Lecture 8-14-2020

Sub-Gaussian Distributions

I. Motivation:

① Exploration of Concentration Inequalities:

$$\mathbb{P}(|X-\mu|>t) \xrightarrow{\text{Markov}} \frac{\mathbb{E}|X-\mu|}{t} \leq \frac{\mathbb{E}|X|+\mu}{t},\ \frac{(\mathbb{E}|X-\mu|^2)^{1/2}}{t}$$

$$\mathbb{P}(|X-\mu|>t) = \mathbb{P}(|X-\mu|^2>t^2) \xrightarrow{\text{Chebyshev}} \frac{\mathbb{E}|X-\mu|^2}{t^2}$$

$\mu := \mathbb{E}X$

$$= \mathbb{P}(e^{(X-\mu)^2} > e^{t^2}) \xrightarrow{\text{Hoeffding}} e^{-t^2}\,\mathbb{E}(e^{(X-\mu)^2})$$

exponential decay for tail prob.

Q. $\mu := \mathbb{E}X$ matters or not?

A: Centering Lemma

② Summary:

$$\text{Hoeffding's} \Downarrow\Uparrow \text{sub-gaussian}$$

Examples: {Gaussian, Bernoulli, Bounded}

II. Definition and Examples

1) Def. 2.5.6 (Sub-gaussian random variables):

A random variable X that satisfies one of the equivalent properties (i)–(iv) in Proposition 2.5.2 is called a sub-gaussian random variable.

The sub-gaussian norm of X, denoted $\|X\|_{\psi_2}$, is defined to be the smallest K_4 in property (iv). I.e. we define

$$\|X\|_{\psi_2} = \inf\left\{ t>0 : \mathbb{E}\, e^{\frac{X^2}{t^2}} \le 2 \right\}.$$

Remark: (Relation to standard deviation of normal distribution)

$$\mathbb{P}(|X|>t) = \mathbb{P}\left(\frac{|X|}{K_4} > \frac{t}{K_4}\right) = \mathbb{P}\left(e^{\frac{|X|^2}{K_4^2}} > e^{\frac{t^2}{K_4^2}}\right)$$

$$\le e^{-\frac{t^2}{K_4^2}}\, \mathbb{E}\left(e^{\frac{|X|^2}{K_4^2}}\right)$$

$$\overset{(iv)}{\le} 2 e^{-\frac{t^2}{K_4^2}}$$

Recall Propo 2.5.2: Equivalent properties of sub-gaussian introduce absolute constants s.t. $K_i \le C K_j$, for any $1 \le i, j \le 5$

If $X \sim N(0, \sigma^2)$, then $\mathbb{P}(|X|>t) \le e^{-\frac{t^2}{2\sigma^2}}$

$\Rightarrow \quad C\sigma^2 \in \left\{ K_4 : \mathbb{E}\, e^{\frac{X^2}{t^2}} \le 2 \right\}$

$\Rightarrow \quad \|X\|_{\psi_2}^2 \le C\sigma^2.$

Hence

(iv) $\mathbb{E}(\exp(X^2/\|X\|_{\psi_2}^2)) \le 2$

(i) $\mathbb{P}(|X| \ge t) \le 2\exp\left(-\frac{ct^2}{\|X\|_{\psi_2}^2}\right)$, $\forall\ t \ge 0$.

(ii) $\|X\|_{L^p} \le C\|X\|_{\psi_2}\sqrt{p}$, $\forall\ p \ge 1$.

(v) if $\mathbb{E}X = 0$, $\mathbb{E}(\lambda X) \le \exp(C\lambda^2\|X\|_{\psi_2}^2)$, $\forall\ \lambda \in \mathbb{R}$.

2) Other examples of sub-gaussian distributions:

a) Bernoulli

Let X be a random variable with symmetric Bernoulli distribution.

