Rotation Invariance

consider a finite number of independent centered subgaussian r.v. X_i

Then $\sum_i X_i$ is also a centered subgaussian r.v. Moreover, $\|\sum_i X_i\|_{\psi_2}^2 \leq C \sum_i \|X_i\|_{\psi_2}^2$

if not squared, this just follows from $\|\ \|_{\psi_2}$ is a norm.

where C is an absolute constant.

Pf: A direct consequence of rotation invariance Hoeffding-type inequality for sums of independent subgaussian r.v.

the sum is still sub-gaussian, so we can use the result about sub-gaussian to this sum

Propo 2.15

(Hoeffding-type ineq for subgaussian)

Let $X_1, \cdots, X_n$ be centered subgaussian r.v. that are inde, and let $K = \max_i \|X_i\|_{\psi_2}$ ← $\|X_i\|_{\psi_2}$ are bounded

Then for every $a = (a_1, \cdots, a_N) \in \mathbb{R}^N$ and every $t \geq 0$, we have

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

where $c > 0$ is an absolute constant.

Pf: $$\left\| \sum_i a_i X_i \right\|_{\psi_2}^2 \leq C \sum_i \|a_i X_i\|_{\psi_2}^2$$

$$= C \sum_i |a_i|^2 \|X_i\|_{\psi_2}^2$$

$$\leq C K^2 \|a\|_2^2$$

The result follows from Lemma 2.12 (tail decay)

same works for moments instead of tails:

Corollary 2.16 (Khintchine-type ineq)

Let X_i be as in Prop 2.15

Then for $p \geq 2$ and any sequence a

$$\left(E\left|\sum_i a_i X_i\right|^p\right)^{\frac{1}{p}} \leq C K \sqrt{p}\, \|a\|_2 \quad (*)$$

Furthermore, if the X_i's also have unit variance, for $p \geq 2$

$$\left(E\left|\sum a_i X_i\right|^p\right)^{\frac{1}{p}} \geq \|a\|_2 . \quad (+)$$

Pf: $(*)$ follows from Prop 2.15, just using moment method.

(+):

$$\left(E\left|\sum_i a_i X_i\right|^p\right)^{\frac{1}{p}} \underset{\substack{\uparrow \\ \text{Jensen with } f(x)=x^{\frac{p}{2}}}}{\geq} \left(E\left|\sum_i a_i X_i\right|^2\right)^{\frac{1}{2}}$$

$$\underset{\substack{\uparrow \\ \text{indep}}}{=} \left(E\, a_i^2 X_i^2 + \sum_{\substack{i,j \\ i\neq j}} a_i a_j \underbrace{E X_i}_{=0} \underbrace{E X_j}_{=0}\right)^{\frac{1}{2}}$$

$$= \left(\sum a_i^2 E X_i^2\right)^{\frac{1}{2}}$$

$$\underset{\substack{\uparrow \\ \text{unit variance}}}{=} \|a\|_2$$

why we require r.v. to be bounded but still analyze heavy tail?

because the constant bounds of r.v.s can be various.

(depend on the dim, that's the heavy tail)

(not we really hope to obtain)

$\downarrow$

2.5 Sub-exp r.v.

what if r.v. are not subgaussian?

We can formulate an equivalence also for slower tail decay, faster growing moments, ...

Homework: establish similar equivalence between tails, moments and superexponential moments also for tail decay rate: $e^{-ct^{\alpha}}$, $\alpha \geq 1$.

"Heaviest" tail in this framework is $\alpha = 1$ ($\alpha \downarrow$ decay slower)

similar to the normal distribution

the prototype distribution for subgaussians is the exponen distri.

given $P\{X \geq t\} = e^{-t}, t \geq 0$

after substituting $\alpha = 1$

Lemma 2.17 Let X be a r.v.

Then the following properties are equiv with parameter $K_i > 0$ differing from each other by at most an absolute constant factor:

1. $P(|X| > t) \leq \exp(-\frac{t}{K_1})$ for all $t \geq 0$

2. $(E|X|^p)^{\frac{1}{p}} \leq K_2 p \quad (p \geq 1)$

3. $E \exp(\frac{X}{K_3}) \leq e$.

Proof: special case of H.W.

Def 2.18 (Subexponential r.v.)

A random variable X that satisfies one of the properties $1 \sim 3$ in Lemma 2.17 is called a subexpo r.v.

The subexponen norm of X, denoted by $\|X\|_{\psi_1}$, is defined by

$$\|X\|_{\psi_1} = \sup_{p \geq 1} p^{-1} (E|X|^p)^{\frac{1}{p}}$$

Lemma 2.19 (Sub expo is subgaussian squared)

A r.v. X is subgaussian iff X^2 is subexpo.

Moreover,

$$\|X\|_{\psi_2}^2 \leq \|X^2\|_{\psi_1} \leq 2\|X\|_{\psi_2}^2$$

Pf: $\|X\|_{\psi_2}^2 = \left(\sup_{p\geq 1} p^{-\frac{1}{2}} (E|X|^p)^{\frac{1}{p}}\right)^2$

$$= \sup_{p\geq 1} p^{-1} (E|X|^p)^{\frac{2}{p}}$$

$$= \frac{1}{2} \sup_{p\geq 1} \left(\frac{p}{2}\right)^{-1} (E|X^2|^{\frac{p}{2}})^{\frac{2}{p}}$$

$$= \frac{1}{2} \sup_{q\geq \frac{1}{2}} q^{-1} (E|X^2|^q)^{\frac{1}{q}}$$

$$\geq \frac{1}{2} \sup_{q\geq 1} q^{-1} (E|X^2|^q)^{\frac{1}{q}}$$

$$= \frac{1}{2} \|X^2\|_{\psi_1}$$

other: $(E|X|^p)^{\frac{2}{p}} \underset{\text{Jensen}}{\leq} (E|X^2|^p)^{\frac{1}{p}}$

Observation: For equiv defining property for a centered varia to be subgaussian, only three for subexp.

