

## Sub-Exponentials (cont.)

Recall:  $X$  is a subE( $v^2, \alpha$ ) iff

$$\forall |\lambda| < \frac{1}{\alpha} : E[e^{\lambda(X-\mu)}] \leq e^{v^2 \lambda^2 / 2}$$

---

\* sum of independent sub-exponentials

$n$  independent r.v.  $\rightarrow X_1, X_2, \dots, X_n$

$$E[X_i] = \mu_i \text{ and } X_i \in \text{SubE}(v_i^2, \alpha_i)$$

$$\Rightarrow E[e^{\lambda(X_i - \mu_i)}] \leq e^{v_i^2 \lambda^2 / 2}$$

for every  $\lambda$  s.t.  $|\lambda| < \frac{1}{\alpha_i}$

what can be said about the sum?

$$Y := \sum X_i - \mu_i$$

$$\mathbb{E} [ e^{\lambda Y_i} ] = \mathbb{E} [ e^{\lambda \sum_{i=1}^n (X_i - \mu_i)} ]$$

$$= \mathbb{E} \left[ \prod_{i=1}^n e^{\lambda (X_i - \mu_i)} \right]$$

$$= \prod_{i=1}^n \mathbb{E} [ e^{\lambda (X_i - \mu_i)} ] \quad (\text{by independent})$$

$$\leq \prod_{i=1}^n e^{\sqrt{v_i \lambda^2 / 2}}$$

$$|\lambda| \leq \frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \dots, \frac{1}{\alpha_n}$$

$$|\lambda| < \frac{1}{\max \alpha_i}$$

$$= e$$

$$\Rightarrow \gamma \in \text{Sub}\mathcal{E} \left( \sum v_i^2, \max \alpha_i \right)$$

\* Scaling : if  $X$  is zero-mean and  $\text{SubE}(v^2, \alpha)$   
 $\Rightarrow c \cdot X$  is zero-mean and  $\text{SubE}(c^2 v^2, c\alpha)$

Proof :

$$MGF_{cX}(\lambda) = E[e^{\lambda c X}] = MGF_X(\lambda c)$$

$$= e^{v^2 \lambda^2 c^2} = e^{(cv)^2 \lambda^2}$$

$$\leq e$$



$$\text{if } |\lambda c| < \frac{1}{\alpha}$$

$$\Rightarrow cX \in \text{SubE}(c^2 v^2, c\alpha).$$

Bernstein's inequality :

Suppose  $X_1, \dots, X_n$  are  $n$  independent zero-mean r.v. and  $X_i \sim \text{Sub } E(1, 1)$ .

Let  $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ . Then

$$\sum a_i X_i \in \text{Sub } E(\|a\|_2^2, \|a\|_\infty).$$

Hence,

$$\Pr \left[ \left| \sum_{i=1}^n a_i X_i \right| \geq t \right] \leq 2 \exp \left( -\min \left( \frac{t^2}{2\|a\|_2^2}, \frac{t}{2\|a\|_\infty} \right) \right)$$

\* when  $a_i = \frac{1}{n}$

$$\Pr \left[ \left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq t \right] \leq \exp \left( -n \min \left( -\frac{t^2}{2}, \frac{t}{2} \right) \right)$$

for only one sub-exponential r.v.  $X_i$

if  $X_i \in \text{SubE}(1, 1)$ , for  $t \geq 1$

$$\Pr \left[ \left| \frac{1}{n} \sum X_i \right| \geq t \right] = \Pr \left[ |X_i| \geq nt \right]$$

$$\leq 2 \exp \left( - \frac{nt}{2} \right) \quad t \geq 1$$

- Observe that the tail bound in the previous page is exactly the same for the sum.

