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

- Notation とこの章で使う preliminary results の説明から.

- 行列 $A \in \mathbb{R}^{n\times m}$ with $n \geq m$ に対し,(順序付き)特異値を
$$\sigma_{\max}(A) = \sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_m(A) = \sigma_{\min}(A) \geq 0$$
と書く.
- 最小・最大特異値は次のように characterize される:
$$\sigma_{\max}(A) = \max_{\nu\in\mathbb{S}^{m-1}} \|A\nu\|_2 \quad \text{and} \quad \sigma_{\min}(A) = \min_{\nu\in\mathbb{S}^{m-1}} \|A\nu\|_2, \tag{6.1}$$
ただし $\mathbb{S}^{d-1} := \left\{\nu \in \mathbb{R}^d \mid \|\nu\|_2 = 1\right\}$ は $\mathbb{R}^d$ 上の Euclidean unit sphere.
- また次の同値性が成り立つ: $|||A|||_2 = \sigma_{\max}(A)$.

- 対称行列の集合を $\mathcal{S}^{d\times d} := \left\{Q \in \mathbb{R}^{d\times d} \mid Q = Q^{\mathrm{T}}\right\}$ とし, その半正定値行列からなる部分集合を
$$\mathcal{S}_{+}^{d\times d} := \left\{Q \in \mathcal{S}^{d\times d} \mid Q \succeq 0\right\} \tag{6.2}$$
と書く.
- 任意の対称行列 $Q \in \mathcal{S}^{d\times d}$ は対角化可能であり, その固有値を
$$\gamma_{\max}(Q) = \gamma_1(Q) \geq \gamma_2 \geq \cdots \geq \gamma_d(Q) = \gamma_{\min}(Q)$$
とする.
- このとき, $Q \succeq 0 \iff \gamma_{\min}(Q) \geq 0$.

- 最小・最大固有値の “Rayleigh – Ritz variational characterization”:

$$\gamma_{\max}(Q) = \max_{\nu \in \mathbb{S}^{d-1}} \nu^{\mathrm{T}} Q \nu \quad \text{and} \quad \gamma_{\min}(Q) = \min_{\nu \in \mathbb{S}^{d-1}} \nu^{\mathrm{T}} Q \nu. \tag{6.3}$$

- 任意の対称行列 Q に対し, その ℓ_2-operator norm は,

$$|||Q|||_2 = \max\{\gamma_{\max}(Q), |\gamma_{\min}(Q)|\} = \max_{\nu \in \mathbb{S}^{d-1}} \left|\nu^{\mathrm{T}} Q \nu\right|. \tag{6.4}$$

- 最後に, 行列 $A \in \mathbb{R}^{n \times m}$ with $n \geq m$ に対し, m-次元対称行列 $R := A^{\mathrm{T}} A$ を考えると,

$$\gamma_j(R) = (\sigma_j(A))^2 \quad \text{for } j = 1, \ldots, m.$$

- $\{x_1, \ldots, x_m\}$ は, $\mathbb{R}^d$ 上の zero-mean・covariance $\Sigma = \mathrm{cov}(x_1) \in \mathbb{S}_+^{d\times d}$ なる分布からの n 個の i.i.d. サンプルとする.
- Σ の standard estimator は, 次の *sample covariance matrix* である:

$$\widehat{\Sigma} := \frac{1}{n}\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}. \tag{6.5}$$

- 各 x_i は zero-mean なので $\mathbb{E}[x_i x_i^{\mathrm{T}}] = \Sigma$ であり, $\widehat{\Sigma}$ は Σ の unbiased estimator.
- したがって $\widehat{\Sigma} - \Sigma$ は期待値ゼロとなり, その ℓ_2-operator norm によって測った error の bound を求めることがこの章の goal となる.

- (6.4) の ℓ_2-operator norm の表現より, $|||\widehat{\Sigma} - \Sigma|||_2 \leq \epsilon$ は以下と同値:

$$\max_{\nu \in \mathbb{S}^{d-1}} \left| \frac{1}{n} \sum_{i=1}^{n} \langle x_i, \nu_i \rangle^2 - \nu^{\mathrm{T}} \Sigma \nu \right| \leq \epsilon. \tag{6.6}$$

- つまり, $|||\widehat{\Sigma} - \Sigma|||_2$ をコントロールすることは, ν で indexed された関数クラス $x \mapsto \langle x, \nu \rangle^2$ の uniform law of large numbers を示すことと同値になる.

- その ℓ_2-operator norm をコントロールすることは, $\widehat{\Sigma}$ の固有値の一様収束も意味する: Weyl's theorem の corollary より,

$$\max_{j=1,\dots,d} \left| \gamma_j(\widehat{\Sigma}) - \gamma_j\left(\Sigma\right) \right| \leq |||\widehat{\Sigma} - \Sigma|||_2. \tag{6.7}$$

- また最後に, ランダム行列 $X \in \mathbb{R}^{n \times d}$ が, 第 i 行に x_i^{T} を持つものとする

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

と,

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^{\mathrm{T}} = \frac{1}{n} X^{\mathrm{T}} X$$

なので, $\widehat{\Sigma}$ の固有値は $X/\sqrt{n}$ の特異値の 2 乗となる.

- 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,

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

  は, Σ-*Gaussian ensemble* から引かれると言う.
- Sample covariance $\widehat{\Sigma} = \frac{1}{n} X^{\mathrm{T}} X$ 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:

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

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:

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

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

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

  となる.
- よって, 同じ確率で

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

- したがって, $d/n \to 0$ なら, sample covariance $\widehat{\Sigma} = \frac{1}{n} W^{\mathrm{T}} W$ は identiry 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,

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

- 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

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

- よって, $|||\widehat{\Sigma} - \Sigma|||_2 / |||\Sigma|||_2$ は $d/n \to 0$ 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:

$$\frac{\mathrm{tr}(\Sigma)}{|||\Sigma|||_2} = \frac{\sum_{j=1}^d \gamma_j(\Sigma)}{\gamma_1(\Sigma)} \le C. \tag{6.13}$$

- C は Σ の（実質的な）rank と見なせる（$\because$ (6.13) は $C = \mathrm{rank}(\Sigma)$ 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:

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

  - $q = 0$ 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:

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

- $\Sigma = I_d$ 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.

-