mathbb{E}[f(Y)]}$

$Z(u) = \sqrt{u}X + \sqrt{1-u}\,Y$

$Z(0) = Y$ $\quad Z(1) = X$

note, weaker Thm that uses an approx of the maximum $h(t) = \mathbb{1}_{(-\infty, \tau)}(t)$, naff!

(70) Thm 7.14 (Sudakov - Fernique Inequality): $(X_t)_{t\in T}$, $(Y_t)_{t\in T}$ two mean zero Gaussian processes. Assume that $\forall t, s \in T$ (i) $\mathbb{E}[(X_t - X_s)^2] \leq \mathbb{E}[(Y_t - Y_s)^2]$. Then $\mathbb{E}[\sup_{t\in T}\{X_t\}] \leq \mathbb{E}[\sup_{t\in T}\{Y_t\}]$

Pf: Use an approximation for the max: $f(x) = \frac{1}{\beta}\log \sum_{i=1}^{n} e^{\beta x_i}$ $\underset{x\in\mathbb{R}}{}$, & note $f(x) \longrightarrow \max_{1\leq i\leq n}\{x_i\}$

$\frac{d}{du}\mathbb{E}[f(Z(u))] \leq 0$ [tediously apply $\boxed{68e}$] $\Rightarrow$ decreasing in u

so $\mathbb{E}[f(Z(1))] \leq \mathbb{E}[f(Z(0))] \Rightarrow \mathbb{E}[\max_{t\in T_0}\{X_t\}] \leq \mathbb{E}[\max_{t\in T_0}\{Y_t\}] \xrightarrow{T_0 \to T} \max \to \sup$

$g(\beta) = \frac{\beta}{2} + \frac{\log N}{\beta}$
$g'(\beta) = \frac{1}{2} - \frac{\log N}{\beta^2} = 0$
$g'(\beta) = 0 \Leftrightarrow \beta = \sqrt{2\log N}$

(71) Prop 7.18 (Max of normally distributed R.V.s): $Y_i \in N(0,1)$ $i = 1, \ldots, N$ independent. Then

(a) $\mathbb{E}[\max_{1\leq i\leq n}\{Y_i\}] \leq \sqrt{2\log N}$

just add all the terms so bigger!

Pf: $\mathbb{E}[\max_{1\leq i\leq n}\{Y_i\}] = \frac{1}{\beta}\mathbb{E}[\log e^{\beta\max_{1\leq i\leq n}\{Y_i\}}] \leq \frac{1}{\beta}\mathbb{E}[\log \sum_{i=1}^{N} e^{\beta Y_i}] \overset{?}{\leq} \frac{1}{\beta}\log \sum_{i=1}^{N} \underbrace{\mathbb{E}[e^{\beta Y_i}]}_{= e^{\beta^2/2}} = \frac{\beta}{2} + \frac{\log N}{\beta}$ $\beta > 0$

take $\beta = \sqrt{2\log N}$

Jensen w/ concave, not convex so switched

$= Ne^{\beta^2/2}$

(b) $\mathbb{E}[\max_{1\leq i\leq n}\{|Y_i|\}] \leq \sqrt{2\log 2N}$

$X \sim N(0,1)$
$\mathbb{E}[e^{tX}] = e^{t^2/2}$
$e^{\beta^2/2} + e^{\beta^2/2} = 2e^{\beta^2/2}$

Pf: $\mathbb{E}[\max_{1\leq i\leq N}\{|Y_i|\}] \leq \frac{1}{\beta}\mathbb{E}[\log \sum_{i=1}^{N}(e^{\beta Y_i} + e^{-\beta Y_i})] \leq \frac{1}{\beta}\log \sum_{i=1}^{N}\mathbb{E}[e^{\beta Y_i} + e^{-\beta Y_i}] = \overbrace{\frac{1}{\beta}\log N \cdot 2e^{\beta^2/2}}^{= \frac{\log 2N}{\beta} + \frac{\beta}{2}}$, $\beta = \ldots$ $\beta > 0$

(c) $\mathbb{E}[\max_{1\leq i\leq N}\{|Y_i|\}] \geq c\sqrt{2\log N}$ for some $c > 0$

Pf: ① Tail bd for normal distribution $\overset{\text{Prop 2.1}}{\mathbb{P}(Y_i > \delta) \geq \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\delta} - \frac{1}{\delta^3}\right)e^{-\delta^2/2}}$ now pick a nice $\delta \ldots$

② Analysis $\Rightarrow$ $(1 - \frac{1}{N})^N \leq \frac{1}{e}$ $\forall N \in \mathbb{N}$ $\because \lim_{n\to\infty}(1 + \frac{x}{n})^n = e^x$