By def, $|X| \equiv 1 \Rightarrow$

$$\|X\|_{\psi_2} := \inf\left\{K:\ 2 \ge \mathbb{E}\left(\exp\left(\frac{X^2}{K}\right)\right) = \mathbb{E}\left(e^{\frac{1}{K^2}}\right) = e^{\frac{1}{K^2}}\right\}$$
$$= \frac{1}{\sqrt{\ln 2}}.$$

b) Bounded

Any bounded random variable X is sub-gaussian with

$$\|X\|_{\psi_2} \le c\,\|X\|_\infty$$

Recall def. of $\|X\|_{\psi_2}$ and apply similar reasoning in (a). ✓

III. General Hoeffding's Inequality

1) Recall : Let $X_1, \dots, X_N$ be independent normal random variables, then

$$\sum_{i=1}^{N} X_i \sim N\left(0, \sum_{i=1}^{N} \sigma_i^2\right)$$

Proposition 2.6.1 (Sums of independent sub-gaussians)

Let $X_1, \dots, X_N$ be independent mean-zero sub-gaussian random variables. Then $\sum_{i=1}^{N} X_i$ is also a sub-gaussian random variable, and

$$\left\| \sum_{i=1}^{N} X_i \right\|_{\psi_2}^2 \le C \sum_{i=1}^{N} \| X_i \|_{\psi_2}^2$$

Pf We analyze the MGF of $\sum_{i=1}^{N} X_i$.

For any $\lambda \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E} \exp\left(\lambda \sum_{i=1}^{N} X_i\right) &= \prod_{i=1}^{N} \mathbb{E}\, e^{\lambda X_i} \quad \text{by independence} \\ &\le \prod_{i=1}^{N} \exp\left(C \lambda^2 \| X_i \|_{\psi_2}^2\right) \\ &= \exp\left(C \lambda^2 \sum_{i=1}^{N} \| X_i \|_{\psi_2}^2\right). \end{aligned}$$

Hence, Prop 2.5.2 implies $\sum_{i=1}^{N} X_i$ is sub-gaussian, and

$$\left\| \sum_{i=1}^{N} X_i \right\|_{\psi_2}^2 \le C \sum_{i=1}^{N} \| X_i \|_{\psi_2}^2 .$$

Q.E.D.

Theorem 2.6.2 (General Hoeffding)

Let $X_1, \cdots, X_N$ be independent mean-zero sub-gaussian random variables. Then for every $t \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^N X_i\right| \geq t\right) \leq 2\exp\left(-\frac{ct^2}{\sum_{i=1}^N \|X_i\|_{\psi_2}^2}\right)$$

Sketch of pf: $\forall \lambda > 0$,

$$\mathbb{P}\left(\sum_{i=1}^N X_i \geq t\right) = \mathbb{P}\left(e^{\lambda \sum_{i=1}^N X_i} \geq e^{\lambda t}\right)$$

$$\leq e^{-\lambda t} \, \mathbb{E}\, e^{\lambda \sum_{i=1}^N X_i}$$

$$\leq e^{-\lambda t} e^{c\lambda^2 \sum_{i=1}^N \|X_i\|_{\psi_2}^2} \quad (\forall \lambda)$$

Minimize the exponent and let $\lambda := \dfrac{t}{2c \sum_{i=1}^N \|X_i\|_{\psi_2}^2}$

$$\mathbb{P}\left(\sum_{i=1}^N X_i \geq t\right) \leq e^{-\frac{ct^2}{\sum_{i=1}^N \|X_i\|_{\psi_2}^2}}.$$

Corollary (Khintchine's inequality)

Let $X_1, \cdots, X_N$ be independent sub-gaussian random variables with zero means and unit variances, and let $a = (a_1, \cdots, a_N) \in \mathbb{R}^N$. Prove that for any $p \in [2, \infty)$,

$$\left(\sum_{i=1}^N a_i^2\right)^{\frac{1}{2}} \leq \left\|\sum_{i=1}^N a_i X_i\right\|_{L^p} \leq CK\sqrt{p}\left(\sum_{i=1}^N a_i^2\right)^{\frac{1}{2}}$$

where $K = \max \|X_i\|_{\psi_2}$ and C is an absolute constant.