---

Compare with CLT:

$$\Pr \left[ \left| \frac{1}{\sqrt{n}} \sum X_i \right| \geq t \right] = \begin{cases} 2 \exp \left( - \frac{t^2}{2} \right) & t \leq \sqrt{n} \\ 2 \exp \left( - \frac{t\sqrt{n}}{2} \right) & t \geq \sqrt{n} \end{cases}$$

CLT: for very large  $n$ , we have a Gaussian tail

(not true when  $t$  depends on  $n$ )

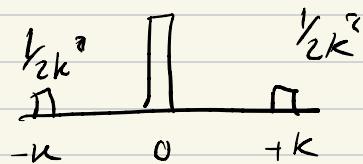
## Bernstein Condition

A more general condition that implies

Sub exponentiality with  $s^2 \leq \text{var}$

A better fit when we have fat tail but low variance.

Recall our example



Definition :

we say that a r.v.  $X$  with mean  $\mu := E[X]$  has Bernstein condition with parameter  $b$  iff,

$$\text{for } i=3,4,5,\dots /E[(X-\mu)^i]/ \leq \frac{1}{2} i! \propto b^{i-2}$$

centered moment

why this weird inequality?

It helps us to bound MGF

$$M.G.F_{X-\mu}(\lambda) = E[e^{\lambda(X-\mu)}]$$

$$\begin{aligned} e^x &= \sum_{i=0}^{\infty} \frac{x^i}{i!} \\ &= E\left[\sum_{i=0}^{\infty} \frac{\lambda^{(X-\mu)^i}}{i!}\right] \\ &= 1 + \underbrace{\lambda E[X-\mu]}_{=0} + \lambda \underbrace{\frac{E[(X-\mu)^2]}{2}}_{\sigma^2 \text{Var}(X)} \\ &\quad + \sum_{i=3}^{\infty} \lambda^i \frac{E[(X-\mu)^i]}{i!} \end{aligned}$$

$$\leq 1 + \frac{\lambda^2 \sigma^2}{2} + \sum_{i=3}^{\infty} |\lambda|^i \frac{\sigma^i}{2} b^{i-2}$$

Bernstein condition  $\rightarrow$   
bounds centered moments

$$= 1 + \frac{\lambda^2 \sigma^2}{2} + \frac{\lambda^2 \sigma^2}{2} + \sum_{i=3}^{\infty} |\lambda|^i b^{i-2}$$

$$\leq 1 + \frac{1}{2} \frac{\alpha^2}{\lambda^2} \left( 1 + \sum_{i=1}^{\infty} |\lambda|^i b^i \right)$$

$$\leq 1 + \lambda \frac{\alpha^2}{2} \left( \sum_{i=0}^{\infty} |\lambda|^i b^i \right)$$

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x} \quad \text{if } |x| < 1$$

$$= 1 + \lambda \frac{\alpha^2}{2} \frac{1}{1 - |\lambda| b}$$

*converges if  $|\lambda|b < 1$*

$$\leq \exp \left( \frac{\lambda^2 \alpha^2}{2(1 - |\lambda|b)} \right) \quad *$$

$$\text{if } |\lambda|b < \frac{1}{2} \quad \leq \exp \left( \frac{\lambda^2 \alpha^2}{4} \right)$$

$X$  is sub- $E$   $((2\alpha)^2, 2b)$

we just proved:

Lemma

Bernstein condition w. parameter  $b$

$\Rightarrow$  Sub $E((2\alpha)^2, 2b)$

Using equation  $*$  directly, or

using the properties of sub-exponential r.v.s

we can get other versions of Bernstein's inequality:

$$\Pr \{ |X - \mu| \geq t \} \leq 2 \exp \left( \frac{-t^2/2}{\sigma^2 + tb} \right)$$

bernstien's condition for bounded variables

Suppose  $|X - \mu| < B \Rightarrow$  then  $X$  satisfies Bernstein condition for  $b = B/3$

$$\mathbb{E}[(X - \mu)^i] \leq \mathbb{E}[(X - \mu)^2 | X - \mu|^{i-2}]$$

$$\leq \mathbb{E}[(X - \mu)^2] \cdot \mathbb{E}[|X - \mu|^{i-2}]$$

$$\leq B^{i-2} \cdot \sigma^2 \leq \sigma^2 B^{i-2} \left( \frac{i!}{2 \cdot 3^{i-2}} \right)$$

$$= \frac{\sigma^2}{2} i! \left( \frac{B}{3} \right)^{i-2}$$

show this is at least one.