③ Pick $\delta = \beta\sqrt{\log N}$, $\beta \in (0,1)$ $(\log N)^{3/2} < N^{\frac{1}{2}}$ for N large

$$\mathbb{P}(Y_i \geq \beta\sqrt{\log N}) \geq \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\beta\sqrt{\log N}} - \frac{1}{\beta^3(\log N)^{3/2}}\right)\exp\left(-\beta^2\log N/2\right) = \cdots \geq \frac{1}{N}$$

④ max: $\mathbb{P}(\max_{1\leq i\leq n}\{Y_i\} \geq \beta\sqrt{\log N}) = 1 - \mathbb{P}(\max_{1\leq i\leq n}\{Y_i\} \leq \beta\sqrt{\log N}) = 1 - \mathbb{P}(Y_i \leq \beta\sqrt{\log N})^N$

⑤ $\mathbb{E}[\max_{1\leq i\leq N}\{Y_i\}] \geq \mathbb{E}[\max_{1\leq i\leq n}\{Y_i\}\mathbb{1}\{\max_{1\leq i\leq N}\{Y_i\} \geq \beta\sqrt{\log N}\}] = 1 - (1 - \mathbb{P}(Y_i \geq \beta\sqrt{\log N}))^N$

$\geq \underbrace{(1 - \frac{1}{e})\beta}_{c}\sqrt{\log N}$ For any β, find an N. $\geq 1 - (1 - \frac{1}{N})^N$

$\geq 1 - \frac{1}{e}$

(72) Def 7.19: The Gaussian width of a subset $T \subset \mathbb{R}^n$ is $w(T) = \mathbb{E}[\sup_{x\in T}\langle Y, x\rangle]$ w/ $Y \sim N(0, \mathbb{1}_n)$

(73) Prop 7.21 (Properties of Gaussian width): Recall that $\text{diam}(T) = \sup_{x,y\in T}\|x - y\|_2$, $T \subset \mathbb{R}^n$

(a) $w(T)$ finite $\Leftrightarrow$ T bounded

Pf: '$\Rightarrow$' $|\langle Y, x\rangle| \leq \|Y\|_2\|x\|_2$ $\forall x \in T$. If $w(T) < \infty \Rightarrow \|x\|_2 < \infty$ $\forall x \in T \Rightarrow T$ bdd $\overset{w(T)<\infty}{\Uparrow}$

'$\Leftarrow$' T bdd $\Rightarrow \|x\|_2 \leq C$ $\forall x \in T$ for some $C > 0$. Hence $\mathbb{E}[\langle Y, x\rangle] \leq \mathbb{E}[\|Y\|_2]C \leq \sqrt{n}C < \infty$

(b) $w(T) = w(UT)$ $\forall U \in O(n)$ orthogonal Pf: rotational $\overset{(34)}{\text{invariance}}$, $Y \sim N(0, \mathbb{1}_n) \Rightarrow UY \sim N(0, \mathbb{1}_n)$

(c) $w(T + S) = w(T) + w(S)$, $S, T \subset \mathbb{R}^n$ and $w(aT) = |a|w(T)$ $\forall a \in \mathbb{R}$

Pf: $w(T+S) = \mathbb{E}[\sup_{x\in T, y\in S}\langle Y, x + y\rangle] = \overbrace{\mathbb{E}[\sup_{x\in T}\langle Y, x\rangle]}^{= w(T)} + \overbrace{\mathbb{E}[\sup_{y\in S}\langle Y, y\rangle]}^{= w(S)}$

If $a \geq 0$, $a = |a|$, $\langle Y, ax\rangle = |a|\langle Y, x\rangle$,

If $a < 0$, $|a| = -a$, Y symmetric $\Rightarrow$ $|a|\langle Y, x\rangle = -a\langle Y, x\rangle$

$\underbrace{\langle -Y, ax\rangle \sim \langle Y, ax\rangle}_{\text{same distribution}}$

(d) $w(T) = \frac{1}{2}w(T - T) = \frac{1}{2}\mathbb{E}[\sup_{x,y\in T}\langle Y, x - y\rangle]$

Pf: $w(T) = \frac{1}{2}(w(T) + w(T)) = \frac{1}{2}(w(T) + w(-T)) = \frac{1}{2}w(T - T)$

(e) $$\frac{1}{\sqrt{2\pi}} \text{diam}(T) \leq w(T) \leq \frac{\sqrt{n}}{2} \text{diam}(T)$$

