MATH888: High-dimensional probability and statistics	Fall 2021

Lecture 8 — September 24, 2021

Sebastien Roch, UW-Madison *Scribe: Mingchen Ma*

1 Review of the Last Lecture

In the last lecture, we introduce Sub-Exponential random variables. Let X be a random variable. We define its ψ_1 norm to be

$$\|X\|_{\psi_1} := \inf\{t \geq 0 \mid \mathbb{E}\exp(|X|/t) \leq 2\}.$$

If $\|X\|_{\psi_1} < \infty$, then we say X is a Sub-Exponential random variable. An important fact for Sub-Exponential random variable is that if X is a Sub-Exponential random variable, then for every $t \geq 0$, we have

$$\mathbb{P}(|X| \geq t) \leq 2\exp(-ct/\|X\|_{\psi_1}).$$

This implies that a Sub-Exponential variable decays slower than a Sub-Gaussian variable.

2 This Lecture

In this lecture, we introduce more properties of Sub-Exponential variables and Sub-Gaussian variables. The topics of this lecture includes the relation between Sub-Exponential variables and Sub-Gaussian variables, Bernstein's inequality and concentration of the norm.

2.1 Sub-Exponential and Sub-Gaussian

Let X be a random variable. We will see that X is a Sub-Gaussian if and only if X^2 is a Sub-Exponential.

Lemma 1. *(Lemma 2.7.6 in [1]) Let X be a random variable. Then $\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$.*

Proof

$$\begin{aligned}\|X^2\|_{\psi_1} &= \inf\{t \geq 0 \mid \mathbb{E}\exp(X^2/t) \leq 2\} \\ &= \inf\{(\sqrt{t})^2 \mid \mathbb{E}\exp(X/\sqrt{t})^2 \leq 2\} \\ &= \|X\|_{\psi_2}^2.\end{aligned}$$

◇

2.2 Bernstein's inequality

Hoeffding's inequality gives a concentration result for Sub-Gaussian random variables. For Sub-Exponential random variables, Bernstein's inequality gives a similar concentration result.

Theorem 2. *(Bernstein's inequality) If $X_1, \dots, X_n$ are independent zero-mean Sub-Exponential random variables, then for every $t \geq 0$,*

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

where c is a constant.

Before we give the proof of the above theorem, we first give an example for the theorem.

Example 1. *Suppose that for $i \in [N]$, we have $\|X_i\|_{\psi_1} = 1$, then*

$$\mathbb{P}(|\sum_{i=1}^{N} X_i| \geq t) \leq \begin{cases} \exp(-\frac{ct^2}{N}) \text{ if } t \leq N \\ \exp(-ct) \text{ if } t > N. \end{cases}$$

The above example shows that if t is small, the behavior of $\sum_{i=1}^{N} X_i$ is like Gaussian, because of the central limit theorem. On the other hand, if t is large, the behavior of $\sum_{i=1}^{N} X_i$ is like a Sub-Exponential.

Now, we present the proof of the Bernstein's inequality, which was not covered in the lecture.

Proof (Proof of Theorem 2.8.1 in [1]) To deal with the sum of independent random variables, we consider the MGF. Let $S := \sum_{i=1}^{N} X_i$. Let $\lambda > 0$ be a parameter. Then we have

$$\begin{aligned}\mathbb{P}(S \geq t) &= \mathbb{P}(\exp(\lambda S) \geq \exp(\lambda t)) \\ &\leq \exp(-\lambda t)\mathbb{E}\exp(\lambda S) \\ &= \exp(-\lambda t)\prod_{i=1}^{N}\mathbb{E}\exp(\lambda X_i)\end{aligned}$$

The inequality follows by Markov's inequality. Using the property of Sub-Exponential, we know that there is a constant c such that when $\lambda \leq 1/c\max_{i\in[N]}\|X_i\|_{\psi_1}$, we have $\mathbb{E}\exp(\lambda X_i) \leq \exp(c\lambda^2\|X_i\|_{\psi_1}^2)$. So we get

$$\mathbb{P}(S \geq t) \leq \exp(c\lambda^2\sum_{i=1}^{N}\|X_i\|_{\psi_1}^2 - \lambda t)$$

By choosing the optimal $\lambda := \min(\frac{t}{2c\sum_{i=1}^{N}\|X_i\|_{\psi_1}^2}, \frac{1}{c\max_{i\in[N]}\|X_i\|_{\psi_1}})$, we can see

$$\mathbb{P}(S \geq t) \leq \exp(-c\min(\frac{t^2}{\sum_{i=1}^{N}\|X_i\|_{\psi_1}^2}, \frac{t}{\max_{i\in[N]}\|X_i\|_{\psi_1}})).$$

Using a similar trick for $-S$, we can finish the proof.

