36-709: Advanced Statistical Theory	**Spring 2019**

Lecture 8: February 21

Lecturer: Alessandro Rinaldo *Scribes: Shenghao Wu*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

This lecture's notes illustrate some uses of various LaTeX macros. Take a look at this and imitate.

8.1 Euclidean norm of sub-Gaussian random vectors

Definition 8.1 ***(Sub-Gaussian random vectors)*** *A random vector $X \in \mathbb{R}^d$ is a sub-Gaussian random vector with parameter σ^2 if*

$$v^T X \in SG(\sigma^2), \forall v \in \mathbb{S}^{d-1}$$

where $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : ||x|| = 1\}$ is the $d-1$ unit sphere. We write $X \in SG_d(\sigma^2)$.

Lemma 8.2 *$X \in \mathbb{R}^d$ is a sub-Guassian random vector with parameter $||\Sigma||_{op}$ if $X \sim \mathcal{N}(0, \Sigma)$*

Proof: For any $v \in \mathbb{S}^{d-1}$, $v^T\Sigma v \leq ||\Sigma||_{op}$. Take MGF: $\mathbb{E}[e^{\lambda v^T X}] = e^{\lambda^2 v^T \Sigma v/2} \leq e^{\lambda^2 ||\Sigma||_{op}/2}$ ■
Notice that sub-Guassian vector does not need to be a vector of independent Gaussians (but the vice is true).

We now prove the theorem from last time:

Theorem 8.3 *Let $X \in SG_d(\sigma^2), ||X|| = \sqrt{\sum_{i=1}^d X_i^2}$, then:*

$$\mathbb{E}[||X||] \leq 4\sigma\sqrt{d}$$

Moreover, with probability at least $1-\delta$ for $\delta \in (0,1)$:

$$||X|| \leq 4\sigma\sqrt{d} + 2\sigma\sqrt{log(\frac{1}{\delta})}$$

Proof: Let $N_{\frac{1}{2}}$ be a $\frac{1}{2}$-minimal cover of B_d in Euclidian norm, that is:

$$\forall \theta \in B_d, \exists z = z(\theta) \in N_{\frac{1}{2}} \ s.t. \ ||\theta - z|| \leq \frac{1}{2}$$

. Equivalently, $\forall \theta \in B_d$, we can write $\theta = z + w$ where $z = z(\theta) \in N_{\frac{1}{2}}$ and $||w|| \leq \frac{1}{2}$. Also, by the volumetric rate bounds,

$$|N_{\frac{1}{2}}| \leq (1 + \frac{2}{1/2})^d = 5^d$$

Hence,

$$\max_{v\in B_d} v^T X \leq \max_{z\in N_{\frac{1}{2}}} z^T X + \max_{w\in \frac{1}{2}B_d} w^T X = \max_{z\in N_{\frac{1}{2}}} z^T X + \frac{1}{2}\max_{w\in B_d} w^T X$$

Hence $\underbrace{\max_{v\in B_d} v^T X}_{||X||} \leq 2\max_{z\in N_{\frac{1}{2}}} z^T X$.

In general, some argument will lead to the following bound:

$$||X|| \leq \frac{1}{1-\epsilon}\max_{z\in N_{\frac{1}{2}}} z^T X \ for\ \epsilon \in (0,1)$$

Therefore,

$$\mathbb{E}[||X||] \leq 2\mathbb{E}[\max_{z\in N_{\frac{1}{2}}} \underbrace{z^T X}_{SG(\sigma^2)}] \leq 2\sigma\sqrt{2log|N_{\frac{1}{2}}|} \leq 2\sigma\sqrt{2dlog5} \leq 4\sigma\sqrt{d}$$

The second inequality is due to the maximal inequality for sub-Gaussian random variables we have proved in class.

To prove the high probability bound, for $t > 0$:

$$\mathbb{P}(||X|| \geq t) \leq \mathbb{P}(\max_{z\in N_{\frac{1}{2}}} z^T X \geq \frac{t}{2}) \leq |N_{\frac{1}{2}}|exp\{-\frac{t^2}{8\sigma^2}\} \leq 5^d exp\{-\frac{t^2}{8\sigma^2}\}$$

The desired bound is obtained by setting the right hand side equal to $\sigma \in (0,1)$ and solve for t ■

Note: we have already seen from HW1 that under some regularity condition [Y10]:

$$||\hat{\Sigma}_n - \Sigma||_\infty \leq C\sqrt{\frac{t+log\delta}{n}}$$

with probability at least $1-e^{-t}$, where $\hat{\Sigma}_n$ is the empirical covariance matrix.

