LECTURE 11

- Last class, we introduced the class of <u>subgaussian distributions</u>

$X \sim$ Subgaussian if:

(1) $\exists K_1$: $\mathbb{P}\{|X| \geq t\} \leq 2\exp(-t^2/K_1^2) \quad \forall t \geq 0$

$\Updownarrow$

(2) $\exists K_2$: $\|X\|_{L^p} \leq K_2\sqrt{p} \quad \forall p \geq 1$

$\Updownarrow$

(3) $\exists K_3$: $\mathbb{E}\exp(X^2/K_3^2) \leq 2$

$\Updownarrow$ if $\mathbb{E}X = 0$

(4) $\exists K_4$: $\mathbb{E}\exp(\lambda X) \leq \exp(K_4^2\lambda^2) \quad \forall \lambda \in \mathbb{R}$.

All K_i's are equivalent: $\boxed{K_i \leq 10 K_j \quad \forall i,j}$

The smallest K_3 is called <u>subgaussian norm</u> $\|X\|_{\psi_2}$ (Orlicz norm for $\psi_2(x) = e^{x^2} - 1$)

- <u>Ex</u> : $N(0,1)$, $Ber(p)$, $Unif[-1,1]$, $\forall$ bounded on $[-1,1]$ are all subgaussian r.v.'s, and $\|X\|_{\psi_2} \leq C$ **(Check!)**

- Fact: $\boxed{X_i \text{ subgaussian} \Rightarrow \sum X_i \text{ subgaussian}}$ [L_{ψ_2} is a <u>linear</u> normed space]

(may be dependent ☺)

- $\|X\|_{\psi_2}^2$ is a "variance proxy"

$\mathrm{Var}(\sum X_i) = \sum \mathrm{Var}(X_i)$ for indep. X_i. Similarly:

<u>Prop</u> If X_i are independent, mean zero, subgaussian, then

$$\Big\| \sum_{i=1}^N X_i \Big\|_{\psi_2}^2 \le C \sum_{i=1}^N \|X_i\|_{\psi_2}^2$$

$$\mathbb{E}\exp\Big(\lambda \sum_i X_i\Big) \underset{\text{indep}}{=} \mathbb{E}\prod_i \exp(\lambda X_i) = \prod_i \underbrace{\mathbb{E}\exp(\lambda X_i)}_{\le\, \exp(CK_i^2\lambda^2)} \le$$

where $K_i = \|X_i\|_{\psi_2}$ due to subgauss. property (4)

$$\le \exp\Big(C\lambda^2 \underbrace{\sum_i K_i^2}_{=\,K^2}\Big)$$

$\Rightarrow$ $\sum_i X_i$ is subgaussian and $\|\sum_i X_i\|_{\psi_2} \le CK$, due to subgauss. property (4).

<u>THM</u> (<u>Subgaussian Hoeffding ineq.</u>) If X_i are independent, mean zero, subgaussian, then

$$\mathbb{P}\Big\{ \Big|\sum_{i=1}^N X_i\Big| \ge t \Big\} \le 2\exp\Big(-\frac{ct^2}{\sigma^2}\Big)$$

where $\sigma^2 = \sum_{i=1}^N \|X_i\|_{\psi_2}^2$

variance proxy of the sum

Apply Prop and then the subgauss. tail (property 1).

Generalizes Hoeffding inequality from <u>bounded</u> to <u>subgaussian</u> r.v.'s

Subexponential distributions.

- $X \sim$ Subgaussian $\not\Rightarrow$ $X^2 \sim$ Subgaussian:

  e.g. $g \sim N(0,1)$. $\mathbb{P}\{g^2 > t\} = \mathbb{P}\{g > \sqrt{t}\} \sim \exp\left(-\frac{\sqrt{t}^2}{2}\right) = \boxed{\exp(-t/2)}$

  exponential tail heavier than $\exp(-ct^2)$

- Fix: define subexponential distr. similarly:

Prop $\forall$ r.v. X, TFAE:

(1) $\exists K_1$: $\mathbb{P}\{|X| \geq t\} \leq 2\exp(-t/K_1)$ $\forall t \geq 0$

$\Updownarrow$

(2) $\exists K_2$: $\|X\|_{L^p} \leq K_2 p$ $\forall p \geq 1$

$\Updownarrow$

(3) $\exists K_3$: $\mathbb{E}\exp(|X|/K_3) \leq 2$

$\Updownarrow$ if $\mathbb{E}X = 0$

(4) $\exists K_4$: $\mathbb{E}\exp(\lambda X) \leq \exp(K_4^2 \lambda^2)$ $\forall |\lambda| \leq \frac{1}{K_4}$

Such X is called subexponential. The smallest K_3 is called the subexponential norm $\|X\|_{\psi_1}$

Orlicz norm for $\psi_1(x) = e^x - 1$

Proof – DIY (book)

Examples

(a) $\forall$ subgaussian, e.g. $g \sim N(0,1)$ (by property 2, since $\sqrt{p} \le p$)

(b) $\forall$ subgaussian squared, e.g. g^2.