2) Centering

"De-mean" does not harm the sub-gaussian property:

<u>Observation</u>: $\|X - \mathbb{E}X\|_{L^2} \leq \|X\|_{L^2}$ (Check this!)

Lemma 2.6.8 (Centering):

If X is a sub-gaussian random variable,

then $X - \mathbb{E}X$ is also sub-gaussian and

$$\|X - \mathbb{E}X\|_{\psi_2} \leq C\|X\|_{\psi_2},$$

where C is an absolute constant.

<u>Pf</u> Since $\|\cdot\|_{\psi_2}$ is a norm (Ex. 2.5.7)
by the triangle inequality,

$$\|X - \mathbb{E}X\|_{\psi_2} \leq \|X\|_{\psi_2} + \underbrace{\|\mathbb{E}X\|_{\psi_2}}_{\text{constant and thus bounded}}$$

By discussion on bounded random variables,

$$\|\mathbb{E}X\|_{\psi_2} \leq \|X\|_{\infty} = |\mathbb{E}X| \leq \mathbb{E}|X| \overset{\text{prop. 2.5.2}}{\leq} \|X\|_{\psi_2}.$$

Q.E.D.

Sub - Exponential Distributions

I. Motivation:

Recall that the class of sub-gaussian distributions is wide and satisfies nice concentration inequalities. In particular,

$$\mathbb{P}(|X-\mu|>t) = \mathbb{P}(e^{(X-\mu)^2} > e^{t^2}) \xrightarrow{\text{Hoeffding}} e^{-t^2}\,\mathbb{E}(e^{(X-\mu)^2})$$

$\mu := \mathbb{E}X$

On the other hand,

① $$\mathbb{P}(|X-\mu|>t) = \mathbb{P}(e^{|X-\mu|} > e^{t}) \leq e^{-t}\,\mathbb{E}\, e^{|X-\mu|}$$

Q: Are there distributions generating the above concentration inequality?

A: Sub - exponential distributions

② Consider $g = (g_1, \cdots, g_N)$ be a random vector in $\mathbb{R}^N$, g_i are independent $N(0,1)$ random variables.
It is useful to have a concentration inequality for

$$\|g\|_2 := \left(\sum_{i=1}^{N} g_i^2\right)^{\frac{1}{2}}.$$

However, note that $g_i \sim N(0,1)$, but not g_i^2:

$$\mathbb{P}(g_i^2 > t) = \mathbb{P}(|g_i| > \sqrt{t}) \lesssim e^{-\frac{t}{2}}.$$

The tails of g_i^2 resemble the exponential distribution and are strictly heavier than sub-gaussian.

Analogy of Proposition 2.5.2:

Proposition 2.7.1 (Sub-exponential properties)

Let X be a random variable. Then the following properties are equivalent; the parameter $K_i > 0$ appearing in these properties differ from each other by at most an absolute constant factor.

(i) The tails of X satisfy

$$\mathbb{P}(|X| > t) \leq 2\exp\left(-\frac{t}{K_1}\right), \quad \forall\, t \geq 0.$$

(ii) The moments of X satisfy

$$\|X\|_{L^p} = \mathbb{E}(|X|^p)^{1/p} \leq K_2\, p, \quad \forall\, p \geq 1.$$

(iii) The MGF of $|X|$ satisfies

$$\mathbb{E}\exp(\lambda|X|) \leq \exp(K_3 \lambda), \quad \forall\, \lambda \text{ s.t. } 0 \leq \lambda \leq \frac{1}{K_3}$$

(iv) The MGF of $|X|$ is bounded at some point, namely

$$\mathbb{E}\exp\left(\frac{|X|}{K_4}\right) \leq 2.$$

Moreover, if $\mathbb{E}X = 0$, then (i)-(iv) are equivalent to

(v) The MGF of X satisfies

$$\mathbb{E}\, e^{\lambda X} \leq e^{K_5^2 \lambda^2}, \quad \forall\, \lambda \text{ s.t. } |\lambda| \leq \frac{1}{K_5}.$$

Remark: (MGF near origin)