Missing: Moment generating func (Mgf)

Prob: Even for the prototype (exp) Mgf is not finite for $t \geq 1$.

Need "local" version:

Lemma 2.20: (Mgf of subexp r.v.)

There exist absolute constant c, C s.t. the following holds:

Let X be a centered subexp r.v. Then, for t s.t. $|t| \leq \frac{c}{\|X\|_{\psi_1}}$, one has $E\exp(tX) \leq \exp(Ct^2\|X\|_{\psi_1}^2)$

Pf: As in the subgau case, w.l.o.g $\|X\|_{\psi_1} = 1$

We showed by Taylor series for centered variable

$$E\exp(tX) \leq 1 + \sum_{p=2}^{\infty} \frac{t^p EX^p}{p}$$

$$\overset{\text{subexp}}{\leq} 1 + \sum_{p=2}^{\infty} |t|^p \frac{p^p}{p!}$$

$$\overset{p! \geq \frac{p^p}{e^p}}{\leq} 1 + \sum_{p=2}^{\infty} (e|t|)^p$$

$$= 1 + e^2 t^2 \sum_{p=0}^{\infty} (e|t|)^p$$

$$\underset{\substack{\uparrow \\ \text{for } e|t|<1}}{=} 1 + e^2 t^2 \frac{1}{1 - e|t|}$$

$$\underset{\text{assume } e|t|<\frac{1}{2}}{<} 1 + 2e^2 t^2$$

$$\leq \exp(2e^2 t^2)$$

we have a subgaussian decay, after sum, we still have a subgaussian decay

We want to get tail bound

sum up subgau r.v. is an easy case because of asymptotic analysis (CLT).

By the CLT, a sum of indep subexp r.v. will be (sub)gaussian in the limit.

↓ but

For non-asym regime, we will have comb of subgau (from limiting behavior) and subexp (from individual r.v.)

In Rademacher case, we make diagonal entries vanished,
the challenge of gaussian case is in addition to the case what we have done in Rade case, we also need to handle the concentration of diagonal term, which is the sum of gaussian squared (minus expectation)
this new part is exactly the sum of subexp r.v.!

$\downarrow$

Proposition 2.21: (Bernstein-type ineq)
Let $X_1, \cdots, X_N$ be inde centered subexp r.v., and $K = \max_i \|X_i\|_{\psi_1}$,
Then for every $a = (a_1, \cdots, a_N) \in \mathbb{R}^N$ and every $t \geq 0$
we have $P\left(\left|\sum_{i=1}^{N} a_i X_i\right| \geq t\right) \leq 2\exp\left(-c\min\left(\frac{t^2}{K^2\|a\|_2^2}, \frac{t}{K\|a\|_\infty}\right)\right)$
where $c > 0$ is an absolute constant.

Pf: Using homog, ass again that $K=1$

set $S = \sum_i a_i X_i$

Then $P\{S \geq t\} = P(e^{\lambda S} \geq e^{\lambda t})$

$$\leq e^{-\lambda t} E e^{\lambda S}$$

$$= e^{-\lambda t} \prod_{i=1}^{N} E \exp(\lambda a_i X_i)$$

We want to apply lemma 2.20

Need: $|\lambda a_i| \leq C \quad \forall i$

$\rightarrow$ choose $\lambda \leq \frac{C}{\|a\|_\infty}$

Thm. $P\{S \geq t\} \underset{\substack{\uparrow \\ \text{Lemma 2.20}}}{\leq} e^{-\lambda t} \prod_i \exp(c\lambda^2 a_i^2)$

$$= \exp(-\lambda t + c\lambda^2 \|a_i\|_2^2)$$

To control the expo, we want $\lambda \leq \frac{t}{2c\|a\|_2^2}$

choose $\lambda = \min\left(\frac{t}{2c\|a\|_2^2}, \frac{c}{\|a\|_\infty}\right)$, $\lambda c \|a\|_2^2 \leq \frac{t}{2}$

Then

$$P(S \geq t) \leq \exp\left(-\frac{\lambda}{2} t\right)$$

$$= \exp\left(-\min\left(\frac{t^2}{4c\|a\|_2}, \frac{ct}{2\|a\|_\infty}\right)\right)$$

similarly,

$P(S \geq t) \leq \cdots, \Rightarrow P(|S| \geq t) \leq 2 \cdots$.

Corollary 2.22

Let $X_1, \cdots, X_N$ be inde central subexp r.v., and let $K = \max_i \|X_i\|_{\psi_1}$, then for every $\varepsilon \geq 0$, we have

$$P\left(\left|\sum_{i=1}^{N} X_i\right| \geq \varepsilon N\right) \leq 2\exp\left(-c\min\left(\frac{\varepsilon^2}{K^2}, \frac{\varepsilon}{K}\right)N\right)$$

$\Rightarrow$ divide N at both sides will get an average (similar to CLT)

Proof: $a = (1, \cdots, 1)$; $t = \varepsilon N$

subexp does not apply for exp

Remark:

result for centered subexp (or subgau)

r.v. also have direct implications for non-centered ones

Indeed, if X is subexp (or subgau) then so is $X - EX$ and (for $\alpha = 1, 2$):

$$\|X - EX\|_{\psi_\alpha} \leq \|X\|_{\psi_\alpha} + E|X| \leq 2\|X\|_{\psi_\alpha}$$

for non-centered r.v. we only need double the constant