Session – 3

26 / 05 / 25

Recall:

1) $X \sim sG(\sigma^2)$ if $\mathbb{E}e^{\lambda(X-\mathbb{E}X)} \leq e^{\lambda^2\sigma^2/2}$ $\forall\ \lambda \in \mathbb{R}$

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{\lambda X} \geq e^{\lambda a}) \leq e^{-\lambda a}\,\mathbb{E}e^{\lambda X}$$

[Cramer-Chernoff bd]

Hoeffding's Inequality: Let $(X_i)_i$ be indep rvs.

1) If $X_i \sim sG(\sigma_i^2)$, then for $S = \sum w_i X_i$,

$$\mathbb{P}(S - \mathbb{E}S \geq t) \leq \exp\left(-\frac{t^2}{2\sum w_i^2\sigma_i^2}\right) \quad \forall\ t \geq 0$$

2) If $\mathrm{Ran}(X_i) \subseteq [a_i, b_i]$, then for $S = \sum X_i$

$$\mathbb{P}(S - \mathbb{E}S \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

2) $X \sim s\mathcal{E}(\nu, \alpha)$ if $\mathbb{E}e^{\lambda(X-\mathbb{E}X)} \leq e^{\frac{\lambda^2\nu}{2}}$ for $|\lambda| \leq \frac{1}{\alpha}$.

Bernstein's Inequality: Let $(X_i)_i$ be indep rvs.

1) If $X_i \sim s\mathcal{E}(\nu_i, \alpha_i)$, then for $S = \sum w_i X_i$,

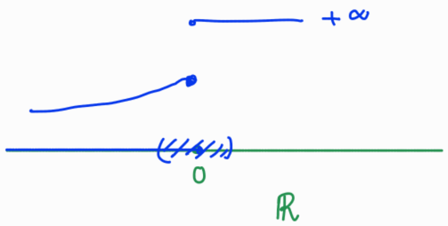
$$\mathbb{P}(S - \mathbb{E}S \geq t) \leq \begin{cases} \exp\left(-\dfrac{t^2}{2\sum w_i^2\nu_i}\right) & 0 \leq t \leq \dfrac{\sum w_i^2\nu_i}{\max |w_i|\alpha_i} \\[2ex] \exp\left(-\dfrac{t}{2\max |w_i|\alpha_i}\right) & t \geq \dfrac{\sum w_i^2\nu_i}{\max |w_i|\alpha_i} \end{cases}$$

2) If $|X_i| \leq c$, then for $S = \sum X_i$,

$$\mathbb{P}(S - \mathbb{E}S \geq t) \leq \begin{cases} \exp\left(-\dfrac{t^2}{4\sum \nu_i}\right) & 0 \leq t \leq \dfrac{\sum \nu_i}{c} \\[2ex] \exp\left(-\dfrac{t}{4c}\right) & t \geq \dfrac{\sum \nu_i}{c} \end{cases}$$

MGFs and CGFs: i) $X \geq 0$, then $\forall\, \lambda \leq 0$, $\mathbb{E} e^{\lambda X} \leq 1$

$X \leq 0$, then $\forall\, \lambda \geq 0$, $\mathbb{E} e^{\lambda X} \leq 1$



Say $X \geq 0$, If $\mathbb{E} e^{sX} < \infty$ for any $s > 0$

Q: For what rvs X is $M_X(s) < \infty$ for $s \in (-\delta, \delta)$.

eg. $X \sim$ Cauchy, then $\mathbb{E} e^{sX} = +\infty$ $\forall\, s \neq 0$.

☆ **Taylor Series Exp:** Say $M_X(s) < \infty$ for $s \in (-\delta, \delta)$.

$$\cdot\ M_X(s) = \mathbb{E} e^{sX} = \mathbb{E}\left(1 + \frac{sX}{1!} + \frac{s^2}{2!} X^2 + \cdots\right)$$

$$\overset{?\checkmark}{=} 1 + \frac{s}{1!}\mathbb{E}X + \frac{s^2}{2!}\mathbb{E}X^2 + \cdots \longrightarrow (1)$$

i) $M_X^{(n)}(0) = \mathbb{E}X^n \longrightarrow$ Moment gen. fn.

