随机矩阵理论的研究（一）：元素独立随机矩阵

朱梦

初稿于 2025-06-12，修改于 2025-06-16

1. 奇异值与奇异向量

（一）奇异值分解（SVD）。对于任意矩阵 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$，可分解为：

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^* \tag{1}$$

其中，$\boldsymbol{U} \in \mathbb{C}^{m\times m}$ 表示左奇异向量组成的正交矩阵，对应 $\boldsymbol{A}\boldsymbol{A}^*$ 的特征向量，$\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ 表示对角矩阵，对角线元素为奇异值 $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r$（$r$ 为矩阵的秩），$\boldsymbol{V} \in \mathbb{C}^{n\times n}$ 表示右奇异向量组成的正交矩阵，对应 $\boldsymbol{A}^*\boldsymbol{A}$ 的特征向量。

（二）奇异值的性质。（1）非负性：$\sigma_i \geq 0$。（2）降序排列：$\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r$。

（三）奇异向量的性质。（1）正交归一性：$\boldsymbol{U}$ 的列向量两两正交且模长为 1，$\boldsymbol{V}$ 的列向量两两正交且模长为 1。（2）方向性：左奇异向量指向数据的主成分方向，右奇异向量指向原始特征的主方向。

（四）几何解释。SVD 将矩阵变换分解为旋转（$\boldsymbol{V}^{\mathrm{T}}$）、缩放（$\boldsymbol{\Sigma}$）和再旋转（$\boldsymbol{U}$）三个步骤。奇异值越大，对应方向的“拉伸”效果越显著。

2. 特征值与特征向量

（一）特征值与特征向量的定义。对于方阵 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，若存在标量 $\lambda \in \mathbb{C}$ 和非零向量 $\boldsymbol{x} \in \boldsymbol{C}^n$，使得：

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} \tag{2}$$

则称：λ 为 $\boldsymbol{A}$ 的特征值，$\boldsymbol{x}$ 为 $\boldsymbol{A}$ 的对应于 λ 的特征向量。

（二）特征值的代数性质。（1）特征值之和等于矩阵的迹（对角线元素之和）。特征值之积等于矩阵的行列式。（2）不同特征值对应的特征向量线性无关。重根特征值可能对应线性相关特征向量。（3）实对称矩阵的特征值为实数，且不同特征值对应的特征向量正交。

（三）特征值的几何行为。当 λ 为实数，矩阵 $\boldsymbol{A}$ 在特征向量方向上的变换为纯拉伸或压缩，无旋转成分。当 λ 为纯虚数，矩阵 $\boldsymbol{A}$ 在特征向量方向上的变换为纯旋转，无拉伸或压缩。当 λ 为实数，矩阵 $\boldsymbol{A}$ 在特征向量方向上的变换结合了拉伸/压缩与旋转。

3. 谱范数与谱半径

谱范数为矩阵的一种诱导范数，定义为矩阵的最大奇异值：

$$||\boldsymbol{A}||_2 = \sigma_{\text{max}}(\boldsymbol{A}) = \sqrt{\lambda_{\text{max}}(\boldsymbol{A}^*\boldsymbol{A})} \tag{3}$$

其中，$\sigma_{\text{max}}(\boldsymbol{A})$ 表示矩阵 $\boldsymbol{A}$ 的最大奇异值，$\sqrt{\lambda_{\text{max}}(\boldsymbol{A}^*\boldsymbol{A})}$ 表示矩阵 $\boldsymbol{A}^*\boldsymbol{A}$ 的最大特征值。对于线性变换 $\boldsymbol{Ax}$，谱范数 $||\boldsymbol{A}||_2$ 表示向量 $\boldsymbol{x}$ 在变换后可能被拉伸的最大倍数，即 $||\boldsymbol{Ax}||_2 \leq ||\boldsymbol{A}||_2||\boldsymbol{x}||_2$。

