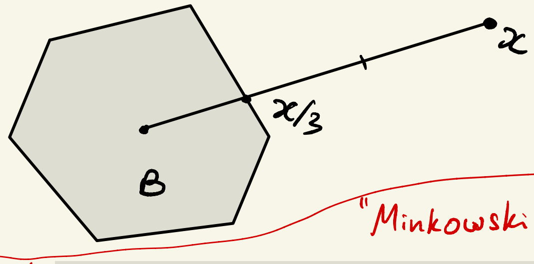
LECTURE 10

- Since $\|X\|_{L^p} = (\mathbb{E}|X|^p)^{1/p}$ is a norm, maybe $\|X\| = \ln(\mathbb{E}e^X)$ is a norm?

  NO: homogeneity fails (HW)

- A fix: Instead of defining a norm $\|x\|$, define the unit ball B first;
  (must be a convex, origin-symmetric set ← HW);
  - B determines which vectors x have norm $\|x\| \le 1$
  - then extend $\|\cdot\|$ to all vectors by homogeneity:



"Minkowski functional" $\|x\| := \inf\{K > 0 : \frac{x}{K} \in B\}$ (*)

SKIP

THM Let B be a convex, origin-symmetric set in a vector space V. ← HW
If Minkowski functional is finite $\forall x \in V$,
then it defines a norm on V. Moreover, $B = \{x \in V : \|x\| \le 1\}$ is a closed ball.
(?)

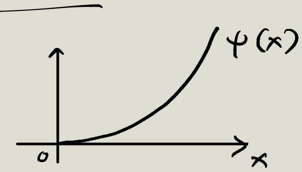
$B := \{\text{r.v's } X : \mathbb{E}e^{|X|} \le 2\}$ is convex, origin symmetric $\Rightarrow$

$$\|X\| = \inf\{K : \tfrac{X}{K} \in B\} = \inf\{K : \mathbb{E}\exp(|X|/K) \le 2\} \text{ is a norm! } ☺$$

More generally:

Def A function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ is called an Orlicz function if
ψ is convex, increasing, and

$$\psi(x) \to \begin{cases} 0 & \text{as } x \to 0 \\ \infty & \text{as } x \to \infty \end{cases}$$



Ex: x^p, $e^x - 1$, $e^{x^2} - 1$.

SKIP

Prop If ψ is an Orlicz function, the set $\{\text{r. variables } X : \mathbb{E}\psi(|X|) \le 1\}$ is convex and origin-symmetric. (HW)

Thus:

Def If ψ is an Orlicz function, the "Orlicz norm"

$$\|X\|_\psi := \inf\{K > 0 : \mathbb{E}\psi(|X|/K) \le 1\}$$

defines a norm on the "Orlicz space"

$$L_\psi := \{\text{r.v's } X : \|X\|_\psi < \infty\}$$

- Ex: (a) $\psi(x) = x^p \Rightarrow L^p$ (b) $\psi(x) = e^x - 1 \Rightarrow L^p \supset L_\psi \supset L^\infty$ $\forall p \ge 1$
  $X \sim N(0,1)$ belongs to L_ψ.

SUBGAUSSIAN DISTRIBUTIONS.

- Recall Hoeffding's inequality : if $X_i \sim \text{Sym Ber}$ iid, then

$$\mathbb{P}\left\{ \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \right| \geq t \right\} \leq 2\exp\left(-\frac{t^2}{2}\right) \qquad \forall t \geq 0. \qquad (*)$$

gaussian tail

if $X_i \sim N(0,1)$, $X_i \sim$ Symmetric Bernoulli, or $X_i \sim \text{Unif}[-\frac{1}{2}, \frac{1}{2}]$.

- What is the biggest class of distributions of X_i that obey H.I.?
- For $N=1$, $(*)$ becomes

$$\boxed{\mathbb{P}\{ |X_i| \geq t \} \leq 2\exp(-t^2/2)} \qquad (**)$$

$\Rightarrow$ X_i's must satisfy that.

- We call such X_i $\boxed{\text{subgaussian}}$ r.v's

& we will show that H.I. holds for such X_i. (PLAN)

$(**)$ = "X_i is as good as $N(0,1)$" in many ways.