Proof of Analyticity: $e^{s|X|} \leq e^{-sX} + e^{sX}$ $\quad s \geq 0$

$$\cdot\ \mathbb{E} e^{s|X|} \leq \mathbb{E} e^{-sX} + \mathbb{E} e^{sX} < \infty \quad \forall\, s \in (-\delta, \delta)$$

$$\| $$

$$\mathbb{E}\left(1 + s|X| + \frac{s^2}{2}|X|^2 + \cdots\right) \qquad [\text{Fubini - Tonelli}]$$

$$= 1 + s\,\mathbb{E}|X| + \frac{s^2}{2}\mathbb{E}|X|^2 + \cdots < \infty$$

$\implies$ Series in (1) is abs. conv.

Exc: Show that $M_X(s)$ is analytic on the open set $\{s \mid M_X(s) < \infty\}^\circ$

Q: Why would one expect $\psi_X(s) = \log \mathbb{E} e^{sX} \leq \frac{s^2 \nu}{2}$ Variance Repr.

Say $M_X(s) < \infty \quad \forall \; s \in (-\delta, \delta)$. Then,

$$\psi_X(s) = \log\left(1 + \frac{\mathbb{E}X}{1!}s + \frac{\mathbb{E}X^2}{2!}s^2 + O(s^3)\right)$$

$$\psi_{X-\mathbb{E}X}(s) = \log\left(1 + \frac{\operatorname{Var} X}{2!}s^2 + O(s^3)\right)$$

$$= \frac{\operatorname{Var} X}{2}s^2 + O(s^3) \qquad \forall \; s \in (-\delta, \delta)$$

$$\leq (\operatorname{Var} X)s^2 \qquad \forall \; |s| \leq s_0 .$$

---

3. <u>Sub Gaussian Norm</u>: $X \sim sG(\sigma^2)$ if $\mathbb{E}e^{tX} \leq e^{\frac{t^2\sigma^2}{2}}$

<u>Thm</u>: Let X be a r.v., $\mathbb{E}X = 0$.

Has an analogue for $s\mathcal{E}(\nu, \alpha)$

1) $\mathbb{P}(|X| \geq t) \leq 2e^{-\frac{t^2}{K_1^2}} \quad \forall \; t \geq 0$ [Tails]

2) $(\mathbb{E}|X|^p)^{1/p} = \|X\|_p \leq K_2\sqrt{p} \quad \forall \; p \geq 1$ [Moments]

3) $\mathbb{E}e^{X^2/K_3^2} \leq 2$ [MGF of X^2]

4) $\mathbb{E}e^{\lambda X} \leq e^{\frac{\lambda^2 K_4^2}{2}}$ [MGF of X]

[Vershynin – HDP]

(1) $\Rightarrow$ (2):

$$\mathbb{E}|X|^p = \int_0^\infty \mathbb{P}(|X|^p \geq t)\,dt \qquad t = u^p$$

$$= \int_0^\infty \mathbb{P}(|X| \geq u)\, p u^{p-1}\,du$$

$$\leq 2\int_0^\infty e^{-\frac{t^2}{K_1^2}} p u^{p-1}\,du \asymp p\,\Gamma\left(\frac{p}{2}\right)$$

$$\asymp \left(\frac{p/2}{e}\right)^{p/2}$$

$\therefore \; (\mathbb{E}|X|^p)^{1/p} \leq K_2\sqrt{p}$

$$\boxed{\begin{array}{c} f \asymp g \\ cg \leq f \leq Cg \text{ for large } n \end{array}}$$

Def$\underline{n}$: $X \sim sG(\sigma^2)$. Then

$$\|X\|_{\psi_2} = \inf\left\{K > 0 \;\middle|\; \mathbb{E}e^{X^2/K^2} \leq 2\right\}$$

— This is a norm. $\quad \|X+Y\|_{\psi_2} \leq \|X\|_{\psi_2} + \|Y\|_{\psi_2}$.

Similarly, we have subexp norm (Vershynin Ch 2).

Thm: Let X be a r.v., $\mathbb{E}\underline{X} = 0$.