$\Rightarrow$  Bernstein condition for  $b = B/3$

Using this information we get the following version of Bernstein's inequality:

Let  $X_1, X_2, \dots, X_n$  be independent with  $E[X_i] = \mu$ ,  $\text{var}(X_i) = \sigma^2$ , and range  $|X_i - \mu| \leq B$ . Then,

$$\Pr \left[ \left| \sum_{i=1}^n (X_i - \mu) \right| \geq t \right] \leq 2 \exp \left( - \frac{t^2/2}{n\sigma^2 + Bt/3} \right)$$

or, the normalized version:

$$\Pr \left[ \frac{1}{n} \left| \sum_{i=1}^n (X_i - \mu) \right| \geq t \right] \leq 2 \exp \left( - \frac{nt^2/2}{\sigma^2 + Bt/3} \right)$$

$2\delta$

If we set the right side to  $2\delta$ , and some manipulation  $\rightarrow$  the following holds with probability  $1-\delta$ :

$$\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \leq \sqrt{\frac{2\sigma^2 \log^2 \delta}{n}} + \frac{B \log^2 \delta}{n}$$

\*\*

Note: if  $\sigma^2 \ll B$ . Then, the error goes down with  $\frac{1}{n}$  instead of  $\frac{1}{\sqrt{n}}$  (as we would get by CLT)

Application: mis-labeling probability

Suppose we have a function  $f: X \rightarrow \{0, 1\}$   
(fixed and pick independent of data)

Suppose we have  $n$  labeled examples:

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n) \sim P$$

↑      ↑      ↑

feature      label  $\in \{0, 1\}$       data distribution

if  $f$  labels all examples correctly,  
can we find an upper bound for  
the classification error of  $f$ ?

$$\text{err}(f) = \Pr_{(X, Y) \sim P} [f(x) \neq Y]$$

$$\text{Let } Z_i = \begin{cases} 1 & \text{if } f(x_i) \neq Y \\ 0 & \text{otherwise} \end{cases}$$

bounded  $\rightarrow B=1$

$Z_i$  is a Bernoulli r.v. with :

$$p := E[Z_i] = \Pr[f(x_i) = Y] = \text{err}(f)$$

$$\text{Var}[Z_i] = p(1-p)$$

Using \*\*, with probability  $1-\delta$ ,

we have:

$$\text{err}(f) = \left| \underbrace{\frac{\sum Z_i}{n}}_{\text{---}} - p \right| \leq \sqrt{\frac{2p(1-p)u}{n}} + \frac{u}{3n}$$

must be zero, since  $f$  label them perfectly

$$\text{where } u = \log(2/\delta)$$

$$P \leq \sqrt{\frac{2Pu}{n}} + \frac{u}{3n}$$

Let  $x = \sqrt{P}$

we want to find the range of  $x$  s.t.

$$x^2 \leq \sqrt{\frac{2u}{n}} x + \frac{u}{3n}$$

$$\Rightarrow x^2 - \sqrt{\frac{2u}{n}} x - \frac{u}{3n} \leq 0$$



$$x \leq \left( \sqrt{\frac{5}{6}} + \frac{1}{\sqrt{2}} \right) \cdot \sqrt{\frac{u}{n}}$$

$$\approx 1.7 \sqrt{\frac{u}{n}}$$

$$\Rightarrow P = x^2 \leq 3 \sqrt{\frac{u}{n}}$$

$x$  lies between zero

and the positive root