谱半径是矩阵所有特征值的模的最大值：

$$\rho(\boldsymbol{A}) = \max_i |\lambda_i| \tag{4}$$

其中，λ_i 表示是矩阵 $\boldsymbol{A}$ 的特征值。

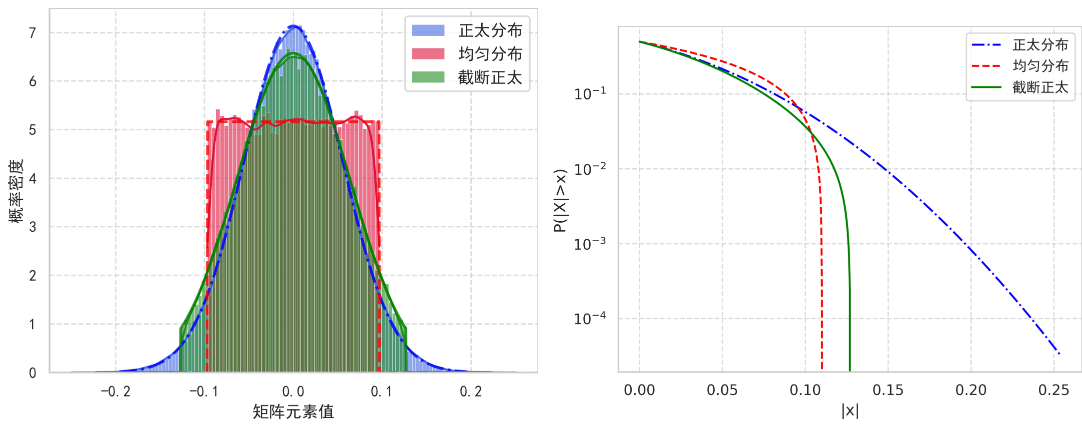
对于任意复数矩阵 $\boldsymbol{A}$，都有：

$$\rho(\boldsymbol{A}) \leq ||\boldsymbol{A}||_2 \tag{5}$$

当且仅当矩阵 $\boldsymbol{A}$ 为正规矩阵（即 $\boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^*$）时，谱半径等于谱范数：

4. 独立元素随机矩阵

对于一个随机矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，其元素独立重复采样于均值为 0，方差为 $2/(m+n)$ 的随机分布（随机分布可以是正太分布、均匀分布或截断正太分布）。图1展示了从三类随机分布采样所得到矩阵的元素分析及其比较。



(a) 直方图　　(b) 尾部行为（对数坐标）

图 1　从三类随机分布采样所得到矩阵的元素分析及其比较

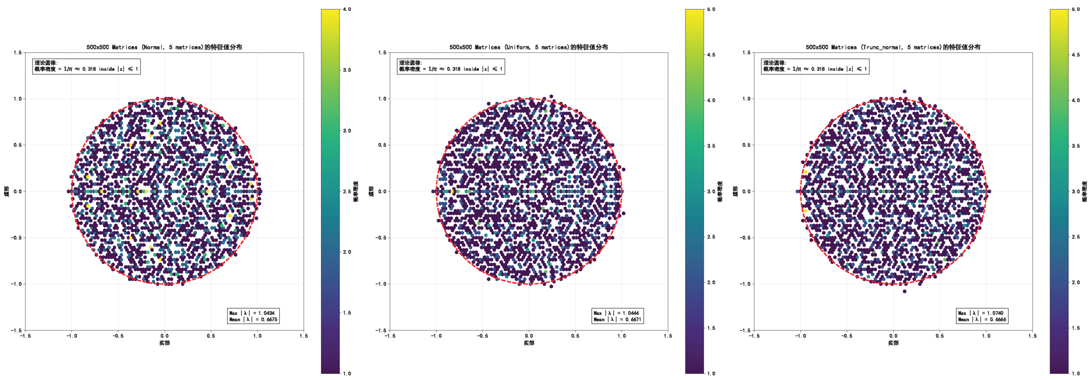
4.1　独立元素随机方阵

当 $m = n$ 时，矩阵为方阵。当 n 很大时，其特征值分布服从圆律（Circular Law），即当 n 很大时，特征值在复平面上均匀分布在以原点为中心、半径为 1 的圆盘内：

