Today's Lecture

Sub-Gaussian

(Adapted from Sash Rakhlin's

lecture notes

what random variables behave similar to Gaussians?

we saw in the last lecture:

$$\Pr[\, z > t \,] \simeq \Theta\left(\frac{1}{t} e^{-t^2/2} \right)$$

Sub-Gaussian random variables mimic the same behavior.

* We say X is a sub-Gaussian random variable with variance proxy (a.k.a. variance factor or sub-Gaussianity parameter) K^2 iff

$$\Pr[\, |X| \geq t \,] \leq 2\exp\left(-t^2/K^2 \right)$$

our notation: $X \in \mathrm{subG}(K)$

The following properties are equivalent.

K_i's appearing below differ from each other by at most an absolute constant factor.

1) for all $t \geq 0$ (tail)

$$\Pr[\,|X| \geq t\,] \leq 2\exp(-t^2/K_1^2)$$

2) for all $p \geq 1$ (moment)

$$\|X\|_{L^p} = (\mathbb{E}[\,|X|^p\,])^{1/p} \leq K_2\sqrt{p}$$

3) $|\lambda| \leq \frac{1}{K_3}$ MGF of X^2

$$\mathbb{E}[e^{\lambda^2 X^2}] \leq e^{K_3^2 \lambda^2}$$

4) 〃

$$\mathbb{E}[e^{X^2/K_4^2}] \leq 2$$

5) $\forall \lambda \in \mathbb{R}$ (if X is zero mean) MGF of X

$$\mathbb{E}[e^{\lambda X}] \leq e^{K_5^2 \lambda^2}$$

which one is the main one? all are correct.

In the literature, you may see various versions.

We stick to the definition in the Vershynin's book.

When we say $X \in \mathrm{SubG}(K^2)$ we mean:

$$X \in \mathrm{SubG}(\Theta(K^2))$$

$\boxed{\text{Examples}}$

+ $Z \sim N(0,1)$

$$\Pr[|Z| \geq t] \leq 2e^{-t^2/2}$$

$$Z \in \mathrm{subG}(\Theta(1))$$

$$Z \sim N(0,\sigma^2) \Rightarrow \mathrm{subG}(\Theta(\sigma^2))$$

* Bernouli / Radamacher

$$X = \begin{cases} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}$$

$$\mathbb{E}[e^{\lambda X}] = \frac{1}{2} e^{\lambda} + \frac{1}{2} e^{-\lambda} \overset{?}{\leq} e^{\lambda^2/2}$$

↳ we hope to show

$$\leq \frac{1}{2} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + \frac{(-\lambda)^k}{k!}$$

$$= \frac{1}{\cancel{2}} \sum_{k=0}^{\infty} \cancel{2} \frac{\lambda^{2k}}{(2k)!}$$

$$\leq \sum_{k=0}^{\infty} \frac{(\lambda^2)^k}{2^k k!} = e^{\lambda^2/2}$$

$$\Rightarrow \quad \mathbb{E}[e^{\lambda X}] \leq e^{\lambda^2/2} \Rightarrow \boxed{X \in \text{subG}(1)}$$

by 5

* bounded variables $a \le b$

first : $Y := \begin{cases} a & \frac{1}{2} \\ b & \frac{1}{2} \end{cases}$

$\rightarrow$ centered $Y' := Y - E[Y]$ $\begin{cases} \frac{b-a}{2} & \frac{1}{2} \\ -\frac{b-a}{2} & \end{cases}$

by rescaling $Y' = \frac{(b-a)}{2} X$ (like a Ber)

$\Rightarrow$ $Y' \in SubG\left(\frac{(b-a)^2}{4}\right)$

In fact for any bounded r.v. Z in $\left[-\frac{b-a}{2}, \frac{b-a}{2}\right]$. if $E[Z]=0$, we have:

$$Z \in SubG\left(\frac{(b-a)^2}{4}\right)$$

Hoeffding lemma

Suppose X is a zero-mean r.v in $[a, b]$

$$\mathbb{E}[\lambda X] \leq \exp\left(\frac{\lambda^2 (b-a)^2}{8}\right) \quad \forall \lambda \in \mathbb{R}$$

$$\Rightarrow \quad X \text{ is in } \operatorname{subG}\left(\frac{(b-a)^2}{8}\right)$$

(with a slightly more compilicated proof that we omit here.)

+ for two independent $X_1 \in subG(\sigma_1)$ and $X_2 \in subG(\sigma_2)$

$$\Rightarrow \quad X_1 + X_2 \in subG(\sigma_1^2 + \sigma_2^2)$$

(prove it for the problem set)

+ Also, $X_1, \ldots, X_n$ $\quad X_i \in SubG(\sigma_i^2)$

$$\Rightarrow \sum X_i \in SubG\left(\sum \sigma_i^2\right)$$

+ X_i are zero-mean r.v. in $[a,b]$

$$\sum_{i=1}^{n} X_i \text{ is in } SubG\left(n \cdot \frac{(b-a)^2}{8}\right)$$

by Hoeffding lemma.