Consider general X with $\mathbb{E}X=0$ and $\mathbb{E}X^2=1$. Assume X bounded.

$$\mathbb{E}\exp(\lambda X) \approx \mathbb{E}\left(1+\lambda X+\frac{\lambda^2X^2}{2}+o(\lambda^2X^2)\right) \underset{\mathbb{E}X=0,\ \mathbb{E}X^2=1}{=} 1+\frac{\lambda^2}{2} \underset{\lambda\to 0}{\approx} e^{\lambda^2/2}$$

① $X\sim N(0,1)$: exact equality

② X sub-gaussian: bound holds for all λ (Prop. 2.5.2) and characterizes sub-gaussian

③ X sub-exponential: bound holds for small λ (Prop. 2.7.1) and characterizes sub-exponential
For larger λ, no general bound can exist for sub-exponentials. (Think about this.)

__Partial Pf of Prop. 2.7.1__ (ii) $\Rightarrow$ (v)

WLOG, assume $K_2=1$. (Why?)
Expanding exponential function in a Taylor series:

$$\mathbb{E}\exp(\lambda X) = \mathbb{E}\left(1+\lambda X+\sum_{k=2}^{\infty}\frac{(\lambda X)^k}{k!}\right)$$

$$= 1+\sum_{k=2}^{\infty}\frac{\lambda^k\,\mathbb{E}X^k}{k!} \quad (\text{linearity of } \mathbb{E} \text{ and } \mathbb{E}X=0)$$

$(**)$

By Prop 2.7.1 (ii),

$$|\mathbb{E}X^k| \le \mathbb{E}|X|^k \le k^k, \ \forall k\ge 1. \quad (*1)$$

Meanwhile, $k! \ge \left(\frac{k}{e}\right)^k$ by Stirling. $(*2)$

Apply $(*1)$ and $(*2)$ to $(**)$:

$$\mathbb{E}(e^{\lambda X}) \le 1+\sum_{k=2}^{\infty}\frac{(\lambda k)^k}{(k/e)^k} = 1+\sum_{k=2}^{\infty}(e\lambda)^k = 1+\frac{(e\lambda)^2}{1-e\lambda} \quad \text{for } |e\lambda|<1$$

If $|e\lambda|\le\frac{1}{2}$, then $1+\frac{(e\lambda)^2}{1-e\lambda}\le 1+2(e\lambda)^2$

To conclude, $\mathbb{E}\,e^{\lambda X} \le e^{2e^2\lambda^2}$, $\forall\lambda$ with $|\lambda|\le\frac{1}{2e}$.

II. Definition and Relationship with Sub-Gaussians

1) Def. 2.7.5:

A random variable X that satisfies one of the equivalent properties (i)-(iv) in Proposition 2.7 is called sub-exponential random variable. The sub-exponential norm of X, denoted by $\|X\|_{\psi_1}$, is defined to be

$$\|X\|_{\psi_1} := \inf \{ t > 0 : \mathbb{E} e^{\frac{|X|}{t}} \leq 2 \}.$$

2) Relation with sub-gaussian

① Sub-gaussian $\Rightarrow$ Sub-exponential

② square of sub-gaussian $\Rightarrow$ Sub-exponential

The reverse is also true:

Lemma 2.7.6:

A random variable X is sub-gaussian if and only if X^2 is sub-exponential. Moreover,

$$\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2.$$

Pf: Follows easily from the definitions:

$$\|X^2\|_{\psi_1} := \inf \{ K : \mathbb{E} e^{\frac{X^2}{K}} \leq 2 \}$$

$$\|X\|_{\psi_2} := \inf \{ L : \mathbb{E} e^{\frac{X^2}{L^2}} \leq 2 \}.$$

More generally, the product of sub-gaussians is sub-exponential.

<u>Lemma 2.7.7</u>: Let X and Y be sub-gaussian random variables. Then $X \cdot Y$ is sub-exponential. Moreover,

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2} \|Y\|_{\psi_2}.$$

Pf W.L.O.G. assume $\|X\|_{\psi_2} = \|Y\|_{\psi_2} = 1$. (Why?)
It suffices to show

if $\mathbb{E}(e^{X^2}), \mathbb{E}(e^{Y^2}) \leq 2$,
then $\mathbb{E}(e^{|XY|}) \leq 2$.

