MATH888: High-dimensional probability and statistics	Fall 2021
Lecture 8 — September 24, 2021	
Sebastien Roch, UW-Madison	*Scribe: Jeremy Johnson*

1 Overview

In the last lecture we expand our understanding of sub-exponential distributions and prove useful concentration inequalities for the sum of sub-exponential random variables and for the norm of random vectors with sub-gaussian coordinates.

2 Sub-Exponential Random Variables

We recall the class of sub-exponential random variables; a class of random variables whose tails decay slightly slower than those of sub-gaussian random variables. Similar to sub-gaussian random variables, this class of random variable can be understood through the following norm:

Definition 1. *The **sub-exponential norm** of $X \in \mathbb{R}$ is*

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

*If $\|X\|_{\psi_1}$ is finite, we say that X is **sub-exponential**.*

Fact 2 (See 2.7.1 in V for greater detail)**.** *The above is equivalent (up to some absolute constant) to $\mathbb{P}(|X| \ge t) \le 2\exp(-ct/\|X\|_{\psi_1})$ for all $t \ge 0$. Moreover, the MGF of $|X|$ is bounded at some point, namely*

$$\mathbb{E}\exp(|X|/K_4) \le 2.$$

We will see that sub-exponential distributions and sub-gaussian distributions are closely related. First, note that any sub-gaussian distribution is sub-exponential. Second, the square of a sub-gaussian random variable is sub-exponential:

Lemma 3 (2.7.6 in V)**.** *A random variable X is sub-gaussian if and only if X^2 is sub-exponential. Moreover,*

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

Idea of Proof.
We have that

$$\|X^2\|_{\psi_1} = \inf\left\{t > 0 : \mathbb{E}\exp(|X^2|/t) \le 2\right\}.$$

For any t for which this is true, take $(\sqrt{t})^2$ in the definition of the sub-gaussian norm.

With the above, we can now state a concentration inequality for sums of independent sub-exponential random variables.

Theorem 4 (Bernstein's inequality)**.** *Let $X_1, \dots, X_N$ be independent, mean zero, sub-exponential random variables. Then, for every $t \geq 0$, we have*

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

where $c > 0$ is an absolute constant.

Proof. See 2.8.1 in Vershynin. □

Let us see what this says for random variables satisfying $\|X_i\|_{\psi_1} = 1$:

$$\mathbb{P}\left\{\left|\sum_{i=1}^N X_i\right| \geq t\right\} \leq \begin{cases} \exp(-ct^2/N) & t < N \\ \exp(-ct) & t \geq N. \end{cases}$$

This says that, for sums of sub-exponential random variables, small deviations from the mean have sub-gaussian tails whereas large deviations have heavier sub-exponential tails.

3 Concentration of the norm

Given a random vector $X = (X_1, \dots, X_n) \in \mathbb{R}^n$, we may ask where its norm is likely to be located. This is answered by the following theorem:

Theorem 5 (3.1.1 in V)**.** *Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector with independent, sub-gaussian coordinates X_i that satisfy $\mathbb{E}X_i^2 = 1$. Then*

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

where $K = \max_i \|X_i\|_{\psi_2}$.

Proof. In the following, we assume $K \geq 1$ (this is justified below). We will apply Bernstein's inequality to $\frac{1}{n}\|X\|_2^2 - 1$ by first noting that this quantity is equivalent to $\frac{1}{n}\sum_{i=1}^n (X_i^2 - 1)$ and that this is indeed a sum of sub-exponential random variables:

$$\begin{aligned} \|X_i^2 - 1\|_{\psi_1} &\lesssim \|X_i^2\|_{\psi_1} \qquad \text{(by centering lemma)} \\ &= \|X_i\|_{\psi_2}^2 \\ &\lesssim K^2. \end{aligned}$$

Apply Bernstein's inequality with $t = nu$:

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n}\|X\|_2^2 - 1\right| \geq u\right) &\leq 2\exp\left(-c\min\left(\frac{n^2u^2}{nK^4}, \frac{nu}{K^2}\right)\right) \\ &\leq 2\exp\left(\frac{-cn}{K^4}\min(u^2, u)\right) \quad (K \leq 1). \end{aligned} \tag{1}$$

This is a concentration inequality for $\|X\|_2^2$. To make the link towards a concentration inequality for $\|X\|_2$, observe (this is an exercise) that for $z \geq 0$:

$$|z-1| \geq \delta \quad \text{implies} \quad |z^2 - 1| \geq \max(\delta, \delta^2).$$

We obtain, for any $\delta \geq 0$, that

$$\begin{aligned}\mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\|X\|_2 - 1\right| \geq \delta\right\} &\leq \mathbb{P}\left\{\left|\frac{1}{n}\|X\|_2^2 - 1\right| \geq \max(\delta, \delta^2)\right\} \\ &\leq 2\exp\left(-\frac{cn}{K^4}\cdot\delta^2\right) \quad \text{(by (1) for } u = \max(\delta,\delta^2)).\end{aligned}$$

Changing variables to $t = \delta\sqrt{n}$, we obtain the desired concentration inequality

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

$\square$

Remark. Why may we assume $K \geq 1$? Note that $e^x \geq 1 + x$. Hence

$$\begin{aligned}2 &\geq \mathbb{E}\left(\exp\left(\frac{|X_i|^2}{\|X_i\|_{\psi_2}^2}\right)\right) \\ &\geq \mathbb{E}\left(\frac{|X_i|^2}{\|X_i\|_{\psi_2}^2}\right) + 1\end{aligned}$$

Recalling that $\mathbb{E}X_i^2 = 1$ and rearranging, we have

$$\mathbb{E}\left(\frac{1}{\|X_i\|_{\psi_2}^2}\right) \leq 1.$$

That is, $\|X_i\|_{\psi_2}$ is at least 1.

References

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