$\forall$ r.v. X: $\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$ trivially (property 3)

(c) Exponential $X \sim \text{Exp}(1)$ $X \ge 0$,

$$\mathbb{P}\{X \ge t\} = e^{-t} \quad \forall t \ge 0$$

SKIP

$$\text{pdf} = \frac{d}{dt}(\text{cdf}) = \frac{d}{dt}(1 - e^{-t}) = e^{-t}, \quad t \ge 0$$

$$\Rightarrow \mathbb{E}\exp(\lambda X) = \int_0^\infty e^{\lambda x} \cdot e^{-x}\, dx < \infty \text{ only if } \underline{\lambda \le 1}$$

$\Rightarrow$ restriction on λ in (4) is essential.

(d) Poisson: not subgaussian **(HW)** but subexponential (tail t^{-t})

Q Does Hoeffding hold for subexponential X_i?

- NO! Even for $N=1$: $\mathbb{P}\{|X| \ge t\} \le \underbrace{\exp\left(-\frac{ct}{K}\right)}$ where $K = \|X\|_{\psi_1}$

too heavy! $\Rightarrow$ Hoeffding's $\exp(-ct^2)$

- Surprisingly, $N=1$ is "the only obstruction":

BERNSTEIN'S INEQUALITY If X_i are indep, mean 0, subexponential, then

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

where $\sigma^2 = \sum_{i=1}^N \|X_i\|_{\psi_1}^2$, $K = \max_i \|X_i\|_{\psi_1}$

Proof: the same MGF method as in Hoeffding, Chernoff: $S = \sum X_i$

$$\mathbb{P}\{S \ge t\} = \mathbb{P}\{e^{\lambda S} \ge e^{\lambda t}\} \le e^{-\lambda t}\,\mathbb{E}e^{\lambda S} = e^{-\lambda t}\boxed{\mathbb{E}e^{\lambda X_i}}$$

$\leftarrow$ now use Property 4 p.3

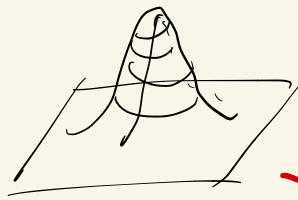
Restriction on λ leads to

THE THIN SHELL PHENOMENON

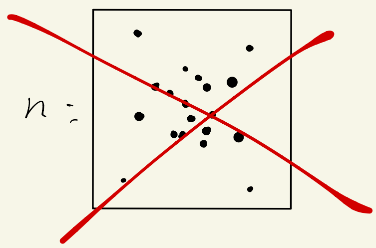
- Gaussian random vector $g \sim N(0, I_n)$: $g = (g_1, \dots, g_n)$, $g_i \overset{\text{iid}}{\sim} N(0,1)$

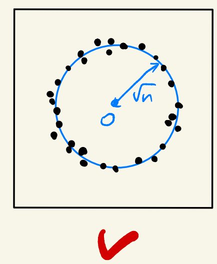
- pdf: $f(x) = f(x_1) \cdots f(x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2}$

Rotation invariant:



- Sample from $N(0, I_n)$ for large n:





THM (Thin shell)

$$\boxed{\text{For } g \sim N(0, I_n): \quad \mathbb{P}\left\{0.99\sqrt{n} \le \|g\|_2 \le 1.01\sqrt{n}\right\} \ge 1 - 2e^{-cn}}$$

exponentially close to 1 !

Proof:

$$\|g\|_2^2 - n = \sum_{i=1}^{n} \underbrace{(g_i^2 - 1)}$$

iid, mean zero, subexponential:

$$\|g_i^2 - 1\|_{\psi_1} \le \|g_i^2\|_{\psi_1} + 1 \qquad (\Delta \text{ inequality})$$

$$= \|g_i\|_{\psi_2}^2 + 1 \le C \qquad (\text{abs. constant})$$

Bernstein's inequality $\Rightarrow$

$$\mathbb{P}\left\{\left|\|g\|_2^2 - n\right| > \underbrace{0.01n}_{t}\right\} \le 2\exp\left(-c \cdot \min\left(\frac{n^2}{\sigma^2}, \frac{n}{K}\right)\right) \le$$

where $\sigma^2 = \sum_{i=1}^{n} \|g_i^2 - 1\|_{\psi_1}^2 \le Cn$; $\quad K = \max_i \|g_i^2 - 1\|_{\psi_1} \le C$.

$$\le 2\exp(-c'n).$$

$\Rightarrow$ with prob. $\ge 1 - 2\exp(-c'n)$,

$$0.99n \le \|g\|_2^2 \le 1.01n.$$

QED.