MATH888: High-dimensional probability and statistics Fall 2021

Lecture 13 — October 6, 2021

Sebastien Roch, UW-Madison *Scribe: Shi Chen*

1 Overview

In the last lecture, we define the sub-Gaussian vector and its norm:

Definition 1 (Sub-Gaussian random vectors)**.** *A random vector $X \in \mathbb{R}^n$ is a sub-Gaussian vector if its projection $\langle X, y\rangle$ is a sub-Gaussian random variable for any $y \in \mathbb{R}^n$. The sub-Gaussian norm of X is defined as*

$$\|X\|_{\psi_2} = \sup_{y\in S^{n-1}} \|\langle X, y\rangle\|_{\psi_2}. \tag{1}$$

We saw one example last time where the coordinates of $X \in \mathbb{R}^n$ were independent (Lemma 3 in Lecture 12). In this example, we were able to bound the norm of a sub-Gaussian vector by the maximum of the norm of its coordinates. In general, sub-Gaussian vectors may not have independent coordinates. This can be seen from the uniform distribution on the sphere (Theorem 4 in Lecture 12). Further more, in this example, we obtain the scaling of its sub-Gaussian norm in terms of dimension, which is itself interesting in the study of high dimensional probability.

In this lecture, we are going back to two interesting high dimensional examples presented at the very beginning of the course. The first example is the concentration of Gaussian vector $X \sim \mathcal{N}(0, I_p)$ near the sphere of radius $R = \sqrt{p-1}$ (see page 17 in Lecture 1 slides). Using the Hanson-Wright inequality, we can show that the Euclidean norm of a sub-Gaussian vector with independent coordinates is concentrated near $\sqrt{n}$, where n is the dimension of the vector. However, we will see that the concentration is lost if the independence is removed. Intuitively, we can view $\|X\|_2^2$ as a sum of random variables and the concentration follows from the central limit theorem for independent random variables, while it is in general not true for correlated random variables.

The second example is the mean estimate problem (see page 9 in Lecture 1 slides), where the mean square error bound is computed but is useful only if the dimension of the vector is much smaller than the number of samples. Using the Hanson-Wright inequality, we can obtain a more useful non-asymptotic bound for the mean estimator of sub-Gaussian random vectors.

2 Hanson-Wright inequalities for sub-Gaussian vectors

We begin by introducing the Hanson-Wright inequality inequalities for sub-Gaussian vectors.

Theorem 2 (Exercise 6.3.5 in [1])**.** *Suppose $X \in \mathbb{R}^n$ is mean-zero with sub-Gaussian norm $\|X\|_{\psi_2} = K$. Let B be an $n \times n$ matrix. Then for any $t \geq 0$,*

$$\mathbb{P}\left(\|BX\|_2^2 \geq CK^2\|B\|_F^2 + t\right) \leq \exp\left(-\frac{ct}{K^2\|B\|_2^2}\right) \tag{2}$$

and

$$\mathbb{P}\left(\|BX\|_2 \geq CK\|B\|_F + t\right) \leq \exp\left(-\frac{ct^2}{K^2\|B\|_2^2}\right) . \tag{3}$$

The second inequality (3) shall follow from the first (2) by using the tricks we used before. See Theorem 3 in Lecture 8.

Remark 3. *When sub-Gaussian vector $X \in \mathbb{R}^n$ has zero mean and independent coordinates with unit variance, stronger tail bounds can be proved. We can show*

$$\mathbb{E}\left[\|BX\|_2^2\right] = \|B\|_F^2 , \tag{4}$$

and the concentration inequality (see Theorem 6.3.2 in [1])

$$\mathbb{P}\left\{\left|\|BX\|_2 - \|B\|_F\right| \geq t\right\} \leq \exp\left(-\frac{ct^2}{K^4\|B\|_2^2}\right) , \forall t \geq 0 , \tag{5}$$

which also implies (check!)

$$\left|\mathbb{E}\left[\|BX\|_2 - \|B\|_F\right]\right| \leq CK^2\|B\|_2 . \tag{6}$$

Concentration near the expectation can be seen from (5) *and* (6)*, while it is not guaranteed in* (3)*. In fact, only upper tails is preserved and concentration is lost in* (3)*. Counterexamples with dependent coordinates can be constructed. See Remark 5.*

Remark 4. *When $t \gtrsim K\|B\|_F$, the upper tails of* (3) *and* (5) *are essentially equivalent.*

Remark 5 (Exercise 6.3.6 in [1])**.** *The concentration is lost for general sub-Gaussian vectors without assuming independence. We show an counterexample in the following . Take $B = I_n$, then $\|B\|_F = \sqrt{n}$ and $\|B\|_2 = 1$. We construct $X \in \mathbb{R}^n$ as follows. Let w be uniform over the unit sphere S^{n-1}. Let z be independent of w and satisfies*

$$z = \begin{cases} 0 , & w.\ p.\ \ 1/2 \\ 1 , & w.\ p.\ \ 1/2 \end{cases} . \tag{7}$$

We define $X = \sqrt{2n}zw$. We show that X is a sub-Gaussian vector with zero mean, and its coordinates have unit variance. It is clear that $\mathbb{E}X = 0$. To compute the variance of X_i, we note that by symmetry

$$1 = \mathbb{E}\left[\|w\|_2^2\right] = n\mathbb{E}\left[w_1^2\right] \Longrightarrow \mathbb{E}\left[w_1^2\right] = \frac{1}{n} . \tag{8}$$