1) $\mathbb{P}(|X| \geq t) \leq 2e^{-\frac{c_1 t^2}{\|X\|_{\psi_2}}} \quad \forall\, t \geq 0$ [Tails]

2) $\left(\mathbb{E}|X|^p\right)^{1/p} = \|X\|_p \leq c_2 \|X\|_{\psi_2} \sqrt{p} \quad \forall\, p \geq 1$ [Moments]

→ 3) $\mathbb{E}e^{X^2/\|X\|^2_{\psi_2}} \leq 2$ [MGF of X^2]

4) $\mathbb{E}e^{\lambda X} \leq e^{\frac{c_3 \lambda^2 \|X\|^2_{\psi_2}}{2}}$ [MGF of X]

HIGH DIMENSIONAL GEOMETRY

• $C^d = \left[-\frac{1}{2}, \frac{1}{2}\right]^d \qquad S^{d-1} \subseteq \mathbb{R}^d \qquad B^d \subseteq \mathbb{R}^d$

$B^d(R)$

Volume: $1 \times 1 \times \cdots \times 1 = 1$ — $\qquad \mathrm{Vol}(B^d(R)) = \underline{\underline{V_d}}\, R^d$

$\mathrm{SA}(B^d(R)) = \underline{\underline{s_d}}\, R^{d-1}$

Calculating s_d: $\quad f(\underset{\sim}{x}) = e^{-(x_1^2 + \cdots + x_d^2)} \qquad \int_{\mathbb{R}^d} f(\underset{\sim}{x})\, d\underset{\sim}{x}$

$$\int_{\mathbb{R}^d} f = \left(\int_{\mathbb{R}} e^{-x_1^2} dx_1\right)\left(\int_{\mathbb{R}} e^{-x_2^2} dx_2\right) \cdots \left(\int_{\mathbb{R}} e^{-x_d^2} dx_d\right)$$

$$= \pi^{\frac{d}{2}}$$

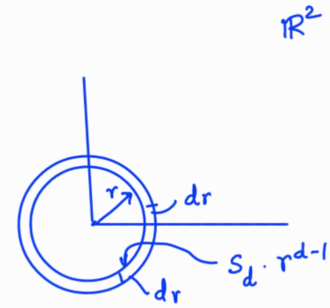
$$\int_{\mathbb{R}^d} f = \int_0^\infty e^{-r^2}\left(s_d \cdot r^{d-1}\,dr\right)$$

$$= s_d \int_0^\infty e^{-r^2} r^{d-1}\,dr$$

$$(r^2 = u) \qquad 2r\,dr = du$$

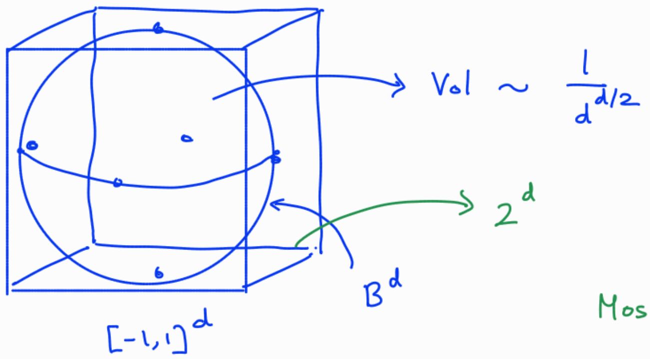
$$= \frac{s_d}{2}\int_0^\infty e^{-u}\, u^{\frac{d}{2}-1}\,du$$

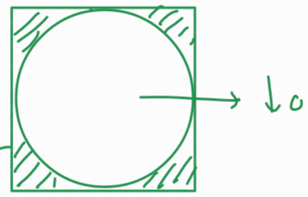
$$= \frac{s_d}{2}\cdot\Gamma\left(\frac{d}{2}\right) = \pi^{\frac{d}{2}} \implies \boxed{s_d = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \qquad v_d = \frac{\pi^{d/2}}{\frac{d}{2}\Gamma\left(\frac{d}{2}\right)}}$$



$$\mathrm{Vol}\left(B^d(R)\right) = v_d R^d = \frac{\pi^{d/2}}{\frac{d}{2}\left(\frac{d/2}{e}\right)^{d/2}}\cdot\frac{1}{\sqrt{\pi d}}\cdot R^d$$

$$= \frac{1}{\sqrt{\pi d}\cdot\frac{d}{2}}\cdot\left(\frac{2\pi e R^2}{d}\right)^{d/2} \xrightarrow{\text{as } d\to\infty} 0$$

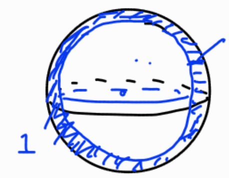




Most of the volume is in these corners

① Most of the volume of a hypercube is in the corners.