8.2 Matrix norm

Definition 8.4 ***(Operator Norm)*** *Let $A \in \mathbb{R}^{m\times n}, rank(A) = r \leq \min\{m,n\}$. The singular value decomposition (SVD) of A is $A = UDV^T$ where*

1. *$D = diag(\sigma_1, \cdots, \sigma_r)$. $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are the singular values of A.*
2. *$U \in \mathbb{R}^{m\times r}$ whose columns are orthonormal and are called singular vectors*
3. *$V \in \mathbb{R}^{n\times r}$ whose columns are orthonormal and are called singular vectors*

Then $AA^T u_j = \sigma_j^2 u_j$, $A^T A v_j = \sigma_j^2 v_j$ where u_j, v_j are the j-th column of U and V respectively. The operator norm of A is:

$$||A||_{op} = \max_i \sigma_i = \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{||Ax||}{||X||} = \max_{\substack{x\in\mathbb{S}^{n-1}\\ y\in\mathbb{S}^{n-1}}} x^T Ay$$

Remarks:

- When $A \in S^n$(symmetric), $||A||_{op} = \max_{x\in\mathbb{S}^{n-1}} |x^T Ax|$.
- We say$A \in S^n_+$(positive semi-definite (PSD)) if and only if $\forall x \in \mathbb{R}^n, x^T Ax \geq 0$. As an example, any covariance matrix Σ is PSD because $\mathbb{V}[a^T x] = a^T \geq 0, \forall a \in \mathbb{R}^n$
- If $A \in S^n_+$, $\sigma_i = \lambda_i$ where λ_i's are the eigenvalues of A, $||A||_{op} = \max_i \lambda_i = \max_{x\in\mathbb{S}^{n-1}} x^T Ax$

The following two types of norms are also common in practice.

Definition 8.5 ***(Frobenius Norm)***

$$||A||_F = \sqrt{\sum_{i,j=1} A_{ij}^2}$$

Definition 8.6 ***(p-Schatten Norm)***

$$||A||_p = (\sum_{i=1}^{n\wedge m} \sigma_i^p(A))^{1/p}$$

where $\sigma_i(A)$'s are the singular values of A. When $p = 1$, $||A||_p$ is the nuclear norm. When $p = \infty$, $||A||_p$ is the spectral norm.

The following two inequality are often useful in practice:

Lemma 8.7

$$||Ax|| \leq ||A||_{op}||x|| \; \forall x$$

Lemma 8.8 ***(Weyl's inequality)*** *Assume $A, B \in \mathbb{R}^m$ have singular values $\sigma_i(A), \sigma_j(B)$ for $i = 1, \cdots, n \wedge m; j = 1, \cdots, n \wedge m$, then:*

$$\max_i |\sigma_i(A) - \sigma_i(B)| \leq ||A - B||_{op}$$

Corollary $||A - B||_{op} \to 0 \Rightarrow |x^T(A - B)y| \to 0$ uniformly for every $x \in \mathbb{S}^{m-1}, y \in \mathbb{S}^{n-1}$.

8.3 Covariance matrix estimation in the operator norm

Theorem 8.9 *Let $X_1, \cdots, X_n$ be iid samples from a distribution with mean 0 and covariance matrix Σ. Assume $X_i \in SG_d(\sigma^2)$ and are centered. Let $\hat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^T$. Then there exists a universal constant $C > 0$ s.t.*

$$\mathbb{P}(\frac{||\hat{\Sigma}_n - \Sigma||_{op}}{\sigma^2} \geq C \max\{\sqrt{\frac{d + log(\frac{2}{\delta})}{n}}, \frac{d + log(\frac{2}{\delta})}{n}\}) \leq \delta, \; \delta \in (0, 1)$$

Remark: Theorem 8.9 indicates that $\hat{\Sigma}_n \xrightarrow{p} \Sigma$ with respect to the operator norm requires $\frac{d}{n} \to 0$
Proof ideas: Use discretization and sub exponential concentration bound. Recall that $X \in SG(\sigma^2) \Rightarrow X^2 - \mathbb{E}[X^2] \in SE(16^2\sigma^4, 16\sigma^2)$.

Lemma 8.10 *Let $A := \hat{\Sigma}_n - \Sigma \in S^n$ and N_ϵ be the ϵ-net of $\mathbb{S}^{d-1}$ for $\epsilon \in (0, \frac{1}{2})$, then:*

$$||A||_{op} = \max_{x\in\mathbb{S}^{n-1}} |x^T Ax| \leq \frac{1}{1 - 2\epsilon} \max_{y\in N_\epsilon} |y^T Ay|$$

References

[VK14] V. Koltchinskii and K. Lounici, "Concentration inequalities and moment bounds for sample covariance operators," *arXiv preprint*, 2014, arXiv:1405.2468.

[Y10] Y. Ming, "High dimensional inverse covariance matrix estimation via linear programming," *Journal of Machine Learning Research* , 2010, pp. 2261–2286.