Sub-Exponentials (cont.)

bound on moments $\overset{\text{implies}}{\Longrightarrow}$ bound on MGF of a sub-Exponential

Lemma Suppose $E[X]=0$

$$\|X\|_{L^p} = \left(E[|X|^p]\right)^{1/p} \leq Cp \quad \forall p \geq 1$$

$$\Downarrow$$

$$E[e^{\lambda X}] \leq \exp(c'^2 \lambda^2) \text{ for } |\lambda| \leq \frac{1}{c'}$$

proof

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

$$E[e^{\lambda X}] = E\left[\sum_{p \geq 1}^{\infty} \frac{(\lambda X)^p}{p!}\right]$$

$$= E\left[1 + \lambda X + \sum_{p=2} \frac{(\lambda X)^p}{p!}\right]$$

$$= 1 + \lambda \underbrace{E[X]}_{=0} + \sum_{p=2}^{\infty} \frac{\lambda^p E[X^p]}{p!}$$

$\mathbb{E}[|X|^p] \leq (cp)^p$

$$\leq 1 + \sum_{p=2}^{\infty} \frac{(\lambda c p)^p}{p!}$$

stirling's approx:

$(p/e)^p \leq p!$

$$\leq 1 + \sum_{p=2}^{\infty} \frac{(\lambda c \cancel{p})^p}{(\cancel{p}/e)^p}$$

$$= 1 + \sum_{p=2}^{\infty} (\lambda c e)^p$$

for $|\lambda c e| < 1$ the series converges

$$= 1 + \frac{(ec\lambda)^2}{1 - ec\lambda}$$

if $|\lambda c e| < \frac{1}{2}$

$$\leq 1 + 2(ec\lambda)^2$$

$$\leq e^{2e^2c^2\lambda^2} \leq e^{(2ec)^2\lambda^2}$$

for all $|\lambda| < \frac{1}{2ce}$

Recall: X is a sub $E(\nu^2, \alpha)$ iff

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

---

* sum of independent sub-exponentials

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

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

$$\Rightarrow \mathbb{E}\left[e^{\lambda(X_i-\mu_i)}\right] \leq e^{\nu_i^2\lambda^2/2}$$

for every λ s.t. $|\lambda| < 1/\alpha_i$

- what can be said about the sum?

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

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

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

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

$$\leq \prod_{i=1}^{n} e^{v_i^2 \lambda^2 / 2} \qquad |\lambda| \leq \frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \ldots, \frac{1}{\alpha_n}$$

$$= e^{\left(\sum_{i=1}^{n} v_i^2\right)\lambda^2/2} \qquad |\lambda| < \frac{1}{\max \alpha_i}$$

$$\Rightarrow \quad Y \in \mathrm{SubE}\left(\sum v_i^2, \max \alpha_i\right)$$

* Scaling : if X is zero-mean and $SubE(v^2, \alpha)$

$\Rightarrow$ $c.X$ is zero-mean and $SubE(c^2v^2, c.\alpha)$

proof:

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

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

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

$$\Rightarrow \quad cX \in SubE(c^2v^2, c\alpha).$$

Bernstein's inequality:

Suppose $X_1, \ldots, X_n$ are n independent zero-mean r.v. and $X_i \sim \text{SubE}(1,1)$. Let $a = (a_1, \ldots, a_n) \in R^2$. Then

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

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 $X_i \in SubE(1,1)$, for $t \geq 1$

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

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

- 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 long n, we have a Gaussian tail

(not true when t depends on n)

Bernstein Condition

A more general condition that implies

Sub exponentiality with $s^2 \simeq var$

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

Recall our example

Definition :

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

$$\text{for } i = 3, 4, 5, \ldots \quad \Big| \underbrace{\mathbb{E}\left[(X-\mu)^i\right]}_{\text{centered moment}} \Big| \leq \frac{1}{2} i! \, \sigma^2 b^{i-2}$$

why this weird inequality?

It helps us to bound MGF.

$$MGF_{X-\mu}(\lambda) = E\left[e^{\lambda(X-\mu)}\right]$$

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

$$= E\left[\sum_{i=0}^{\infty} \frac{\lambda^i (X-\mu)^i}{i!}\right]$$

$$= 1 + \lambda \underbrace{E[X-\mu]}_{=0} + \frac{\lambda^2 \overbrace{E[(X-\mu)^2]}^{\sigma^2 = Var(X)}}{2} + \sum_{i \geq 3}^{\infty} \frac{\lambda^i E[(X-\mu)^i]}{i!}$$

Bernstein condition bounds centered moments

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

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

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

$$\leq 1 + \frac{\lambda^2 \sigma^2}{2}\left(\underbrace{\sum_{i=0}^{\infty} |\lambda|^i b^i}_{\text{converges}}\right)$$

if $|\lambda b| < 1$

$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$

$|x| < 1$

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

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

if $|\lambda b| < \frac{1}{2}$

$$\leq \exp\left(\frac{\lambda^2 \sigma^2}{4}\right)$$

X is sub-$E\left((2\sigma)^2, 2b\right)$

we just proved:

Lemma

Bernstein condition w. parameter b

$\Rightarrow$ Sub$E\left((2\sigma)^2, 2b\right)$

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's condition for $b = B/3$

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

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

$$\leq B^{i-2} \cdot \sigma^2 \leq \sigma^2 B^{i-2} \underbrace{\left(\frac{i!}{2 \cdot 3^{i-2}}\right)}_{\text{show this is at least one.}}$$

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

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

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

Let $X_1, X_2, \ldots, X_n$ be independent with $E[X_i] = \mu$, $var(X_i) = \sigma^2$, 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:

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

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

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

$**$

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

Application : mis-labeling probability

Suppose we have a function $f : \mathcal{X} \to \{0,1\}$

(fixed and pick independent of data)

Suppose we have n labeled examples:

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

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 ?

$$\mathrm{err}(f) = \Pr_{(X,Y)\sim P}\left[f(X) = Y \right]$$

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

bounded $\rightarrow B = 1$

Z_i is a Bernoulli r.v. with :

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

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

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

we have:

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

must be zero, since f label them perfectly

where $u = \log(2/\delta)$

$$p \leq \sqrt{\frac{2pu}{n}} + \frac{u}{3n}$$

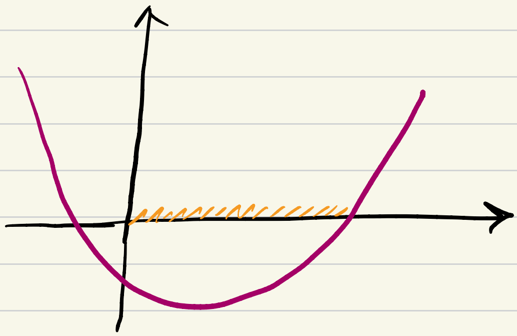
Let $x := \sqrt{p}$

we want to find the

range of x s.t.

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

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



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

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

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

x lies between zero

and the positive root