② $$\frac{\mathrm{Vol}\left(S(\varepsilon)\right)}{\mathrm{Vol}\left(B^d\right)} = \frac{1-(1-\varepsilon)^d}{1} \longrightarrow 1$$



$R_{in} = (1-\varepsilon)R$

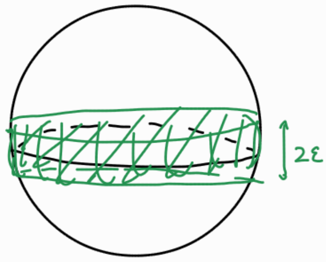
For a fixed ε, most of the vol of B^d is conc. close to the surface

Let $\varepsilon = \frac{k}{d}$, then $\frac{\mathrm{Vol}\left(S(\varepsilon)\right)}{\mathrm{Vol}\left(B^d\right)} = 1-\left(1-\frac{k}{d}\right)^d \approx 1-e^{-k}$ for large d

$\therefore$ To get $\frac{1}{2}$ the vol., it is enough to take

$$\varepsilon = \left(\frac{\log 2}{d}\right). \qquad \square$$

③ Most of volume of B^d is conc. near the equator.



$$\frac{\mathrm{Vol}\left(\{0 \le x_1^2 + \dots + x_d^2 \le 1,\ |x_1| \le \varepsilon\}\right)}{\mathrm{Vol}\left(\{0 \le x_1^2 + \dots + x_d^2 \le 1\}\right)} \longrightarrow 1$$

3. Aside on MGF:

① X, Y are rvs s.t. $\underline{M_X(s) = M_Y(s)}$ on $(-\delta, \delta)$ for any $\delta > 0$.
Then $X \stackrel{d}{=} Y$.

$\Rightarrow$ All moments agree

$\not\Downarrow$

$\to$ Moments cannot grow too fast.

② $\exists\, X, Y$ s.t. $\mathbb{E}X^p = \mathbb{E}Y^p < \infty\ \forall p \in \mathbb{N}$, but $X \stackrel{d}{\neq} Y$.

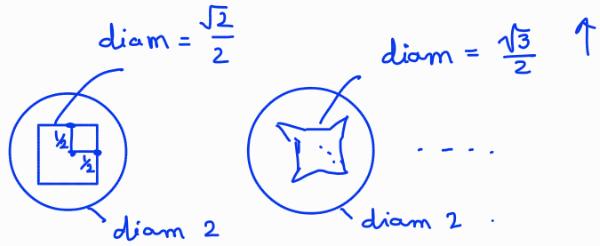
$X \sim$ log-normal , $Y \sim$ Some weird distr$^{\underline{n}}$ $(\sin(\cdots))$

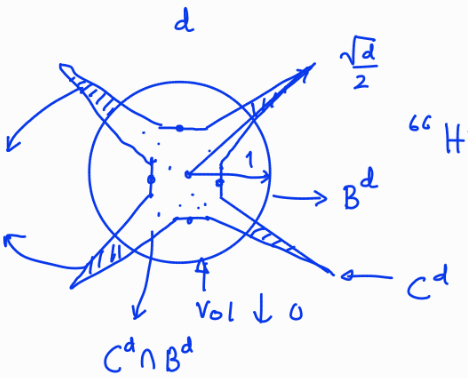
$\longrightarrow$ Moments grow too fast.

③ Riesz Cond$^{\underline{n}}$ etc. $\begin{cases} \lim\limits_{p \to \infty} \frac{1}{p}\left(\mathbb{E}X^{2p}\right)^{\frac{1}{2p}} < \infty \\ \dots \end{cases}$

---

④ Unit sphere vs unit cube.





These 'spikes' or 'corners' contain most of the volume.

"High dim cubes are spiky"

• X_i iid Unif $[-1, 1]$

$$\mathrm{Vol}(C^d \cap B^d) = \mathbb{P}\Big(\sum_{i=1}^{d} \underset{[0,1]}{X_i^2} \leqslant 1 \Big) \qquad \mathbb{E} X_1^2 = \int_0^1 x^2\,dx = \frac{1}{3}$$

$$\mathbb{E} X_1^4 = \int_0^1 x^4\,dx = \frac{1}{5}$$

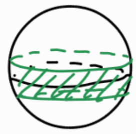
$$= \mathbb{P}\Big(\sum_{i=1}^{d} \big(X_i^2 - \tfrac{1}{3}\big) \leqslant -\big(\tfrac{d}{3} - 1\big) \Big)$$

$$\leqslant \exp\Big(- \frac{2\left(\frac{d}{3}-1\right)^2}{d} \Big) \leqslant \exp\Big(-\frac{2d}{9}\Big)$$

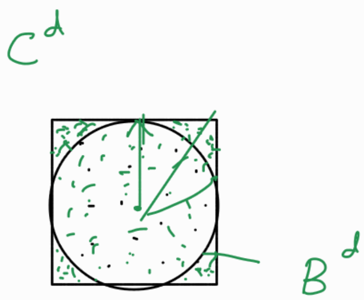
- [H.W.] Give a prob. proof for ③.

$$\mathbb{P}\Big(\frac{X_1}{\sqrt{\sum X_i^2}} \leqslant \varepsilon \Big)$$

⑤ Q: How do we obtain a uniform sample from S^{d-1}?



