Lecture 13 — October 6, 2021

Sebastien Roch, UW-Madison *Scribe: Yun Li and Sebastien Roch*

1 Overview

In the last lecture we introduced isotropic random vectors and sub-Gaussian random vectors. We proved that if a vector has independent, mean-zero, sub-Gaussian entries then it is a sub-Gaussian vector. We also showed that the uniform random vector on the n-dimensional sphere of radius $\sqrt{n}$ centered at the origin is sub-Gaussian. Note that the vectors in the second example have dependent coordinates and are scaled according to the dimension.

In this lecture we discussed tail inequalities for the norm of sub-Gaussian vectors and applied the inequality to estimate the mean of sub-Gaussian vectors.

2 Tail inequalities for the norm of sub-Gaussian vectors

Recall that the sub-Gaussian norm of a random vector $X \in \mathbb{R}^n$ is defined as

$$\|X\|_{\psi_2} := \sup_{u \in \mathbb{S}^{n-1}} \|\langle X, u\rangle\|_{\psi_2},$$

where $\mathbb{S}^{n-1} = \{u \in \mathbb{R}^n : \|u\|_2 = 1\}$. We began with the following Hanson-Wright inequality.

Theorem 1 (Norm of sub-Gaussian vectors: anisotropic case)**.** *If X is a mean-zero, sub-Gaussian random vector in $\mathbb{R}^n$ with $\|X\|_{\psi_2} = K$. Let B be an $m \times n$ matrix. Then for any $t \geq 0$ we have*

$$\mathbb{P}\left(\|BX\|_2^2 \geq CK^2\|B\|_F^2 + t\right) \leq \exp\left(-\frac{ct}{K^2\|B\|_2^2}\right), \tag{1}$$

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

where $\|\cdot\|_F$ and $\|\cdot\|_2$ denote the Frobenius norm, and operator norm respectively.

Proof. The proof is left as an exception. See Exercises 6.2.6 and 6.3.5 in [1]. □

Remarks:

1. If we assume further that the coordinates of X are independent, sub-Gaussian, and unit variance, then we have

$$\mathbb{E}(\|BX\|_2^2) = \|B\|_F^2,$$

and for any $t \geq 0$,

$$\mathbb{P}\left(|\|BX\|_2 - \|B\|_F| > t\right) \leq \exp\left(-\frac{ct^2}{K^4\|B\|_2^2}\right). \quad (3)$$

Note that the bound (3) also implies

$$|\mathbb{E}(\|BX\|_2) - \|B\|_F| \leq CK^2\|B\|_2,$$

which can be checked by integrating the tail probabilities.

2. When $t \gtrsim K\|B\|_F$, the upper tails of (2) and (3) exhibit a similar behavior.

3. The next example shows that the concentration inequality (3) might fail if we drop the independence assumption.

   Example: Let $B = I_n$, then $\|B\|_F^2 = n, \|B\|_2 = 1$. Let w be a random vector uniformly distributed on $\mathbb{S}^{n-1}$, and let z be a random variable that satisfies $\mathbb{P}(z = 0) = \mathbb{P}(z = 1) = \frac{1}{2}$ and independent of w. Consider the vector defined by $X = \sqrt{2n}zw$. By symmetry and the assumption that $w \in \mathbb{S}^{n-1}$, we have $\mathbb{E}(w_i) = 0$ and $1 = \mathbb{E}(\sum_{i=1}^n w_i^2) = n\mathbb{E}(w_1^2)$. Therefore, $\mathbb{E}(X_i) = 0$ and $\mathbb{E}(X_i^2) = n\mathbb{E}(w_i^2) = 1$. Note that by definition $|X_1| \leq \sqrt{2n}|w_1|$ a.s., hence

   $$\|X\|_{\psi_2} = \|\langle X, e_1\rangle\|_{\psi_2} = \|X_1\|_{\psi_2} \lesssim \|\sqrt{n}w_1\|_{\psi_2} \lesssim C,$$

   where the step follows from Lecture 12. The coordinates of X are sub-Gaussian and unit variance (but not independent). It is direct to check that the expectation of $\|X\|_2$ satisfies $\mathbb{E}(\|X\|_2) = \sqrt{n/2}$, which is not centered around $\|B\|_F = \sqrt{n}$. The concentration inequality (3) does not hold.

3 An application: mean estimation of sub-Gaussian vectors

Theorem 2 (Mean estimation of sub-Gaussian vectors). *Let $X^{(1)}, X^{(2)}, \cdots, X^{(n)} \in \mathbb{R}^p$ be i.i.d. random vectors with $\|X^{(i)}\|_{\psi_2} = K$, $\mathbb{E}(X^{(i)}) = \mu$. Let $\bar{X} := \frac{1}{n}\sum_{i=1}^n X^{(i)}$ be the empirical mean. Then the event $\{\|\bar{X} - \mu\|_2 \leq \epsilon\}$ happens with probability $1 - \delta$ if $n \gtrsim (\frac{K}{\epsilon})^2(p + \log(\frac{1}{\delta}))$.*

Proof. Define $Y^{(i)} = \frac{1}{n}(X^{(i)} - \mu)$ and $Y = \sum_{i=1}^n Y^{(i)} = \bar{X} - \mu$. Then $\mathbb{E}(Y) = 0$ and Y is a sum of independent, mean-zero, sub-Gaussian vectors. For $u \in \mathbb{S}^{p-1}$, we have

$$\begin{aligned}
\|\langle Y^{(i)}, u\rangle\|_{\psi_2} &= \|\langle \tfrac{1}{n}(X^{(i)} - \mu), u\rangle\|_{\psi_2} \\
&\lesssim \tfrac{1}{n}\|\langle X^{(i)}, u\rangle\|_{\psi_2} \qquad \text{(by Lemma 2.6.8 in [1])} \\
&\lesssim \tfrac{K}{n}.
\end{aligned}$$

Since $\langle Y, u\rangle$ is a sum of independent, mean zero, sub-Gaussian random variables, we have by Proposition 2.6.1 in [1] that

$$\|\langle Y, u\rangle\|_{\psi_2}^2 \lesssim \sum_{i=1}^n \|\langle Y^{(i)}, u\rangle\|_{\psi_2}^2 \lesssim \frac{K^2}{n},$$

which implies that $\|Y\|_{\psi_2} \lesssim K/\sqrt{n}$. Note that a similar estimation was used when we proved the Hoeffding's inequality. Apply Theorem 1 (in particular the inequality (1)) to Y, $B = I_p$ and $t = CK^2 \log(\frac{1}{\delta})/n$, we get

$$\mathbb{P}\left(\|Y\|_2^2 \geq C\frac{K^2}{n}\left(p + \log(\frac{1}{\delta})\right)\right) \leq \exp\left(-\frac{cC\frac{K^2}{n}\log(\frac{1}{\delta})}{\frac{K^2}{n}}\right) \lesssim \delta$$

for large enough C. Therefore, the event $\{\|\bar{X} - \mu\|_2 \geq \epsilon\}$ happens with probability at most δ if $n \gtrsim (\frac{K}{\epsilon})^2(p + \log(\frac{1}{\delta}))$. This completes the proof. □

References

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