Hoeffding bound

Let $X_1, \dots, X_n$ be n i.i.d

r.v. in range $[a, b]$

$$\Pr\left[\left|\frac{\sum X_i - E[X_i]}{n}\right| \geq \varepsilon\right] \leq 2e^{-\theta\left(\frac{\varepsilon n^2}{(b-a)^2}\right)}$$

proof. $Y_i = X_i - E[X_i]$ is zero-mean

in $\left[a - E[X_i], b - E[X_i]\right]$

$$\Rightarrow \sum_{i=1}^{n} Y_i \sim \text{subG}\left(n \cdot \frac{(b-a)^2}{8}\right)$$

$$\Rightarrow \Pr\left[\frac{\sum X_i - E[X_i]}{n} \geq \varepsilon\right] = \Pr\left[\left|\sum Y_i\right| \geq \varepsilon n\right]$$

$$\leq 2\exp\left(-\frac{8\varepsilon^2 n}{(b-a)^2}\right)$$

Some proofs regarding the definitions

Integral identity of expectation for non-negative random variables: $Y \geq 0$

$$E[Y] = \int_0^\infty \Pr[Y > t]\, dt$$

1 => 2

$$E[|X|^p] = \int_0^\infty \Pr[|X|^p > t] . dt$$

$$= \int_0^\infty \Pr[|X| \geq \sqrt[p]{t}\,]\, dt$$

change of variable $u = \sqrt[p]{t} \Rightarrow t = u^p \Rightarrow \frac{dt}{du} = p u^{p-1}$

$$= \int_0^\infty \Pr[|X| \geq u] . p u^{p-1} du$$

$$\leq \int_0^\infty 2e^{-u^2/k_1^2} \cdot p \cdot u^{p-1} \cdot du$$

by ①

another change of variable

$$z := \frac{u^2}{k_1^2} \Rightarrow k_1\sqrt{z} = u \Rightarrow \frac{du}{dz} = \frac{k_1}{2}\frac{1}{\sqrt{z}}$$

$$= \int_0^\infty \cancel{2} e^{-z} \cdot p \cdot (k_1\sqrt{z})^{p-1} \frac{k_1}{\cancel{2}} \frac{1}{\sqrt{z}} dz$$

$$= k_1^p \int_0^\infty e^{-z} p (\sqrt{z})^{p-2} dz$$

$$= p k_1^p \int_0^\infty e^{-z} (z)^{p/2 - 1} dz$$

$$= p k_1^p \cdot \Gamma(p/2) \leq p k_1^p (p/2)^{p/2}$$

$$\Rightarrow \left(E[|X|^p]\right)^{1/p} \leq \frac{p}{\sqrt{2}}^{1/p} k_1 \sqrt{p} < 1.06 k_1 \sqrt{p}$$

(5) $\Rightarrow$ (1)

for $\lambda > 0$

$$\Pr[X \geq t] = \Pr[e^{\lambda X} \geq e^{t\lambda}]$$

$$\underset{\text{markov}}{\leq} \inf_{\lambda>0} e^{-t\lambda} E[e^{\lambda X}]$$

$$\leq \inf_{\lambda>0} e^{K_5^2 \lambda^2 - t\lambda} = e^{-t^2/4K_5^2}$$

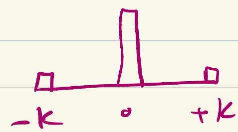
$\lambda = \frac{t}{2K_5^2}$

For the rest of proofs see Vershynin's book.

Is sub-Gaussia tail always describe the behavior of a r.v. well?

$$X = \begin{cases} 0 & \text{w.p. } 1 - 1/k^2 \\ +k & \text{w.p. } 1/2k^2 \\ -k & \text{w.p. } 1/2k^2 \end{cases}$$

for some large k



$$E[X] = 0, \quad \text{Var}[X] = k^2 \cdot \frac{1}{2k^2} + k^2 \frac{1}{2k^2} = 1$$

Suppose we have n i.i.d copy of X:

$$\Pr[X_1 = \dots = X_n = 0] = \left(1 - \frac{1}{k^2}\right)^n$$

$$\simeq e^{-n/k^2}$$

$$\text{or} \simeq 1 - \frac{n}{k^2}$$

for any $t > 0$

$$\Pr\left[\left|\sum X_i\right| \geq t\right] < \Pr[\exists i : X_i \neq 0]$$

$$\leq 1 - \Pr[X_1 = \dots = X_n] \simeq \frac{n}{k^2} = \frac{1}{k}$$

almost 0

for $n = k$

using sub-Gaussianity of bounded random variables:

$$X \in SubG(\theta(k)^2)$$

$$\Rightarrow \quad \Pr[\,|\Sigma X_i| > t\,] \leq \exp\left(-\theta\left(\frac{t^2}{n \cdot k^2}\right)\right)$$

for $n = k$ $\quad \leq \exp\left(-\theta\left(\frac{t^2}{k}\right)\right)$

for small t, this is almost 1.

→ Hoeffding gives us a very bad upper bound on the probability.

Is every random variable subG(k) for some large k? Nope.

Let $Z \sim N(0,1)$

Consider Z^2

centered

The MGF of Z^2-1 is

$$E\left[e^{\lambda(Z^2-1)} \right] =$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda(z^2-1)} \cdot e^{-z^2/2} \, dz$$

$$= \frac{e^{-\lambda}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}(1-2\lambda)} \, dz$$

$$= \begin{cases} \dfrac{e^{-\lambda}}{\sqrt{1-2\lambda}} \overset{\frac{1}{4} \geq \lambda \geq 0}{\leq} e^{\lambda^2/1-2\lambda} \overset{\frac{1}{4} \geq \lambda \geq 0}{\leq} e^{2\lambda^2} & \lambda < 1/2 \\ \infty & \lambda \geq 1/2 \end{cases}$$

↳ unbounded :(

for $|\lambda| \leq \frac{1}{4}$: Z^2 is behaving like a sub-Gaussian