MATH888: High-dimensional probability and statistics Fall 2021

Lecture 7 — September 22, 2021

Sebastien Roch, UW-Madison *Scribe: Nikos Zarifis*

1 Last time

In the last lecture we talked about the sub-Gaussian Random variables. Recall that a random variable X is sub-Gaussian variable if its sub-Gaussian norm is bounded, i.e.,

$$\|X\|_{\psi_2} < \infty ,$$

where

$$\|X\|_{\psi_2} = \inf\{t \geq 0 : \mathbf{E}[\exp(X^2/t^2)] \leq 2\} .$$

If $\|X\|_{\psi_2} \leq K$, there exists an absolute constant $C > 0$ such that the following properties are equivalent (see [1])

1. $\mathbf{Pr}[|X| \geq t] \leq 2\exp(-t^2/(CK^2))$ for all $t > 0$.
2. $\|X\|_p \leq CK\sqrt{p}$ for all $p \geq 1$ (recall that $\|X\|_p = (\mathbf{E}[X^p])^{1/p}$).
3. If $\mathbf{E}[X] = 0$, $\mathbf{E}[\exp(\lambda X)] \leq \exp(C\lambda^2 K^2)$ for all λ.

Moreover, the triangle inequality holds w.r.t. the $\|\cdot\|_{\psi_2}$, i.e.,

$$\|X + Y\|_{\psi_2} \leq \|X\|_{\psi_2} + \|Y\|_{\psi_2} .$$

We denote that $a \lesssim b$ for $a \leq Cb$, for some absolute constant $C > 0$.

2 Hoeffding's Inequality

In this lecture we are going to see the Hoeffding's Inequality for sub-Gaussian random variables. First, we prove the following lemma:

Lemma 1 (Centering lemma 2.6.8 in [1]). *It holds that*

$$\|X - \mathbf{E}[X]\|_{\psi_2} \lesssim \|X\|_{\psi_2} .$$

Proof. From triangle inequality, it holds that

$$\|X - \mathbf{E}[X]\|_{\psi_2} \leq \|X\|_{\psi_2} + \|\mathbf{E}[X]\|_{\psi_2} \lesssim \|X\|_{\psi_2} + |\mathbf{E}[X]| ,$$

where we used that $\|\mathbf{E}[X]\|_{\psi_2} \leq C|\mathbf{E}[X]|$, for some $C > 0$, from the definition of the sub-Gaussian norm (for $t = \mathbf{E}[X]/\sqrt{\ln 2}$). Then using Jensen's inequality we have that $|\mathbf{E}[X]| \leq \|X\|_1$ and from Property 2, we have that $\|X\|_1 \leq C\|X\|_{\psi_2}$, for some $C > 0$. Therefore, we have that

$$\|X - \mathbf{E}[X]\|_{\psi_2} \lesssim \|X\|_{\psi_2} + |\mathbf{E}[X]| \lesssim \|X\|_{\psi_2} .$$

□

Next, we prove the General Hoeffding's inequality.

Theorem 2 (Thm 2.6.2 in [1])**.** *Let $X_1, X_2, \ldots, X_N$ be independent mean zero, sub-Gaussian random variables. Then there exists an absolute constant $c > 0$, so that for all $t \geq 0$, it holds*

$$\mathbf{Pr}\left[\left|\sum_{i=1}^{N} X_i\right| \geq t\right] \leq 2\exp\left(-\frac{ct^2}{\sum_{i=1}^{N} \|X_i\|_{\psi_2}^2}\right) .$$

This is called General Hoeffding's inequality and generalizes the Hoeffding's inequality which assumes that the random variables have bounded variance σ^2.

Proof. First, from triangle inequality it holds that

$$\|\sum_{i=1}^{N} X_i\|_{\psi_2} \leq \sum_{i=1}^{N} \|X_i\|_{\psi_2} \leq \infty ,$$

therefore $\sum_{i=1}^{N} X_i$ is a sub-Gaussian random variable. It suffices to show that the

$$\|\sum_{i=1}^{N} X_i\|_{\psi_2}^2 \lesssim \sum_{i=1}^{N} \|X_i\|_{\psi_2}^2 .$$

One could try the triangle inequality, but this would give that $\| \sum_{i=1}^{N} X_i\|_{\psi_2}^2 \leq (\sum_{i=1}^{N} \|X_i\|_{\psi_2})^2 \sim N \sum_{i=1}^{N} \|X_i\|_{\psi_2}^2$ which is very loose, therefore we need to find a different way to prove this.

