# Spring 2018: STA 6448
# Advanced Probability and Inference II
# Lecture 15

Yun Yang

- Random matrices and covariance estimation

# Covariance matrices from sub-Gaussian ensembles

Our previous development has crucially exploited different properties of the Gaussian distribution. Now, we show a different approach for general sub-Gaussian random matrices.

## Definition

We call a random vector $x \in \mathbb{R}^d$ zero-mean and sub-Gaussian with parameter $\sigma^2$ if for each fixed $v \in \mathcal{S}^{d-1}$,

$$\mathbb{E}[e^{\lambda \langle v, x \rangle}] \leq e^{\frac{\lambda^2 \sigma^2}{2}}, \quad \text{for all } \lambda \in \mathbb{R}.$$

We assume each row $x_i$ of $X$ is zero-mean, and sub-Gaussian with parameter $\sigma^2$.

## Example

- $X \in \mathbb{R}^{n \times d}$ has i.i.d. entries that are zero-mean and sub-Gaussian with parameter $\sigma^2$.
- $x_i \sim \mathcal{N}(0, \Sigma)$ where $\sigma^2 = \|\Sigma\|_{\text{op}}$.

# Concentration of sub-Gaussian ensembles

**Theorem**

*Suppose $x_1, \ldots, x_n$ are i.i.d. samples from a zero-mean sub-Gaussian distribution with parameter $\sigma^2$. Then*

$$\mathbb{E}[e^{\lambda \, \|\widehat{\Sigma} - \Sigma\|_{op}/\sigma^2}] \leq e^{\frac{8\lambda^2}{n} + 4d}, \quad \text{for all } \lambda \in [0, \tfrac{n}{8}].$$

*Moreover, there is some universal constant $c > 0$ such that for all $t > 0$,*

$$\mathbb{P}\Big[\|\widehat{\Sigma} - \Sigma\|_{op}/\sigma^2 \geq c\left(\sqrt{\frac{d}{n}} + \frac{d}{n} + \sqrt{\frac{t}{n}} + \frac{t}{n}\right)\Big] \leq e^{-t}.$$

An equivalent concentration inequality: there are universal constants $c_1, c_2 > 0$ such that for all $\delta > 0$,

$$\mathbb{P}\Big[\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}}/\sigma^2 \geq c_1\left(\sqrt{\frac{d}{n}} + \frac{d}{n}\right) + \delta\Big] \leq e^{-c_2 \, n \, \min\{\delta, \delta^2\}}.$$

# Proof: Concentration of sub-Gaussian ensembles

Without loss of generality, assume $\sigma = 1$.

Use the shorthand $Q = \widehat{\Sigma} - \Sigma$. Then

$$\|Q\|_{\mathrm{op}} = \max_{v \in \mathcal{S}^{d-1}} |\langle v,\, Qv \rangle|.$$

Let $v^1, \ldots, v^N$ be a $\frac{1}{8}$-cover of $\mathcal{S}^{d-1}$, where $N \leq 17^d$. Then

$$\|Q\|_{\mathrm{op}} = \max_{v \in \mathcal{S}^{d-1}} |v^T Q v| \leq 2 \max_{j=1,\ldots,N} |\langle v^j,\, Qv^j \rangle|.$$

For any $\lambda > 0$ and fixed $u \in \mathcal{S}^{d-1}$,

$$\mathbb{E}[e^{2\lambda \, \langle u,\, Qu \rangle}] = \prod_{i=1}^{n} \mathbb{E}[e^{\frac{2\lambda}{n} \{\langle x_i,\, u \rangle^2 - \langle u,\, \Sigma u \rangle\}}$$

# Proof: Concentration of sub-Gaussian ensembles

Since $z_i = \langle x_i, u \rangle$ is sub-Gaussian with mean $\gamma_i = \langle u, \Sigma u \rangle \le \sigma^2$, we have (why?)

