36-710: Advanced Statistical Theory **Spring 2019**

Lecture 5: February 5

Lecturer: Alessandro Rinaldo *Scribes: Aleksandr Podkopaev*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

In this lecture we continue analyzing the class of Sub-Exponential random variables, derive properties of the tail behavior for this class. We introduce Orlicz norms as a more generic way of dealing with defined before classes of random variables.

5.1 Sub-Exponential Random variables

One can treat class of Sub-Exponential random variables as an extention of the class of Sub-Gaussian random variables.

Definition 5.1 (Sub-Exponential random variable) *Centered random variable $X \in SE(\nu^2, \alpha)$ with parameters $\nu, \alpha > 0$ if:*

$$\mathbb{E}e^{\lambda X} \leq e^{\frac{\lambda^2\nu^2}{2}},\ \forall \lambda : |\lambda| < \frac{1}{\alpha}$$

Observe that the moments of X are still well defined since they can be found as the derivative of the MGF (moment generating function) at zero. Additionally, one can say informally that class of Sub-Gaussian random variable can be viewed as the class of Sub-Exponential random variables when alpha goes to zero.

Example: If $Z \in \mathcal{N}(0,1)$, then $Z^2 \in SE(4,4)$.

5.1.1 Tail behavior for Sub-Exponential Random Variables

Theorem 5.2 (Tail bound for Sub-Exponential Random Variables) *Let $X \in SE(\nu^2, \alpha)$. Then:*

$$\mathbb{P}(|X-\mu| \geq t) \leq \begin{cases} 2e^{-t^2/(2\nu^2)}, & 0 < t \leq \frac{\nu^2}{\alpha} \text{ (\textit{Sub-Gaussian behavior})} \\ 2e^{-t/2\alpha}, & t > \frac{\nu^2}{\alpha} \end{cases}$$

It can be equivalently stated as:

$$\mathbb{P}(|X-\mu| \geq t) \leq e^{-\frac{1}{2}\min\{\frac{t^2}{\nu^2}, \frac{t}{\alpha}\}}$$

Proof:

Assume $\mu = 0$. Then repeating Chernoff argument, one obtains:

$$\mathbb{P}(X \geq t) \leq e^{-\lambda t + \frac{\lambda^2\nu^2}{2}} = e^{g(\lambda, t)},\ \forall \lambda \in (0, \frac{1}{\alpha})$$

To obtain the tightest bound one needs to find:

$$g^*(t) = \inf_{\lambda \in (0, \frac{1}{\alpha})} g(\lambda, t)$$

To do so, notice, firstly, that unconstrained minimum occurs at $\lambda^* = t/\nu^2 > 0$. Consider two cases:

1. If $\lambda^* < 1/\alpha \Leftrightarrow t \leq \frac{\nu^2}{\alpha}$, then unconstrained minimum appears to be also constrained minimum and by plugging in this value one obtains the bound describing sub-Gaussian behavior.
2. If $\lambda^* > 1/\alpha \Leftrightarrow t > \nu^2/\alpha$, then notice that the function $g(\lambda, t)$ is decreasing in λ in the interval $\lambda \in (0, \frac{1}{\alpha})$. Thus, the constrained minimum occurs at the boundary:

$$\lambda^*_{constrained} = \frac{1}{\alpha}$$

and

$$g(\lambda^*_{constrained}, t) = -\frac{t}{\alpha} + \frac{1}{2\alpha}\frac{\nu^2}{\alpha} \leq -\frac{t}{2\alpha}$$

since $\frac{\nu^2}{\alpha} \leq t$.

■

Recall that sufficient conditions for a random variable to be a Sub-Gaussian include:

- Boundedness of a random variable.
- Condition on the moments $\left(\mathbb{E}|X|^k\right)^{1/k}$

One would like to obtain a similar condition allowing unbounded random variables to behave sub-exponentially. One such condition is called Bernstein condition.

Definition 5.3 (Bernstein condition) *Let X be a random variable with mean μ and variance σ^2. Assume that $\exists b > 0$:*

$$\mathbb{E}|X - \mu|^k \leq \frac{1}{2}k!\sigma^2 b^{k-2}, k = 3, 4, \ldots$$

Then one says that X satisfies Bernstein condition.

