MATH888: High-dimensional probability and statistics Fall 2021

Lecture 7 — September 22, 2021

Sebastien Roch, UW-Madison *Scribe: Sixu Li*

1 Overview

In the last lecture we defined the sub-gaussian distributions and talked about a few useful sub-gaussian properties.

In this lecture we are going to discuss the general Hoeffding's inequality, sub-exponential distributions, and Bernstein's inequality.

2 Review of Sub-Gaussian Distributions

In this section, we recall some definitions and properties of sub-gaussian variables and norm.

- (Sub-gaussian norm) The sub-gaussian norm of X, denoted $\|X\|_{\psi_2}$ is defined as

$$\|X\|_{\psi_2} := \inf\{t > 0 : \mathbb{E}\exp(X^2/t^2) \leq 2\} \tag{1}$$

  if $\|X\|_{\psi_2} < +\infty$, we say X is sub-gaussian.

- (Restate Proposition 2.5.2 in [1] in terms of $\|\cdot\|_{\psi_2}$) $\|X\|_{\psi_2} \leq K$ is equivalent to (up to constants, see more details in [1] page 28)

  (i) The tails of X satisfy

$$\mathbb{P}(|X| \geq t) \leq 2\exp(-\frac{ct^2}{K^2}), \quad \text{for all } t \geq 0 \tag{2}$$

  (ii) The moments of X satisfy

$$\|X\|_p \leq CK\sqrt{p}, \quad \text{for all } p \geq 1 \tag{3}$$

  (v) If $\mathbb{E}X = 0$, the MGF of X satisfies

$$\mathbb{E}\exp(\lambda X) \leq \exp(C\lambda^2 K^2), \quad \text{for all } \lambda \in \mathbb{R} \tag{4}$$

- (Triangle inequality) For any two sub-gaussian random variables X and Y,

$$\|X + Y\|_{\psi_2} \leq \|X\|_{\psi_2} + \|Y\|_{\psi_2} \tag{5}$$

3 Main Section

3.1 Centering

Lemma 1 (Centering lemma (Lemma 2.6.8 in [1])). *If X is a sub-gaussian random variable, then $X - \mathbb{E}X$ is also a sub-gaussian and,*

$$\|X - \mathbb{E}X\|_{\psi_2} \lesssim \|X\|_{\psi_2}{}^{1} \tag{6}$$

Proof. By the triangle inequality (5), we get

$$\|X - \mathbb{E}X\|_{\psi_2} \leq \|X\|_{\psi_2} + \|\mathbb{E}X\|_{\psi_2} \tag{7}$$

Also note that, by the definition of sub-gaussian norm, we have $\|\mathbb{E}X\|_{\psi_2} \lesssim |\mathbb{E}X|$ (because $\mathbb{E}X$ is a constant), then

$$\begin{aligned} \|\mathbb{E}X\|_{\psi_2} &\lesssim |\mathbb{E}X| \\ &\lesssim \|X\|_1 \quad \text{(by Jensen's inequality)} \\ &\lesssim \|X\|_{\psi_2} \quad \text{(using (3) with } p = 1) \end{aligned} \tag{8}$$

Substituting this into (7), we complete the proof. □

3.2 Hoeffding's inequality

Here, we consider the concentration inequality to the general *sub-gaussian* distributions.

Theorem 2 (General Hoeffding's inequality (Theorem 2.6.2 in [1])). *Let $X_1, \cdots, X_N$ be independent, mean zero, sub-gaussian random variables. Then, for every $t \geq 0$, we have*

$$\mathbb{P}\Big\{|\sum_{i=1}^{N} X_i| \geq t\Big\} \leq 2\exp\Big(-\frac{ct^2}{\sum_{i=1}^{N} \|X_i\|_{\psi_2}^2}\Big) \tag{9}$$

Observations: It is sufficient to show that

$$\|\sum_{i=1}^{N} X_i\|_{\psi_2}^2 \lesssim \sum_{i=1}^{N} \|X_i\|_{\psi_2}^2 \tag{10}$$

But note that triangle inequality (5) only gives us

$$\|\sum_{i=1}^{N} X_i\|_{\psi_2}^2 \leq (\sum_{i=1}^{N} \|X_i\|_{\psi_2})^2, \tag{11}$$

which is not good enough. Therefore, we consider to use the independent assumption and the moment generating function of the sum to prove it.

[1] The notation $a \lesssim b$ means that $a \leq Cb$ where is C is some absolute constant.

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

$$\begin{aligned}
\mathbb{E}\exp(\lambda \sum_{i=1}^{N} X_i) &= \prod_{i=1}^{N} \mathbb{E}\exp(\lambda X_i) \quad \text{(by independence)} \\
&\leq \prod_{i=1}^{N} \exp(C\lambda^2 \|X_i\|_{\psi_2}^2) \quad \text{(by sub-gaussian property (4))} \\
&= \exp(\lambda^2 K^2), \quad (\text{where } K^2 := C \sum_{i=1}^{N} \|X_i\|_{\psi_2}^2)
\end{aligned} \tag{12}$$

Recall that the bound on MGF we just proved characterizes sub-gaussian distribution (sub-gaussian property (4)), which implies $\sum_{i=1}^{N} X_i$ is sub-gaussian and $\|\sum_{i=1}^{N} X_i\|_{\psi_2}^2 \lesssim \sum_{i=1}^{N} \|X_i\|_{\psi_2}^2$.