$$\mathbb{E}[e^{\frac{t z_i^2}{2\sigma^2}}] \le \frac{1}{\sqrt{1-t}}, \quad |t| \le 1.$$

This implies

$$\mathbb{E}[e^{\frac{t(z_i^2 - \gamma_i^2)}{2\gamma_i^2}}] \le \frac{e^{-t/2}}{\sqrt{1-t}} \le e^{t^2/2}, \quad |t| \le 1/2,$$

$$\text{and} \quad \mathbb{E}[e^{2\lambda \langle u, Qu \rangle}] \le e^{\frac{8\lambda^2}{n^2} \sum_{i=1}^n \gamma_i^2} \le e^{\frac{8\lambda^2}{n}}, \quad |\lambda| \le n/8.$$

Therefore, a union argument yields that for $\lambda \in [0, n/8]$,

$$\mathbb{E}[e^{\lambda \|Q\|_{\mathsf{op}}}] \le 2N\, e^{\frac{8\lambda^2}{n}} \le e^{\frac{8\lambda^2}{n} + 4d}.$$

# Bounds for general matrices: Background

- Matrix-valued function: Any function $f : \mathbb{R}^d \to \mathbb{R}^d$ can be extended to a map from $\mathcal{S}^{d\times d}$ to itself through

$$f(Q) = U\,\mathrm{diag}(\gamma(Q))U^T,$$

  where $Q = U\,\mathrm{diag}(\gamma(Q))U^T$ is the SVD of $Q \in \mathcal{S}^{d\times d}$.
- Unitary invariant property: for any unitary matrix $V$,

$$f(VQV^T) = Vf(Q)V^T.$$

- Spectral mapping property: the eigenvalues of the $f(Q)$ are simply the eigenvalues of $Q$ transformed by $f$.
- Examples: matrix exponential $e^Q = \sum_{k=0}^{\infty} \frac{Q^k}{k!}$, defined for all $Q \in \mathcal{S}^{d\times d}$; matrix logarithm $\log Q$, defined for all $Q \succ 0$.

# Tail conditions for matrices

- Moments: $j$th moment of a symmetric random matrix $Q$ is defined by $\mathbb{E}[Q^j]$.
- Variance: $\mathrm{Var}(Q) = \mathbb{E}[Q^2] - (\mathbb{E}[Q])^2 \succeq 0$ (Exercise).
- If $Q$ has polynomial moments of all orders, then its cumulative generating function $\Pi_Q : \mathbb{R} \to \mathcal{S}^{d\times d}$ is given by

$$\Pi_Q(\lambda) = \log \mathbb{E}[e^{\lambda Q}].$$

**Definition**

A zero-mean symmetric random matrix $Q \in \mathcal{S}^{d\times d}$ is sub-Gaussian with matrix parameter $V \in \mathcal{S}_+^{d\times d}$ if

$$\Pi_Q(\lambda) \preceq \frac{\lambda^2 V}{2} \quad \text{for all } \lambda \in \mathbb{R}.$$

# Example

- $Q = \varepsilon B$, where $B$ is a fixed symmetric matrix, and $\varepsilon$ is a Rademacher random variable.
- We have $\mathbb{E}[Q^k] = 0$ for odd $k$, and $\mathbb{E}[Q^k] = B^k$ for even $k$. Therefore, we have

$$\mathbb{E}[e^{\lambda Q}] = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} B^{2k} \le e^{\frac{\lambda^2 B^2}{2}},$$

  implying that $Q$ is sub-Gaussian with parameter $V = \sigma^2 B^2$.
- Now suppose $B$ is a symmetric random matrix, independent of $\varepsilon$, that satisfies $\|B\|_{\mathrm{op}} \le b$.
- Then $\Pi_Q(\lambda) \le \frac{\lambda^2 b^2}{2} I_d$, implying that $Q$ is sub-Gaussian with parameter $V = b^2 I_d$.