Lemma 5.4 *If random variable X satisfies Bernstein condition with parameter b, then:*

$$\mathbb{E}e^{\lambda(X-\mu)} \leq e^{\frac{\lambda^2\sigma^2}{2}\frac{1}{1-b|\lambda|}}, \ \forall |\lambda| < \frac{1}{b}$$

Additionally, from the bound on the moment generating function one can obtain the following tail bound (also known as Bernstein inequality):

$$\mathbb{P}\left(|X - \mu| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2(\sigma^2 + bt)}\right), \forall t > 0$$

Proof: Pick $\lambda : |\lambda| < \frac{1}{b}$ (allowing interchanging summation and taking expectation) and expand the MGF in a Taylor series:

$$\mathbb{E}e^{\lambda(X-\mu)} = 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty}\frac{\mathbb{E}|X-\mu|^k}{k!}\lambda^k \leq 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2}\sum_{k=3}^{\infty}(|\lambda|b)^{k-2} =$$

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

where we used $1 + x \leq e^x$. To show the final bound, take $\lambda : |\lambda| < \frac{1}{2b}$. Then the bound becomes:

$$e^{\frac{\lambda^2\sigma^2}{2}\frac{1}{1-b|\lambda|}} \leq e^{\lambda^2\sigma^2} = e^{\frac{\lambda^2(2\sigma^2)}{2}}$$

implying that $X \in SE(2\sigma^2, 2b)$. The concentration result then follow by taking $\lambda = \frac{t}{bt+\sigma^2}$. ■

5.1.2 Composition property of Sub-Exponential random variables

Let $X_1, \ldots, X_n$ be independent random variables such that $\mathbb{E}X_i = \mu_i$ and $X_i \in SE(\nu_i^2, \alpha_i)$. Then

$$\sum_{i=1}^{n}(X_i - \mu_i) \in SE\left(\sum_{i=1}^{n}{\nu_i}^2, \max_i \alpha_i\right)$$

In particular, denote $\nu_*^2 = \sum_{i=1}^{n}{\nu_i}^2, \alpha_* = \max_i \alpha_i$. Then:

$$\mathbb{P}\left(\frac{1}{n}\Big|\sum_{i=1}^{n}(X_i - \mu_i)\Big| \geq t\right) \leq \begin{cases} 2\exp\left(-\frac{nt^2}{2\sum_{i=1}^{n}\nu_i^2}\right), & 0 < nt \leq \frac{\nu_*^2}{\alpha_*} \\ 2\exp\left(-\frac{nt}{2\alpha_*}\right), & \text{otherwise} \end{cases}$$

or, equivalently,

$$\mathbb{P}\left(\frac{1}{n}\Big|\sum_{i=1}^{n}(X_i - \mu_i)\Big| \geq t\right) \leq \exp\left(-\frac{n}{2}\min\{\frac{t^2}{\nu_*^2}, \frac{t}{\alpha_*}\}\right)$$

Remark: Notice that that the range over which one obtains sub-Gaussian tail behavior gets smaller.

Example: Let $X \sim \chi_n^2$, i.e. $X = \sum_{i=1}^{n} Z_i^2$ where $Z_i \sim \mathcal{N}(0,1)$. Then $X \in SE(4n, 4)$ and thus,

$$\mathbb{P}\left(\Big|\frac{1}{n}\sum_{i=1}^{n}Z_i^2 - 1\Big| \geq t\right) \leq \begin{cases} 2\exp\left(-\frac{nt^2}{8}\right), & t \in (0,1) \\ 2\exp\left(-\frac{nt}{8}\right), & t \geq 1 \end{cases}$$

5.2 Orlicz norms

Everything said so far can be handled in more general way using Orlicz norms.

Definition 5.5 (ψ-Orlicz norm) *Let function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ satisfy the following properties:*

- *$\psi(x)$ is strictly increasing function*
- *$\psi(x)$ is a convex function*
- *$\psi(0) = 0$*

Then the ψ-Orlicz norm of a random variable X is defined as:

$$\|X\|_\psi = \inf\{t > 0 : \mathbb{E}\psi\left(\frac{|X|}{t}\right) \leq 1\}$$

Let us look at several examples:

1. Let $\psi(x) = x^p,\ p \geq 1$. Then:
$$\|X\|_\psi = \|X\|_p = (\mathbb{E}|X|^p)^{\frac{1}{p}}$$
2. Let $\psi_p(x) = e^{x^p} - 1, p \geq 1$. The corresponding Orlicz has the following properties:
   (a) $p = 1$: then $\|X\|_{\psi_1} < \infty$ is equivalent to X belonging to the class of Sub-Exponential random variables
   (b) $p = 2$: then $\|X\|_{\psi_2} < \infty$ is equivalent to X belonging to the class of Sub-Gaussian random variables

It is easy to show that:
$$\|X^2\|_{\psi_1} = (\|X\|_{\psi_2})^2, \quad \|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2}$$
Using Orlicz norms allows to straightforwardly implies the following facts:

1. squared Sub-Gaussian random variable is Sub-Exponential.
2. product of two Sub-Gaussian random variables is Sub-Exponential.