Consider $e^{|XY|} \leq e^{\frac{X^2+Y^2}{2}}$ (Why?)

Then
$$\mathbb{E}(e^{|XY|}) \leq \mathbb{E}(e^{\frac{X^2+Y^2}{2}}) = \mathbb{E}(e^{\frac{X^2}{2}} e^{\frac{Y^2}{2}})$$
$$\leq \frac{1}{2} \mathbb{E}(e^{X^2} e^{Y^2})$$
$$\leq 2.$$

3) Other examples of sub-exponential:

Exponential, Poisson (Check!)

4) Centering.
An analogue of Lemma 2.6.8:
If X is sub-exponential, then $\|X - \mathbb{E}X\|_{\psi_1} \leq C\|X\|_{\psi_1}$.

IV. Concentration Inequality : Bernstein's Inequality

Theorem 2.8.1 (Bernstein's Inequality)

Let $X_1, \dots, X_N$ be independent mean-zero sub-exponential random variable. Then for every $t \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} X_i\right| \geq t\right) \leq 2\exp\left(-c\min\left(\frac{t^2}{\sum_{i=1}^{N}\|X_i\|_{\psi_1}^2}, \frac{t}{\max_i \|X_i\|_{\psi_1}}\right)\right),$$

where $c > 0$ is an absolute constant.

Pf: $\forall \lambda > 0$,

$$\mathbb{P}\left(\sum_{i=1}^{N} X_i > t\right) \leq e^{-\lambda t}\, \mathbb{E}\exp\left(\lambda \sum_{i=1}^{N} X_i\right)$$

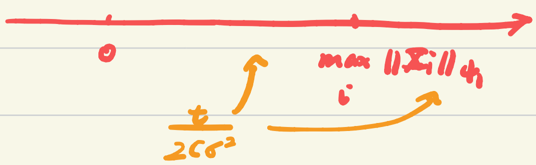
$$= e^{-\lambda t} \prod_{i=1}^{N} \mathbb{E}\, e^{\lambda X_i} \qquad \text{by independence}$$

$$\leq e^{-\lambda t} \prod_{i=1}^{N} e^{c\lambda^2 \|X_i\|_{\psi_1}^2} \qquad \text{if } \lambda \leq \frac{c}{\|X_i\|_{\psi_1}}, \forall i \text{ by Prop 2.7.1 (v)}$$

$$= e^{-\lambda t + c\lambda^2 \underbrace{\sum_{i=1}^{N} \|X_i\|_{\psi_1}^2}_{=:\sigma^2}}$$

We can then minimize λ and let

$$\lambda := \min\left(\frac{t}{2C\sigma^2}, \frac{c}{\max_i \|X_i\|_{\psi_1}}\right)$$



which yields

$$\mathbb{P}\left(\sum_{i=1}^{N} X_i > t\right) \leq \exp\left(-c\min\left(\frac{t^2}{4C\sigma^2}, \frac{ct}{2\max \|X_i\|_{\psi_1}}\right)\right).$$

A more convenient form:

Theorem 2.8.2 (Bernstein's inequality)

Let $X_1, \cdots, X_N$ be independent mean-zero, sub-exponential random variables, and let $a = (a_1, \cdots, a_N) \in \mathbb{R}^N$.

Then, for every $t \geq 0$,

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

where $K := \max_i \|X_i\|_{\psi_1}$.

$$\Rightarrow \|a\|_2^2 = \frac{1}{N} = \|a\|_\infty$$

When $a_i := \frac{1}{N}$, we obtain a quantified law of large numbers.

Corollary 2.8.3

Let $X_1, \cdots, X_N$ be independent mean-zero, sub-exponential random variables. Then for every $t \geq 0$,

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

where $K := \max_i \|X_i\|_{\psi_1}$.