Pf: $\boxed{\text{lower bound}}$: Fix $x, y \in T$, then $x-y,\ y-x \in T-T$ $\max\{a, -a\} = |a|$

just take 1 pair, not supremum

$$w(T) \geq \mathbb{E}[\max\{\langle Y, x-y\rangle, \langle Y, y-x\rangle\}] = \tfrac{1}{2}\mathbb{E}[|\langle Y, x-y\rangle|] = \frac{1}{2}\sqrt{\frac{2}{\pi}}\|x-y\|_2$$

note: $\mathbb{E}[\|Y\|_2] \leq (\mathbb{E}[\|Y\|_2^2])^{\frac{1}{2}}$

$\mathbb{E}[\|Y\|_2^2] = n$

$\Rightarrow \mathbb{E}[\|Y\|_2] \leq \sqrt{n}$

why? $\langle Y, x-y\rangle \sim N(0, \|x-y\|_2)$

$\Rightarrow \langle Y, \frac{x-y}{\|x-y\|_2}\rangle \sim N(0,1)$ and if $X \sim N(0,1)$ then

then take sup over all x, y

$$(✱)\ \mathbb{E}[|X|] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}|x|e^{-x^2/2}dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty}xe^{-x^2/2}dx = \frac{2}{\sqrt{2\pi}}[e^{-x^2/2}]_0^{\infty} = \sqrt{\frac{2}{\pi}}$$

$\boxed{\text{Upper bound}}$: $$w(T) = \frac{1}{2}\mathbb{E}[\sup_{x,y\in T}\{\langle Y, x-y\rangle\}] \leq \frac{1}{2}\mathbb{E}[\sup_{x,y\in T}\overbrace{\|Y\|_2}^{\mathbb{E}}\underbrace{\|x-y\|_2}_{\text{diam}}] \leq \frac{1}{2}\sqrt{n}\,\text{diam}(T)$$

(74) Gaussian width calculations

Homework

(a) $$w(S^{n-1}) = \mathbb{E}[\sup_{x\in S^{n-1}}\langle Y, x\rangle] \leq \mathbb{E}[\sup_{x\in S^{n-1}}\|Y\|_2\overbrace{\|x\|_2}^{=1}] \leq \mathbb{E}[\|Y\|_2] = \sqrt{n} \neq c$$

(b) Cube $B_\infty^n = [-1,1]^n$ wrt L_∞ norm is $w(B_\infty^n) = n\sqrt{\frac{2}{\pi}}$

$\max\{|x_i|\}$

Pf: $w(B_\infty^n) = \mathbb{E}[\sup_{x\in B_\infty^n}\langle Y, x\rangle] = ?$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \cdot \begin{pmatrix} \text{sign}(Y_1) \\ \vdots \\ \text{sign}(Y_n) \end{pmatrix} = \overbrace{|Y_1| + |Y_2| + \dots + |Y_n|}^{=\|Y\|_1}$$

Note $\mathbb{E}[\langle Y, x\rangle] \leq \mathbb{E}[\|Y\|_1\|x\|_\infty] = \mathbb{E}[\|Y\|_1]$

Set $x = (\text{sign } Y_1, \dots, \text{sign } Y_n) \in B_\infty^n$, get lower bd $\mathbb{E}[\langle Y, x\rangle] = \mathbb{E}[\|Y\|_1]$

Hence equality $$w(B_\infty^n) = \mathbb{E}[\|Y\|_1] = \mathbb{E}[\sum_{i=1}^n |Y_i|] = n\mathbb{E}[|Y_1|] \overset{(✱)}{=} n\sqrt{\frac{2}{\pi}}$$

(c) $B_1^n = \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$ unit ball wrt ℓ_1 norm has $w(B_1^n) = \mathbb{E}[\|Y\|_\infty] = \mathbb{E}[\max_{1\leq i\leq n}|Y_i|]$

why? $w(B_1^n) = \mathbb{E}[\sup_{x\in B_1^n}\langle Y, x\rangle] = \mathbb{E}[\|Y\|_\infty] = \mathbb{E}[\max_i |Y_i|]$

cauchy schwartz: $\mathbb{E}[\langle Y, x\rangle] \leq \mathbb{E}[\|Y\|_\infty \overbrace{\|x\|_1}^{=1 \text{ on } B_1^n}]$

pick x to be zero everywhere but for a 1 in the max Y_i column, $x = (\mathbb{1}_i, \dots, \mathbb{1}_i)$ where $i = \{i : \max_{1\leq i\leq n}\{|Y_i|\}\}$

$\Rightarrow \|Y\|_\infty \leq \mathbb{E}[\langle Y, x\rangle]$ so equality

note, ⑦① $\Rightarrow c\sqrt{\log n} \leq w(B_1^n) \leq C\sqrt{\log n}$

✓

✓

✓

✓