To prove this, we need to use the independence of the random variables X_i. From the assumption that the random variables are mean zero, the sum is also from linearity, hence, using Property 3, we have that for some absolute constant $C > 0$ that

$$\mathbf{E}[\exp(\lambda \sum_{i=1}^{N} X_i)] = \prod_{i=1}^{N} \mathbf{E}[\exp(\lambda X_i)] \leq \prod_{i=1}^{N} \exp(C\lambda \|X_i\|_{\psi_2}) = \exp(C\lambda \sum_{i=1}^{N} \|X_i\|_{\psi_2}^2) ,$$

where we used the independence to split the expectation into a product. The proof follows by applying the Property 1. ☐

3 Sub-Exponential Random Variables

To understand the norm of a vector with sub-Gaussian (X) coordinates, we need to understand the square of a sub-Gaussian (X^2). Observe that

$$\mathbf{Pr}[X^2 \geq t] = \mathbf{Pr}[|X| \geq \sqrt{t}] \leq 2\exp(-\frac{C(\sqrt{t})^2}{K^2}) = 2\exp(-\frac{Ct}{K^2}) ,$$

where we used the Property 1. Observe that the decay is smaller than before, these random variables are called sub-exponential.

Definition 3. *We call a random variable X, sub-exponential if*

$$\|X\|_{\psi_1} < \infty \ ,$$

where

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

Similar with the sub-Gaussian random variables, we have some equivalence properties for sub-exponential random variables. We mark with red the main differences with sub-Gaussian random variables.

Proposition 4. *If $\|X\|_{\psi_1} \leq K$, i.e., X is sub-exponential then there exists an absolute constant $C > 0$ such that the following properties are equivalent,*

1. $\mathbf{Pr}[|X| \geq t] \leq 2\exp(-t/(CK))$ *for all $t > 0$.*
2. $\|X\|_p \leq CKp$ *for all $p \geq 1$.*
3. *If $\mathbf{E}[X] = 0$, then $\mathbf{E}[\exp(\lambda X)] \leq \exp(C\lambda^2 K^2)$ for all $|\lambda| \leq \frac{1}{CK}$.*

We have that triangle inequality holds w.r.t. the $\|\cdot\|_{\psi_1}$. We have the Bernstein's inequality.

Theorem 5 (Thm 2.8.1. in [1])**.** *Let $X_1, X_2, \ldots, X_N$ be independent mean zero, sub-exponential random variables. Then there exists an absolute constant $c > 0$, so that for all $t \geq 0$, it holds*

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

Proof. The proof is similar to the Hoeffding's inequality. To prove this, we need to use the independence of the random variables X_i. We use Property 3, we have that

$$\mathbf{E}[\exp(\lambda\sum_{i=1}^{N} X_i)] = \prod_{i=1}^{N}\mathbf{E}[\exp(\lambda X_i)] \leq \prod_{i=1}^{N}\exp(C\lambda\|X_i\|_{\psi_1}) = \exp(C\lambda\sum_{i=1}^{N}\|X_i\|_{\psi_1}^2) \ ,$$

for all $\lambda \leq \frac{1}{C\max_i \|X_i\|_{\psi_1}}$, where we used the independence to split the expectation into a product. Let $\sigma^2 = \sum_{i=1}^{N}\|X_i\|_{\psi_1}^2$. From Markov's inequality we have that

$$\mathbf{Pr}\left[\sum_{i=1}^{N} X_i \geq t\right] = \mathbf{Pr}\left[\exp\left(\lambda\sum_{i=1}^{N} X_i\right) \geq \exp(\lambda t)\right] \leq \exp(-\lambda t)\,\mathbf{E}[\exp(\lambda\sum_{i=1}^{N} X_i)] \leq \exp(-\lambda t + C\lambda^2\sigma^2) \ ,$$

then optimizing with respect of λ we have that $\lambda = \min\left(\frac{t}{2C\sigma^2}, \frac{1}{C\max_i \|X_i\|_{\psi_1}}\right)$ which gives the result, to complete the proof apply Markov's inequality in the $-\sum_{i=1}^{N} X_i$. □

There is another form of Bernstein's inequality, when the random variables are bounded.

References

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