So the variance of the coordinates of X is

$$\mathbb{E}\left[X_1^2\right] = \mathbb{E}[(\sqrt{2n})^2 z^2 w_1^2] = \frac{1}{2} \cdot 2n\mathbb{E}(w_1^2) = 1 . \tag{9}$$

The sub-Gaussian norm of X is finite:

$$\begin{aligned} \|X\|_{\psi_2} &= \|\langle X, e_1\rangle\|_{\psi_2} && \textit{(By symmetry)} \\ &= \|X_1\|_{\psi_2} \\ &\lesssim \|\sqrt{n}w_1\|_{\psi_2} && \textit{(Because } |X_1| \leq \sqrt{2n}|w_1| \textit{ a.s.)} \\ &\lesssim C && \textit{(See Theorem 4 in Lecture 12)} \end{aligned} \tag{10}$$

$\|X\|_2$ takes the value 0 or $\sqrt{2n}$ with probability 1/2, and it is clearly not concentrated near $\|B\|_F$ or its expectation

$$\mathbb{E}\|X\|_2 = \frac{1}{2} \cdot 0 + \frac{1}{2}\sqrt{2n} = \sqrt{\frac{n}{2}} . \tag{11}$$

3 Mean estimation of sub-Gaussian vectors

In this section, we get back to the mean estimation problem and use the Hanson-Wright inequality to obtain an error bound for sub-Gaussian vectors.

Theorem 6. *Let $X^{(1)}, \cdots, X^{(n)} \in \mathbb{R}^p$ be i.i.d. sub-Gaussian vectors with sub-Gaussian norm $\|X^{(1)}\|_{\psi_2} = K$ and expectation $\mu = \mathbb{E}(X^{(1)})$. We denote the mean estimation by*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X^{(i)} .$$

Then we have

$$\|\bar{X} - \mu\|_2 \leq \epsilon \qquad w.\ p.\ 1-\delta \tag{12}$$

if the sample complexity $n \gtrsim (\frac{K}{\epsilon})^2[p + \log(\frac{1}{\delta})]$. That is to say, $\bar{X}$ is a good estimation of μ with high probability.

Proof. Denote by

$$\begin{aligned} Y^{(i)} &= \frac{1}{n}(X^{(i)} - \mu) \\ Y &= \sum_{i=1}^{n} Y^{(i)} = \bar{X} - \mu \,, \end{aligned}$$

the centered variables, that is, $\mathbb{E}Y = 0$. We are going to estimate the sub-Gaussian norm of Y and then apply the Hanson-Wright inequality (2) to it. Since $Y^{(i)}$ is centered, we can apply the Hoeffding inequality (see Theorem 2 in Lecture 7) to Y and obtain

$$\|Y\|_{\psi_2}^2 = \sup_{u\in S^{p-1}} \|\langle Y, u\rangle\|_{\psi_2}^2 \lesssim \sum_{i=1}^{n} \sup_{u\in S^{p-1}} \|\langle Y^{(i)}, u\rangle\|_{\psi_2}^2 \lesssim \sum_{i=1}^{n} \|Y^{(i)}\|_{\psi_2}^2 \,. \tag{13}$$

We thus need to estimate the sub-Gaussian norm of $Y^{(i)}$. To do this, we use the centering lemma (see Lemma 1 in Lecture 7)

$$\begin{aligned} \|Y^{(i)}\|_{\psi_2} &= \sup_{u\in S^{p-1}} \|\langle Y^{(i)}, u\rangle\|_{\psi_2} \\ &= \sup_{u\in S^{p-1}} \frac{1}{n}\|\langle X^{(i)} - \mu, u\rangle\|_{\psi_2} \\ &\lesssim \sup_{u\in S^{p-1}} \frac{1}{n}\|\langle X^{(i)}, u\rangle\|_{\psi_2} \qquad \text{(Centering lemma)} \\ &\lesssim \frac{1}{n}\|X^{(i)}\|_{\psi_2} \\ &\lesssim \frac{K}{n} \,. \end{aligned} \tag{14}$$

Now combining (13) and (14) we obtain the sub-Gaussian norm for Y:

$$\|Y\|_{\psi_2} \lesssim \frac{K}{\sqrt{n}} \,. \tag{15}$$

We take $B = I_p$ in the Hanson-Wright inequality (2) and it provides

$$\mathbb{P}\left(\|Y\|_2 \geq C\frac{K^2}{n}p + t\right) \leq \exp\left(-\frac{cnt}{K^2}\right) . \tag{16}$$

Then we choose $t = \frac{K^2}{cn}\log(\frac{1}{\delta})$ and obtain

$$\mathbb{P}\left(\|Y\|_2 \geq \frac{K^2}{n}\left(Cp + c\log\left(\frac{1}{\delta}\right)\right)\right) \leq \delta \, . \tag{17}$$

Finally, we take $\epsilon = \frac{K^2}{n}\left(Cp + c\log\left(\frac{1}{\delta}\right)\right)$ and it completes the proof. □

References

[1] R. Vershynin, *High-Dimensional Probability: An introduction with Applications in Data Science*, Cambridge University Press, 2008.