# Sub-exponential random matrices and Bernstein condition

### Definition

A zero-mean random matrix is sub-exponential with parameters $(V, b)$ if its cumulant function $\Phi_Q(\lambda) \preceq \frac{\lambda^2 V}{2}$ for all $|\lambda| \leq 1/b$.

The following Bernstein condition for random matrices provides one useful way of certifying the sub-exponential condition.

### Definition

A zero-mean symmetric random matrix $Q$ satisfies a Bernstein condition with parameter $b > 0$ if

$$\mathbb{E}[Q^j] \preceq \frac{1}{2} j!\, b^{j-2} \operatorname{Var}(Q) \quad \text{for } j = 3, 4, \ldots.$$

Similar to the scalar case, the Bernstein condition holds whenever $Q$ has a bounded operator norm, $\|Q\|_{\text{op}} \leq b$. In this case, $\mathbb{E}[Q^j] \preceq b^{j-2} \operatorname{Var}(Q)$.

# Bernstein condition implies sub-exponential condition

## Lemma

*For any symmetric zero-mean random matrix satisfies the Bernstein condition, we have*

$$\Phi_Q(\lambda) \preceq \frac{\lambda^2 \operatorname{Var}(Q)}{1 - b\,|\lambda|} \quad \textit{for all } |\lambda| \leq \tfrac{1}{b}.$$

Proof is similar to the scalar case.

# Matrix-Chernoff approach

**Lemma**

*Let $Q$ be a zero-mean symmetric random matrix whose cumulant function $\Phi_Q$ exists in an open interval $(-a, a)$. Then for any $\delta > 0$, we have*

$$\mathbb{P}[\gamma_{\max}(Q) \geq \delta] \leq \mathrm{Tr}\left(e^{\Phi_Q(\lambda)}\right) e^{-\lambda\delta} \quad \textit{for all } \lambda \in [0, a).$$

*Similarly, we have*

$$\mathbb{P}[\|Q\|_{op} \geq \delta] \leq 2\ \mathrm{Tr}\left(e^{\Phi_Q(\lambda)}\right) e^{-\lambda\delta} \quad \textit{for all } \lambda \in [0, a).$$

Proof is similar to the scalar case.

# Cumulant function of sum of independent matrices

The cumulant function of sum of independent matrices does not decompose additively, because **matrix products need not commute.**

Fortunately, for independent random matrices, it is possible to establish an upper bound in terms of the trace of the cumulant generating functions.

### Lemma

*Let $Q_1, \ldots, Q_n$ be independent symmetric random matrices whose cumulant functions exists for all $\lambda \in I$. Then the sum $S_n = \sum_{i=1}^n Q_i$ satisfies*

$$\mathrm{Tr}\left(e^{\Phi_{S_n}(\lambda)}\right) \le \mathrm{Tr}\left(e^{\sum_{i=1}^n \Phi_{Q_i}(\lambda)}\right) \quad \textit{for all } \lambda \in I.$$

A proof uses Lieb's theorem: for any fixed $H \in \mathcal{S}^{d\times d}$, the following function is concave:

$$A \mapsto \mathrm{Tr}\left(e^{H+\log(A)}\right).$$

# Tail bounds for sub-Gaussian matrices

**Theorem (Hoeffding bound for random matrices)**

*Let $Q_1, \ldots, Q_n$ be independent symmetric random matrices that are sub-Gaussian with parameters $V_1, \ldots, V_n$. Then for any $\delta > 0$, we have*

$$\mathbb{P}\Big[ \| \sum_{i=1}^{n} Q_i \|_{op} \geq \delta \Big] \leq 2\, d\, e^{-\frac{n\delta^2}{2\sigma^2}},$$

*where $\sigma^2 = \| n^{-1} \sum_{i=1}^{n} V_i \|_{op}$.*

This inequality also implies an analogous bound for general independent but potentially non-symmetric and/or non-square matrices in $\mathbb{R}^{d_1 \times d_2}$, with $d$ replaced by $d_1 + d_2$ (why?).