Spring 2018: STA 6448
Advanced Probability and Inference II
Lecture 14

Yun Yang

- Random matrices and covariance estimation

Wishart matrices

- Assume x_i is drawn i.i.d. from a multivariate $\mathcal{N}(0, \Sigma)$ distribution.
- We say X is drawn from a Σ-Gaussian ensemble.
- The sample covariance $\widehat{\Sigma}$ follow a multivariate Wishart distribution.

Theorem (Concentration of Gaussian random matrices)

For each $\delta > 0$, the maximum singular value satisfies

$$\mathbb{P}\left[\frac{\gamma_{\max}(X)}{\sqrt{n}} \geq \gamma_{\max}(\sqrt{\Sigma}\,)(1+\delta) + \sqrt{\frac{\mathrm{Tr}(\Sigma)}{n}}\,\right] \leq e^{-n\delta^2/2}.$$

Moreover, if $n \geq d$, then the minimum singular value satisfies

$$\mathbb{P}\left[\frac{\gamma_{\min}(X)}{\sqrt{n}} \leq \gamma_{\min}(\sqrt{\Sigma}\,)(1-\delta) - \sqrt{\frac{\mathrm{Tr}(\Sigma)}{n}}\,\right] \leq e^{-n\delta^2/2}.$$

Example: Operator norm bounds for the standard Gaussian ensemble

Consider a random matrix $W \in \mathbb{R}^{n \times d}$ with i.i.d. $\mathcal{N}(0,1)$ entries.

This corresponds to $\Sigma = I_d$. The theorem implies that when $n \geq d$,

$$\frac{\gamma_{\max}(W)}{\sqrt{n}} \leq 1 + \delta + \sqrt{\frac{d}{n}}, \quad \text{and} \quad \frac{\gamma_{\min}(W)}{\sqrt{n}} \geq 1 - \delta - \sqrt{\frac{d}{n}}$$

holds with probability at least $1 - 2e^{-n\delta^2/2}$.

These bounds implies that

$$\left\| \frac{1}{n} W^T W - I_d \right\|_{\mathsf{op}} \leq 2\varepsilon + \varepsilon^2, \quad \varepsilon = \delta + \sqrt{\frac{d}{n}},$$

with the same probability.

Example: Gaussian covariance estimation

We reduce the problem to the standard Gaussian ensemble by writing $X = W\sqrt{\Sigma}$, where $W \in \mathbb{R}^{n \times d}$ has i.i.d. $\mathcal{N}(0,1)$ entries.

$$\begin{aligned} \left|\!\left|\!\left| \frac{1}{n}X^T X - \Sigma \right|\!\right|\!\right|_{\mathrm{op}} &= \left|\!\left|\!\left| \Sigma^{1/2}\left(\frac{1}{n}W^T W - I_d\right)\Sigma^{1/2} \right|\!\right|\!\right|_{\mathrm{op}} \\ &\leq \left|\!\left|\!\left| \Sigma \right|\!\right|\!\right|_{\mathrm{op}} \left|\!\left|\!\left| \frac{1}{n}W^T W - I_d \right|\!\right|\!\right|_{\mathrm{op}}. \end{aligned}$$

Consequently,

$$\frac{\left|\!\left|\!\left| \widehat{\Sigma} - \Sigma \right|\!\right|\!\right|_{\mathrm{op}}}{\left|\!\left|\!\left| \Sigma \right|\!\right|\!\right|_{\mathrm{op}}} \leq 2\delta + 2\sqrt{\frac{d}{n}} + \left(\delta + \sqrt{\frac{d}{n}}\right)^2$$

holds with probability at least $1 - 2e^{-n\delta^2/2}$.

Proof: Concentration of Gaussian random matrices

We only prove the upper bound. The proof consists of two steps. Recall $X =$, where $W \in \mathbb{R}^{n \times d}$ has i.i.d. $\mathcal{N}(0, 1)$ entries.

Step one: we use concentration inequalities to argue that the random singular value is close to its expectation with high probability.

Consider the mapping $W \mapsto \gamma_{\max}(W\sqrt{\Sigma})/\sqrt{n}$. It is Lipschitz w.r.t. the Euclidean norm with parameter at most $L = \gamma_{\max}(\sqrt{\Sigma})/\sqrt{n}$. Therefore,

$$\mathbb{P}\left[\gamma_{\max}(X) \geq \mathbb{E}[\gamma_{\max}(X)] + \sqrt{n}\,\gamma_{\max}(\sqrt{\Sigma})\,\delta\right] \leq e^{-n\delta^2}.$$

Proof: Concentration of Gaussian random matrices

Step two: we use Gaussian comparison inequalities to bound the expected value

$$\mathbb{E}[\gamma_{\max}(X)] \le \sqrt{n}\,\gamma_{\max}(\sqrt{\Sigma}) + \sqrt{\mathrm{Tr}(\Sigma)}.$$

We use the variational characterization

$$\gamma_{\max}(X) = \max_{u \in \mathcal{S}^{n-1}} \max_{v \in \mathcal{S}^{d-1}(\Sigma^{-1})} \underbrace{u^T W v}_{Z_{u,v}},$$

where $\mathcal{S}^{d-1}(\Sigma^{-1}) = \{v \in \mathbb{R}^d : \|\Sigma^{-1/2} v\|_2 = 1\}$ is an ellipsoid. $\gamma_{\max}(X)$ is the supremum of the zero-mean GP $Z_{u,v}$.

It can be verified that

$$\mathbb{E}[(Z_{u,v} - Z_{\tilde{u},\tilde{v}})^2] = |||uv^T - \tilde{u}\tilde{v}^T|||_{\mathsf{F}}^2 \le \gamma_{\max}^2(\sqrt{\Sigma})\, \|u - \tilde{u}\|_2^2 + \|v - \tilde{v}\|_2^2.$$

Proof: Concentration of Gaussian random matrices

Define another GP $Y_{u,v}$ by

$$Y_{u,v} = \gamma_{\max}(\sqrt{\Sigma})\langle g,\, u\rangle + \langle h,\, v\rangle,$$

where $g \sim \mathcal{N}(0,\, I_n)$ and $h \sim \mathcal{N}(0,\, I_d)$. Then

$$\mathbb{E}[(Z_{u,v} - Z_{\tilde{u},\tilde{v}})^2] \leq \mathbb{E}[(Y_{u,v} - Y_{\tilde{u},\tilde{v}})^2].$$

We may apply the Sudakov-Fernique bound to obtain

$$\begin{aligned}
\mathbb{E}[\gamma_{\max}(X)] &\leq \mathbb{E}[\max_{u \in \mathcal{S}^{n-1}} \max_{v \in \mathcal{S}^{d-1}(\Sigma^{-1})} Y_{u,v}] \\
&= \gamma_{\max}(\sqrt{\Sigma})\, \mathbb{E}[\|g\|_2] + \mathbb{E}[\|\sqrt{\Sigma}\, h\|_2] \\
&\leq \sqrt{n}\, \gamma_{\max}(\sqrt{\Sigma}) + \sqrt{\mathrm{Tr}(\Sigma)}.
\end{aligned}$$

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 σ^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 σ^2.

Example

- $X \in \mathbb{R}^{n \times d}$ has i.i.d. entries that are zero-mean and sub-Gaussian with parameter σ^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* σ^2*. Then*

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

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

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

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

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

Proof: Concentration of sub-Gaussian ensembles

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

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

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

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

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

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 $\lambda \in [0,\, n/8]$,

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