Lemma 5.6 (Concentraiton of a sub-gaussian random vector) *Let $X = (X_1, \ldots, X_d)^\top \in \mathbb{R}^d$ be such that: $\mathbb{E}X_i = 0, \mathbb{V}(X_i) = 1$ and assume that $X_i \in SG(\sigma^2)$. Then we can show that $\|X\|_2$ concentrates around $\sqrt{d}$.*

Proof: Consider:
$$\|X\|_2^2 = \sum_{i=1}^{n} X_i^2$$
Then $X_i^2 - 1 \in SE(\nu^2, \alpha)$ where both parameters are determined by σ^2. Thus,
$$\mathbb{P}\left(\Big|\frac{\|X\|^2}{d} - 1\Big| \geq t\right) \leq 2\exp\left(-\frac{d}{2}\min\{\frac{t^2}{\nu^2}, \frac{t}{\alpha}\}\right),\ \forall t > 0$$
We will need to use the following fact: fix $c > 0$. Then for any numbers $z > 0$:
$$|z - 1| \geq c \overset{implies}{\Longrightarrow} z^2 - 1 \geq \max\{c, c^2\}$$
Using this fact allows to conclude that:
$$\mathbb{P}\left(\Big|\frac{\|X\|}{\sqrt{d}} - 1\Big| \geq u\right) = \mathbb{P}\left(\Big|\frac{\|X\|^2}{d} - 1\Big| \geq \max\{u, u^2\}\right) \leq 2\exp\left(-\frac{du^2}{2C}\right)$$

■

5.3 Hoeffding vs. Bernstein

One would like to compare two type of bounds/inequalities: Hoeffding's and Bernstein's. Denote $\mu = \mathbb{E}X$ and$\sigma^2 = \mathbb{V}(X)$. Assume that $|X - \mu| \leq b$ a.e. Then:
$$\mathbb{P}(|X - \mu| \geq t) \leq \begin{cases} 2\exp\left(-\frac{t^2}{2b^2}\right), & \text{Hoeffding} \\ 2\exp\left(-\frac{t^2}{2(\sigma^2 + bt)}\right), & \text{Bernstein} \end{cases}$$

For small t (meaning $bt \ll \sigma^2$) Bernstein's inequality gives rise to a bound of the order:

$$\mathbb{P}(|X - \mu| \geq t) \leq 2e^{-\frac{t^2}{c\sigma^2}}$$

while Hoeffding's gives:

$$\mathbb{P}(|X - \mu| \geq t) \leq 2e^{-\frac{t^2}{cb^2}}$$

But $\sigma^2 \leq b^2$ and, thus, Bernstein's bound is better / tighter.

Theorem 5.7 (Classic Bernstein inequality) *Let $X_1, \ldots, X_n$ be independent random variables such that $|X_i - \mathbb{E}X_i| \leq b$ a.e. and $\max_i \mathbb{V}(X_i) \leq \sigma^2$. Then:*

$$\mathbb{P}\left(\Big|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\Big| \geq t\right) \leq 2\exp\left(-\frac{nt^2}{2(\sigma^2 + \frac{bt}{3})}\right)$$

Theorem 5.8 (Laurent-Massart bounds for χ^2) *Let $Z_1, \ldots, Z_d \sim \mathcal{N}(0,1)$ and $a = (a_1, \ldots, a_d)$ with $a_i \geq 0, \forall i \in \{1, \ldots, n\}$. Let $X = \sum_{i=1}^{n} a_i(X_i^2 - 1)$. Then for right-tail behavior is described by:*

$$\mathbb{P}(X \geq 2\|a\|\sqrt{t} + 2\|a\|_\infty t) \leq e^{-t},\ \forall t > 0$$

and for left-tail behavior:

$$\mathbb{P}(X \leq -2\|a\|\sqrt{t}) \leq e^{-t},\ \forall t > 0$$