◇

2.3 Concentration of the Norm

In this part, we apply the property of Sub-Gaussian to obtain a result for a high dimensional random vector.

Theorem 3. *Let $\vec{X} = (X_1, \ldots, X_n) \in \mathbb{R}^n$, where X_i are independent Sub-Gaussian with $\mathbb{E}X_i^2 = 1$. Then for every $t \geq 0$, we have*

$$\mathbb{P}(|\|\vec{X}\|_2 - \sqrt{n}| \geq t) \leq 2\exp(-ct^2/K^4),$$

where $K = \max_{i\in[n]} \|X_i\|_{\psi_2}$.

We notice that $\mathbb{E}\|\vec{X}\|_2^2 = \sum_{i=1}^n X_i^2 = n$. Although this doesn't implies $\mathbb{E}\|\vec{X}\|_2 = \sqrt{n}$, we can expect that $\mathbb{E}\|\vec{X}\|_2$ is close to $\sqrt{n}$.

Proof We first show that $K \geq 1$. By the definition of ψ_2 norm, we have

$$2 \geq \mathbb{E}\exp(X_i^2/\|X_i\|_{\psi_2}^2) \geq 1 + \mathbb{E}X_i^2/\|X_i\|_{\psi_2}^2 = 1 + 1/\|X_i\|_{\psi_2}^2.$$

This implies $\|X_i\|_{\psi_2}^2 \geq 1$ for every i.

Notice that $\frac{1}{n}\|\vec{X}\|_2^2 - 1 = \frac{1}{n}\sum_{i=1}^n (X_i^2 - 1)$. We show that $\|X_i^2 - 1\|_{\psi_1} \lesssim K^2$ for every i. This is because

$$\|X_i^2 - 1\|_{\psi_1} \lesssim \|X_i^2\|_{\psi_1} = \|X_i\|_{\psi_2}^2 \leq K^2,$$

where the first inequality follows by centering and the last inequality holds because $K \geq 1$. So there is some $c > 0$ such that $\|X_i^2 - 1\|_{\psi_1} \leq cK^2$. Now we use Bernstein's inequality and get for every $u > 0$,

$$\mathbb{P}(|\frac{1}{n}\|\vec{X}\|_2^2 - 1| > u) \leq 2\exp(-\frac{cn}{K^4}\min(u^2, u)).$$

A simple observation is that if $z, \delta \geq 0$, then $|z-1| \geq \delta$ implies $|z^2 - 1| \geq \max\{\delta, \delta^2\}$. To see this, we only need to show $|z+1| \geq \max\{1, \delta\}$. From $|z-1| \geq \delta$, we know that either $z \geq 1+\delta$ or $0 \leq z \leq 1-\delta$. In the first case $|z+1| = 2+\delta$, while in the second case $|z+1| = 2-\delta$ and $0 < \delta < 1$. So in both cases, we have $|z+1| \geq \max\{1, \delta\}$.

Now, we take $\delta = t/\sqrt{n}$. Then we get

$$\begin{aligned}
\mathbb{P}(|\|\vec{X}\|_2 - \sqrt{n}| \geq t) &= \mathbb{P}(|\frac{1}{\sqrt{n}}\|\vec{X}\|_2 - 1| \geq \delta) \\
&\leq \mathbb{P}(|\frac{1}{n}\|\vec{X}\|_2^2 - 1| > \max\{\delta, \delta^2\}) \\
&\leq 2\exp(-\frac{cn}{K^4}\min(u^2, u)),
\end{aligned}$$

where $u = \max\{\delta, \delta^2\}$. Notice that $\min(u^2, u) = \delta^2$ always holds true. To see this, if $\delta \geq 1$, then $u = \delta^2$ and $\min(u^2, u) = u = \delta^2$. If $\delta < 1$, then $u = \delta$ and $\min(u^2, u) = u^2 = \delta^2$. Combine the above results, we finally get

$$\mathbb{P}(|\|\vec{X}\|_2 - \sqrt{n}| \geq t) \leq 2\exp(-ct^2/K^4).$$

$\diamond$

References

[1] R.Vershynin, *High-dimensional probability: An introduction with applications in data science*, Cambridge university press, 2008.