$\emptyset$ [H.W.] If $X_1, \dots, X_d \overset{iid}{\sim} N(0,1)$ and $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$, then for $A \in O(d)$, $A\underset{\sim}{X} \overset{d}{=} \underset{\sim}{X}$.



X_i iid $N(0,1)$

- Multivariate Normal $\quad (X_1, X_2, \dots, X_d) \sim N(0, I)$

- $$\frac{(X_1, \dots, X_d)}{\sqrt{X_1^2 + \dots + X_d^2}} \longrightarrow \text{Uniform on } S^{d-1}$$

⑥ Let $\underset{\sim}{X} = (X_1, X_2, \dots, X_d)$ is a d-dim^l random vector s.t. X_i indep $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$.

- $$\mathbb{E}\|\underset{\sim}{X}\|_2^2 = \mathbb{E}(X_1^2 + X_2^2 + \dots + X_d^2) = n.$$

- $$\|\underset{\sim}{X}\|_2 \approx \sqrt{n}$$

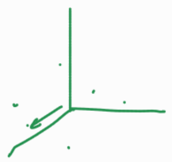
$$\mathbb{P}\Big(\Big| \frac{\|\underset{\sim}{X}\|_2}{\sqrt{n}} - 1 \Big| > \delta \Big) \leqslant \exp\Big(-\frac{n}{8} \max(\delta, \delta^2) \Big) \qquad \text{[H.W.]}$$

[Use $\quad |z-1| > t \Rightarrow |z^2 - 1| > \max(t, t^2)$.]

⊘ Reference: Bandeira, S – Math of Data Science (Ch 2)

⑦ $X = (X_1, \dots, X_d)$, $Y = (Y_1, \dots, Y_d)$ where X_i, Y_j iid $\underset{\text{Rad}(\frac{1}{2})}{\overset{\|}{\text{Ber}_{\pm}(\frac{1}{2})}}$

$$\mathbb{P}\left(|\cos \theta_{x,y}| > \sqrt{\frac{2 \log d}{d}} \right) \leq \frac{2}{d} \leftarrow$$



$$|\cos(\theta_{x,y})| = \frac{|\langle x, y \rangle|}{\underset{\sqrt{d}}{\|X\|} \; \underset{\sqrt{d}}{\|Y\|}} = \frac{|\langle x, y \rangle|}{d}$$

$X_i Y_i \overset{d}{=} \text{Ber}_{\pm}\left(\frac{1}{2}\right)$

$$\mathbb{P}\left(|\cos \theta_{x,y}| \geq \frac{t}{d}\right) = \mathbb{P}\left(|\langle x, y \rangle| \geq t \right) = \mathbb{P}\left(\left| \sum_i X_i Y_i \right| \geq t \right)$$

$$\leq 2 \exp\left(- \frac{\cancel{2} t^2}{2 \cancel{4} d} \right)$$

Take $t = \sqrt{2d \log d}$.

Using this method, we can generate $\sim \overset{m}{\geq} \exp(cd)$ many points $(x_i)_{i=1}^m$ on $\boxed{\{-1, +1\}^d}$ which are approx $\perp^r$.

$\underline{\overline{2^d}}$ χ k ⓓ

$$|\cos \theta_{x_i, x_j}| \leq \frac{2}{\tilde{c}}.$$ $\tilde{c}$ is given by $\frac{1}{c}$

Maximal such set S

|

Can we talk abt dimensions / complexity of S.

⊛ Concentration in High Dimensions

↙ Curse ↘ Blessing

Henceforth:

⟶ Conc. phenomena for $f(X_1, \dots, X_d)$ $X_1, \dots, X_d \sim sG$.

$$\rightarrow \mathbb{P}(|f - \mathbb{E}f| \geq t) \leq \exp(-).$$

↙ Smooth f ↘ f varies 'slowly'

f = chromatic no. of Graph