In what ways?

SKIP

Space of (sub)gaussian r.v's? (Habitat for gaussian & better)

what do we know about gaussian?

Let $g \sim N(0,1)$. Then:

① <u>Tails</u>: $\mathbb{P}\{|g| \geq t\} \leq 2\exp(-t^2/2) \quad \forall t \geq 0$.

② <u>Moments</u>: if p is even,

$$\mathbb{E} g^p \overset{\text{HW1}}{=} (p-1)!! = 1\cdot 3\cdot 5 \cdots (p-3)(p-1) \leq \underbrace{p\cdot p\cdot p \cdots p \cdot p}_{p/2} \leq p^{p/2}$$

(and if p=o, $\mathbb{E} g^p = 0$).

$$\Rightarrow \quad \|g\|_{L^p} = (\mathbb{E}|g|^p)^{1/p} \leq \sqrt{p} \qquad \forall p \geq 1.$$

(MUST grow to ∞)

③ <u>MGF</u>: $$\mathbb{E} e^{\lambda g} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\lambda x} e^{x^2/2} \, dx = e^{\lambda^2/2} \quad \forall \lambda \in \mathbb{R}$$

(DIY)

④ <u>MGF of g^2</u>: $$\mathbb{E} e^{g^2/4} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{e^{x^2/4} e^{-x^2/2}}_{= e^{-x^2/4}} \, dx \leq 2.$$

$\Rightarrow$ for Orlicz function

$\psi_2(x) := e^{x^2} - 1, \quad \mathbb{E}\psi_2(|g|/2) \leq 1 \quad \Rightarrow \quad \|g\|_{\psi_2} \leq 2.$

(1) – (4) are equivalent for $\forall$ distribution!

Lemma (Subgaussian distributions)

$\forall$ random variable X, the following are equivalent:

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

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

(3) (ψ_2): $\exists K_3$: $\mathbb{E}\exp(X^2/K_3^2) \leq 2$

Moreover, if $\mathbb{E}X = 0$ then (1)–(3) are equivalent to:

(4) (MGF): $\exists K_5$: $\mathbb{E}\exp(\lambda X) \leq \exp(K_4^2 \lambda^2)$ $\forall \lambda \in \mathbb{R}$.

A r.v. X that satisfies one (and thus all) of properties (1)–(4) is called subgaussian.

The proof of (1) $\Rightarrow$ (2) is based on

Lem (Integrated Tail Formula) $\forall$ nonnegative r.v. X,

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

Γ $\forall x \in \mathbb{R}$,

$$x = \int_0^x dt = \int_0^\infty \underbrace{\mathbb{1}_{\{t<x\}}}_{= \begin{cases} 1 & \text{if } t<x \\ 0 & \text{if } t \geq x \end{cases}} dt$$

let $x = X$, take expectation $\Rightarrow$

$$\mathbb{E}X = \int_0^\infty \underbrace{\mathbb{E}\mathbb{1}_{\{t<X\}}}_{= \mathbb{P}\{t<X\}.} dt \qquad \text{(Fubini)}$$

$\lrcorner$

Proof of Subgaussian Lemma

<u>(1) ⇒ (2):</u> WLOG $K_1 = 1$

$$\mathbb{E}|X|^p = \int_0^\infty \mathbb{P}\{|X|^p > t\}\,dt \qquad \text{(Integral Lemma)}$$

change var. $t = s^p$ $\quad dt = ps^{p-1}ds$

$$= \int_0^\infty \underbrace{\mathbb{P}\{|X| > s\}}_{\leq\, 2e^{-s^2}} p s^{p-1}\, ds \qquad \text{by Property (1).}$$

$$= p\left(\tfrac{p}{2} - 1\right)! \qquad \text{(by parts -- DIY)}$$

$$\leq p \cdot p^{p/2 - 1} = p^{p/2}$$

$$\Rightarrow \|X\|_{L^p} \leq \sqrt{p}.$$

<u>(2) ⇒ (3):</u> WLOG $K_2 = 1$

$$\mathbb{E}\exp(X^2/100) = \mathbb{E}\left[\sum_{p=0}^\infty \frac{(X^2/100)^p}{p!}\right] \qquad \text{(Taylor series)}$$

$$= \sum_{p=0}^\infty \frac{\mathbb{E}X^{2p}}{100^p\, p!} \leq \sum_{p=0}^\infty \frac{(2p)^p}{100^p\, p!}$$

