1 Overview

In the last lecture, we proved tail inequalities for the norm of sub-Gaussian vectors. We recall this result in a slightly different form:

Given $Z \in \mathbb{R}^p$, $\|Z\|_{\psi_2} \leq K$, $\mathbb{E}Z = 0$ and $\Sigma \in \mathbb{R}^{p \times p}$, $\|\Sigma\|_2 \geq 0$, we have $\forall t \geq 0$,

$$\mathbb{P}(Z^T \Sigma Z \geq CK^2 tr(\Sigma) + t) \leq \exp\left(\frac{-ct}{K^2 \|\Sigma\|_2}\right)$$

In order to connect with the result from last lecture, note that $\|B\|_F^2 = tr(B^T B)$ and $\|B\|_2^2 = \|B^T B\|_2$.

2 Mean estimation of sub-Gaussian vectors

Lemma 1. *Let $Z^{(1)}, ...Z^{(n)}$ be independent, mean-zero, sub-Gaussian vectors, then*

$$\|\Sigma_{i=1}^n \alpha_i Z^{(i)}\|_{\psi_2}^2 \leq \Sigma_{i=1}^n \alpha_i^2 \|Z^{(i)}\|_{\psi_2}^2$$

Proof. For any $\vec{u} \in \mathbb{S}^{p-1}$,

$$\|\langle \vec{u}, \Sigma_{i=1}^n \alpha_i Z^{(i)} \rangle\|_{\psi_2}^2 = \|\Sigma_{i=1}^n \alpha_i \langle \vec{u}, Z^{(i)} \rangle\|^2 \tag{1}$$

$$\leq \Sigma_{i=1}^n \alpha_i^2 \|\langle \vec{u}, Z^{(i)} \rangle\|_{\psi_2}^2 \tag{2}$$

We get the inequality in the second line using Prop. 2.6.1 from [1]. Now, taking supremum wrt $\vec{u}$, we have,

$$\sup_{\vec{u} \in \mathbb{S}^{p-1}} \|\langle \vec{u}, \Sigma_{i=1}^n \alpha_i Z^{(i)} \rangle\|_{\psi_2}^2 =\leq \Sigma_{i=1}^n \alpha_i^2 \sup_{\vec{u} \in \mathbb{S}^{p-1}} \|\langle \vec{u}, Z^{(i)} \rangle\|_{\psi_2}^2 \tag{3}$$

Now, using $\sup_{\vec{u} \in \mathbb{S}^{p-1}} \|\langle \vec{u}, Z^{(i)} \rangle\|_{\psi_2} = \|Z^{(i)}\|_{\psi_2}$, we get the requisite result. □

Theorem 2 (Mean estimation of sub-Gaussian vectors (taking into account covariance))**.** *If $X^{(1)}, ...X^{(n)} \in \mathbb{R}^{\prime}$ be iid random vectors with $\mathbb{E}(X^{((i))}) = \vec{\mu}$, $Cov(X^i) = \mathbb{E}((X^{(i)} - \vec{\mu})^T (X^{(i)} - \vec{\mu})) = \Sigma \geq 0$ and $\bar{X} = \frac{1}{n} \Sigma_{i=1}^n X^{(i)}$. Assume further that $\forall \vec{w} \in \mathbb{R}^p$,*

$$\|\langle \vec{w}, X^{(i)} - \vec{\mu} \rangle\|_{\psi_2} \leq K \sqrt{\mathbb{E}(\langle \vec{w}, X^{(i)} - \vec{\mu} \rangle)}$$

This last requirement is an anisotropic version of a sub-Gaussian norm of a random vector. Then with probability $p \leq 1-\delta$,

$$\|\bar{X}-\vec{\mu}\|_2^2 \leq \frac{K^2}{n}\|\Sigma\|_2\left(\frac{tr\Sigma}{\|\Sigma\|_2}+\log(1/\delta)\right)$$

Note:

1. The term $\frac{tr\Sigma}{\|\Sigma\|_2}$ (stable rank) is an indicator of the rank/effective dimension of sub-Gaussian vector. When $\Sigma = I_p$, $\frac{tr\Sigma}{\|\Sigma\|_2} = p$; if some of the diagonal terms are zero, this would result in $\frac{tr\Sigma}{\|\Sigma\|_2} \leq p$.

2. Given $\Sigma = UDU^T$; $\Sigma^+ = UD^+U^T$ where

$$D_{ii}^+ = \begin{cases} \frac{1}{D_{ii}}, & \text{if } D_{ii} < 0 \\ 0, & \text{otherwise.} \end{cases}$$

   And

$$\sqrt{\Sigma^+} = U\sqrt{D^+}U^T$$

   where we get $\sqrt{D^+}$ by taking the square root of the diagonal terms in D.

Proof. Let $Z^{(i)} = \sqrt{\Sigma^+}(X^{(i)}-\vec{\mu})$, then $X^{(i)} = \sqrt{\Sigma^+}Z^{(i)}+\vec{\mu}$ with probability $p=1$.

So $\|\bar{X}-\vec{\mu}\|_2^2 = \bar{Z}^T\Sigma\bar{Z}$.

Now, it remains to bound $\|Z^{(i)}\|_{\psi_2}$

$$\begin{aligned} \|\langle\vec{w}, Z^{(i)}\rangle\|_{\psi_2} &= \|\langle\vec{w}, \sqrt{\Sigma^+}(X^{(i)}-\vec{\mu})\rangle\|_{\psi_2} && (4)\\ &= \|\langle\sqrt{\Sigma^+}\vec{w}, (X^{(i)}-\vec{\mu})\rangle\|_{\psi_2}, && (5)\end{aligned}$$

where $\vec{w} \in \mathbb{S}^{p-1}$. The second line is due to the symmetry of $\sqrt(\Sigma^+)$.

From Lemma 1, we have

$$\begin{aligned} \|\langle\sqrt{\Sigma^+}\vec{w}, X^{(i)}-\vec{\mu}\rangle\|_{\psi_2} &\leq K\sqrt{\mathbb{E}\langle\sqrt{\Sigma^+}\vec{w}, X^{(i)}-\vec{\mu}\rangle} && (6)\\ &\leq K && (7)\end{aligned}$$

The last line follows from: $\mathbb{E}(\langle X^{(i)}-\vec{\mu}, \vec{w}\rangle^2) = \langle\Sigma\vec{w}, \vec{w}\rangle$. Therefore, we have,

$$\|\bar{Z}\|_{\psi_2}^2 \leq \frac{K^2}{n}$$

. The final result follows from the tail inequality for sub-Gaussian vectors. □

References

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