

6. Random matrices and covariance estimation

担当：みーとみ

2021 年 6 月 30 日, 7 月 7 日

Table of Contents

6.1 Some preliminaries

6.2 Wishart matrices and their behavior

6.1 Some preliminaries

6.2 Wishart matrices and their behavior

- ・ サンプル x_i は, d -次元正規分布 $\mathcal{N}(0, \Sigma)$ から i.i.d. で引かれるとする.
- ・ このとき,

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} \in \mathbb{R}^{n \times d}$$

は, Σ -Gaussian ensemble から引かれると言う.

- ・ Sample covariance $\hat{\Sigma} = \frac{1}{n} X^T X$ は, a multivariate Wishart distribution に従う.

Theorem 6.1

$X \in \mathbb{R}^{n \times d}$ は Σ -Gaussian ensemble から引かれるとする. このとき, 任意の $\delta > 0$ に対し, 最大特異値 $\sigma_{\max}(X)$ は以下の upper deviation inequality を満たす:

$$\mathbb{P} \left[\frac{\sigma_{\max}(X)}{\sqrt{n}} \geq \gamma_{\max} \left(\sqrt{\Sigma} \right) (1 + \delta) + \sqrt{\frac{\text{tr}(\Sigma)}{n}} \right] \leq \exp \left(-\frac{n\delta^2}{2} \right). \quad (6.8)$$

さらに $n \geq d$ なら, 最小特異値 $\sigma_{\min}(X)$ は以下の lower deviation inequality を満たす:

$$\mathbb{P} \left[\frac{\sigma_{\min}(X)}{\sqrt{n}} \leq \gamma_{\min} \left(\sqrt{\Sigma} \right) (1 - \delta) - \sqrt{\frac{\text{tr}(\Sigma)}{n}} \right] \leq \exp \left(-\frac{n\delta^2}{2} \right). \quad (6.9)$$

Example 6.2 (Operator norm bounds for the standard Gaussian ensemble)

- $W \in \mathbb{R}^{n \times d}$ は各成分が $\mathcal{N}(0, 1)$ i.i.d. で引かれる random matrix とする ($\Sigma = I_d$) .
- Thm 6.1 より, $n \geq d$ なら, 確率 $1 - 2 \exp\left(-\frac{n\delta^2}{2}\right)$ 以上で

$$\frac{\sigma_{\max}(W)}{\sqrt{n}} \leq 1 + \delta + \sqrt{\frac{d}{n}} \quad \text{and} \quad \frac{\sigma_{\min}(W)}{\sqrt{n}} \geq 1 - \delta - \sqrt{\frac{d}{n}} \quad (6.10)$$

となる.

- よって, 同じ確率で

$$\left\| \left\| \frac{1}{n} W^T W - I_d \right\|_2 \right\| \leq 2\epsilon + \epsilon^2, \quad \text{where } \epsilon = \sqrt{\frac{d}{n}} + \delta. \quad (6.11)$$

- したがって, $d/n \rightarrow 0$ なら, sample covariance $\hat{\Sigma} = \frac{1}{n} W^T W$ は identity matrix I_d の一致推定量となる. ♣

Example 6.3 (Gaussian covariance estimation)

- $X \in \mathbb{R}^{n \times d}$ は Σ -Gaussian ensemble からの random matrix とする.
- このとき $X = W\sqrt{\Sigma}$ と書ける ($W \in \mathbb{R}^{n \times d}$ は standard Gaussian random matrix) ので,

$$\left\| \left\| \frac{1}{n} X^T X - \Sigma \right\| \right\|_2 = \left\| \left\| \sqrt{\Sigma} \left(\frac{1}{n} W^T W - I_d \right) \right\| \right\|_2 \leq \|\Sigma\|_2 \left\| \left\| \frac{1}{n} W^T W - I_d \right\| \right\|_2.$$

- したがって (6.11) より, 任意の $\delta > 0$ に対して確率 $1 - 2 \exp\left(-\frac{n\delta^2}{2}\right)$ で

$$\frac{\|\hat{\Sigma} - \Sigma\|_2}{\|\Sigma\|_2} \leq 2\sqrt{\frac{d}{n}} + 2\delta + \left(\sqrt{\frac{d}{n}} + \delta\right)^2. \quad (6.12)$$

- よって, $\|\hat{\Sigma} - \Sigma\|_2 / \|\Sigma\|_2$ は $d/n \rightarrow 0$ である限り 0 に収束する.



Example 6.4 (Faster rates under trace constraints)

- $\{\gamma_j(\Sigma)\}_{j=1}^d$ は Σ の固有値列で, $\gamma_1(\Sigma)$ がそのうち最大のもの.
- Σ は, 次元に対して独立な定数 C に対し, 次の “trace constraint” を満たすとする:

$$\frac{\text{tr}(\Sigma)}{\|\Sigma\|_2} = \frac{\sum_{j=1}^d \gamma_j(\Sigma)}{\gamma_1(\Sigma)} \leq C. \quad (6.13)$$

- C は Σ の (実質的な) rank と見なせる (\because (6.13) は $C = \text{rank}(\Sigma)$ では常に成立.)
- パラメータ $q \in [0, 1]$ と半径 $R_q > 0$ の the Schatten q -“balls” を, 以下で定義する:

$$\mathbb{B}_q(R_q) := \left\{ \Sigma \in S^{d \times d} \left| \sum_{j=1}^d |\gamma_j(\Sigma)|^q \leq R_q \right. \right\}. \quad (6.14)$$

- $q = 0$ なら, rank R_q 以下の対称行列の集合.
- $q = 1$ なら, trace constraint になる.
- 任意の非零行列 $\Sigma \in \mathbb{B}_q(R_q)$ は, (6.13) を $C = R_q/(\gamma_1(\Sigma))^q$ で満たす.

- (6.13) を満たす任意の Σ に対し, Thm 6.1 は高確率で X の最大特異値が次のように抑えられることを保証する:

$$\frac{\sigma_{\max}(X)}{\sqrt{n}} \leq \gamma_{\max}(\sqrt{\Sigma}) \left(1 + \delta + \sqrt{\frac{C}{n}} \right). \quad (6.15)$$

- $\Sigma = I_d$ のときの bound (6.10) と比べると, C が d に置き換わって “実行的な rank” となっている.



Proof of Theorem 6.1.