$\mathbb{E}X^{2p} \leq (\sqrt{2p})^{2p}$ $\qquad$ $p! \geq (p/e)^p$ (Stirling)

$$\leq \sum_{p=0}^\infty \Big(\underbrace{\frac{2e}{100}}_{\leq\, 1/10}\Big)^p \leq 2$$

(3) ⇒ (4) $\quad$ WLOG $K_3 = 1$. $\quad$ Use $e^x \leq x + e^{x^2}$ $\forall x \in \mathbb{R}$ ⇒

$$\mathbb{E}e^{\lambda X} \leq \underbrace{\mathbb{E}[\lambda X]}_{=\,0 \text{ by assumption}} + \mathbb{E}e^{\lambda^2 X^2}$$

• If $|\lambda| \leq 1$, $\quad \mathbb{E}e^{\lambda^2 X^2} \leq \left(\mathbb{E}e^{X^2}\right)^{\lambda^2}$

$$\left(\Leftrightarrow \left(\mathbb{E}e^{\lambda^2 X^2}\right)^{1/\lambda^2} \leq \mathbb{E}e^{X^2} \quad \Leftrightarrow \|e^{X^2}\|_{L^{\lambda^2}} \leq \|e^{X^2}\|_{L^1}\right)$$

true since $\lambda^2 \leq 1$

$$\Rightarrow \mathbb{E}e^{\lambda X} \leq \Big(\underbrace{\mathbb{E}e^{X^2}}_{\leq\, 2 \text{ by property (3)}}\Big)^{\lambda^2} \leq \exp(\lambda^2 \ln 2)$$

• If $|\lambda| > 1$ — DIY. (or see the Book).

(4) ⇒ (1) WLOG $K_4 = 1$.

$$\mathbb{P}\{X \geq t\} = \mathbb{P}\{e^{\lambda X} \geq e^{\lambda t}\} \quad \text{(multiply by } \lambda \geq 0 \text{ and exponentiate)}$$

$$\leq e^{-\lambda t} \underbrace{\mathbb{E} e^{\lambda X}}_{\leq e^{\lambda^2}} \quad \text{(Markov)} \quad \text{(property 4)}$$

$$= \exp(\lambda^2 - \lambda t). \quad \text{Minimize in } \lambda \Rightarrow \lambda = t/2 \Rightarrow$$

$$= \exp(-t^2/4). \quad \text{Repeat for } -X \Rightarrow \quad \text{QED}$$

QUANTITATIVE ASPECTS

- Subgaussian is a qualitative notion (yes/no). But:
- K_1, K_2, K_3, K_4 in Subgaussian Lemma are all equivalent up to absolute const. factors:

$$K_i \leq C_{ij} K_j \qquad \forall i, j = 1, 2, 3, 4 \qquad \text{(CHECK!)}$$

⇒ they all capture the ≈ same quantity.

- We call this quantity the subgaussian norm
- Why norm? the smallest K_3 in property ③ equals $\|X\|_{\psi_2}$, the Orlicz norm for $\psi_2(x) = e^{x^2} - 1$

& all other K_i's are equivalent ⇒

Subg. Lemma restated:

(1) (tails): $\mathbb{P}\{|X| \geq t\} \leq 2\exp(-ct^2/\|X\|_{\psi_2}^2) \quad \forall t \geq 0$

(2) (moments): $\|X\|_{L^p} \leq C\|X\|_{\psi_2} \sqrt{p} \quad \forall p \geq 1$

(3) (ψ_2): $\mathbb{E}\exp(cX^2/\|X\|_{\psi_2}^2) \leq 2$

Moreover, if $\mathbb{E}X = 0$ then

(4) (MGF): $\mathbb{E}\exp(\lambda X) \leq \exp(C\|X\|_{\psi_2}^2 \lambda^2) \quad \forall \lambda \in \mathbb{R}$.