□

3.3 Sub-exponential distributions

Motivations: To understand the norm of a vector with sub-gaussian coordinate, we need to understand the square of a sub-gaussian. For example, considering X is sub-gaussian and $Y := X^2$, then for $\forall t \geq 0$,

$$\mathbb{P}(Y \geq t) = \mathbb{P}(X^2 \geq t) = \mathbb{P}(|X| \geq \sqrt{t}) \leq 2\exp(-\frac{c(\sqrt{t})^2}{\|X\|_{\psi_2}^2}) = 2\exp(-\frac{ct}{\|X\|_{\psi_2}^2}) \tag{13}$$

Note that the tail of Y are like for the exponential distribution, and are strictly heavier than sub-gaussian. Therefore, in the following, we consider another important family of distributions, *sub-exponential* distributions, which are quite similar to the sub-gaussian distributions in terms of either definition or properties.

Definition 3 (Sub-exponential random variables)**.** *The sub-exponential norm of X, denoted $\|X\|_{\psi_1}$, is defined as*

$$\|X\|_{\psi_1} := \inf\{t > 0 : \mathbb{E}\exp(|X|/t) \leq 2\} \tag{14}$$

If $\|X\|_{\psi_1} < +\infty$, we say X is sub-exponential.

Proposition 4 (Sub-exponential properties (restate Proposition 2.7.1 in [1] in terms of $\|\cdot\|_{\psi_1}$))**.** *$\|X\|_{\psi_1} \leq K$ is equivalent to (up to constants, see more details in [1] page 32)*

(i) The tails of X satisfy

$$\mathbb{P}(|X| \geq t) \leq 2\exp(-\frac{ct}{K}), \quad \text{for all } t \geq 0 \tag{15}$$

(ii) The moments of X satisfy

$$\|X\|_p \leq CKp, \quad \text{for all } p \geq 1 \tag{16}$$

(v) If $\mathbb{E}X = 0$, the MGF of X satisfies

$$\mathbb{E}\exp(\lambda X) \leq \exp(C\lambda^2 K^2), \quad \text{for all } \lambda \text{ such that } |\lambda| \leq \frac{1}{\sqrt{C}K} \tag{17}$$

Lemma 5 (Triangle inequality)**.** *For any two sub-exponential random variables X and Y,*

$$\|X+Y\|_{\psi_1} \le \|X\|_{\psi_1} + \|Y\|_{\psi_1} \tag{18}$$

Lemma 6 (Centering lemma)**.** *If X is a sub-exponential random variable, then $X - \mathbb{E}X$ is also a sub-exponential and,*

$$\|X - \mathbb{E}X\|_{\psi_1} \lesssim \|X\|_{\psi_1} \tag{19}$$

3.4 Bernstein's inequality

Similar to the concentration inequality of sums of independent *sub-gaussian* random variables (Hoeffding's inequality), for *sub-exponential* random variables, we have

Theorem 7 (Bernstein's inequality (Theorem 2.8.1 in [1]))**.** *Let $X_1, \cdots, X_N$ be independent, mean zero, sub-exponential random variables. Then, for every $t \ge 0$, we have*

$$\mathbb{P}\big\{|\sum_{i=1}^N X_i| \ge t\big\} \le 2\exp\big[-\min\{\frac{ct^2}{\sum_{i=1}^N \|X_i\|_{\psi_1}^2}, \frac{ct}{\max_i \|X_i\|_{\psi_1}}\}\big], \tag{20}$$

where $c > 0$ is an absolute constant.

Proof. Note that

$$\begin{aligned}
\mathbb{P}(\sum_{i=1}^N X_i \ge t) &= \mathbb{P}(\lambda \sum_{i=1}^N X_i \ge \lambda t) \\
&\le e^{-\lambda t}\mathbb{E}\exp(\lambda \sum_{i=1}^N X_i) \quad \text{(by Markov inequality)} \qquad (21) \\
&= e^{-\lambda t}\prod_{i=1}^N \mathbb{E}\exp(\lambda X_i)
\end{aligned}$$

To bound the MGF of each term X_i, we use property (17) in Proposition 4. So, if λ is small enough such that,

$$|\lambda| \le \frac{c}{\max_i \|X_i\|_{\psi_1}} \tag{22}$$

then $\mathbb{E}\exp(\lambda X_i) \le \exp(C\lambda^2\|X_i\|_{\psi_1}^2)$. Then we have,

$$\mathbb{P}(\sum_{i=1}^N X_i \ge t) \le \exp(-\lambda t + C\lambda^2\sigma^2), \quad \text{where } \sigma^2 := \sum_{i=1}^N \|X_i\|_{\psi_1}^2 \tag{23}$$

Now we minimize this expression in λ w.r.t the constraint (22). The optimal choice is $\lambda = \min(\frac{t}{2C\sigma^2}, \frac{c}{\max_i \|X_i\|_{\psi_1}})$, for which we obtain

$$\mathbb{P}(\sum_{i=1}^N X_i \ge t) \le \exp\big[-\min\{\frac{t^2}{4C\sigma^2}, \frac{ct}{2\max_i \|X_i\|_{\psi_1}}\}\big] \tag{24}$$

Repeating this argument for $-X_i$ instead of X_i, we obtain the same bound for $\mathbb{P}(-\sum_{i=1}^N X_i \le t)$. A combination of these two bounds completes the proof. $\square$

References

[1] Vershynin, Roman, *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge university press, 2018,