$$|\lambda_i| \leq 1, \lambda_i \in \mathbb{C} \tag{6}$$

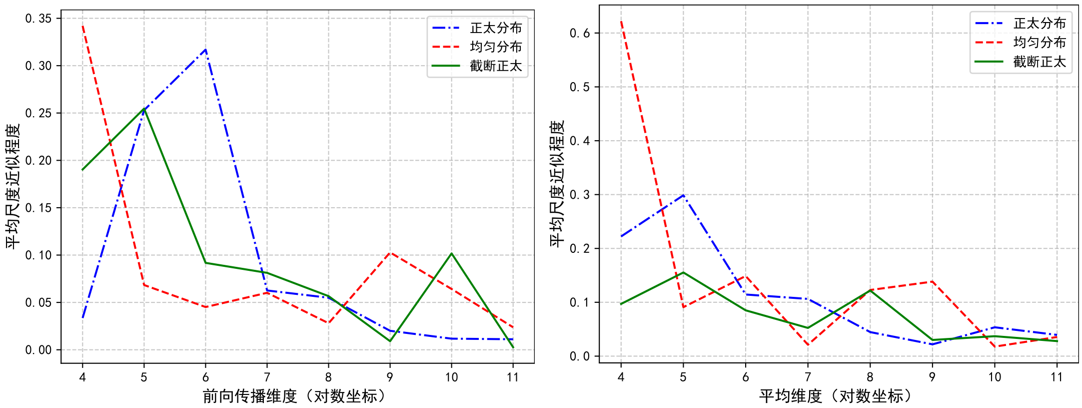
$$f_{\text{pdf}}(\lambda_i) = \begin{cases} \frac{1}{\pi}, & \text{如果} |\lambda_i| \leq 1, \\ 0, & \text{其它} \end{cases} \tag{7}$$

图2展示了从不同随机分布采样所得到方阵的特征值分布，可知：特征值在复平面上均匀分布在以原点为中心、半径为 1 的圆盘内。



(a) 正太分布　　(b) 均匀分布　　(c) 截断正太

图 2　从不同随机分布采样所得到方阵的特征值分布（服从圆率）

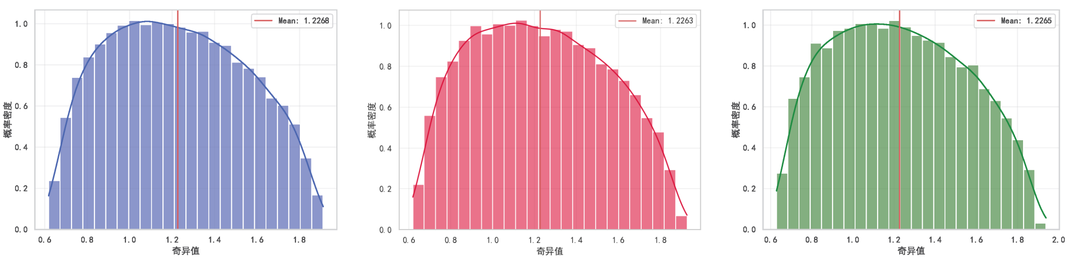


(a) 前向传播维度 *vs.* 平均尺度近似程度　　(b) 平均维度 *vs.* 平均尺度近似程度

图 3　方阵维度（对数坐标）*vs.* 平均尺度近似程度

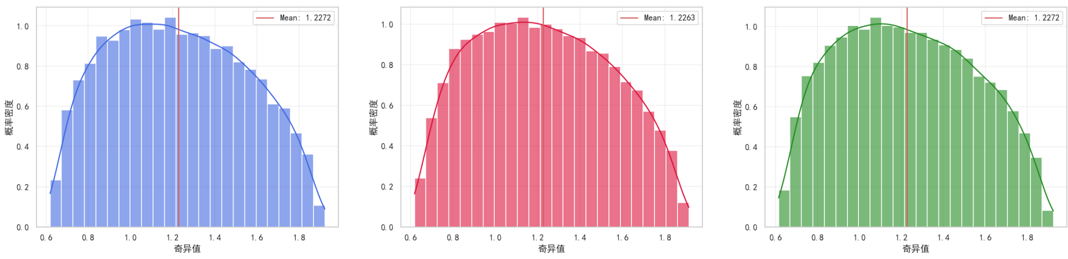
4.2　独立元素随机非方阵

当 $m \neq n$ 时，非方阵没有传统意义上的特征值，但可以分析其奇异值。图4展示了从不同随机分布采样所得到非方阵的奇异值分布，可知：奇异值分布收敛至截断正太分布，且平均值均收敛至 1.2 左右。图5展示了从不同随机分布采样所得到非方阵转置的奇异值分布，可知：奇异值分布也收敛至截断正太分布，且平均值均也收敛至 1.2 左右。



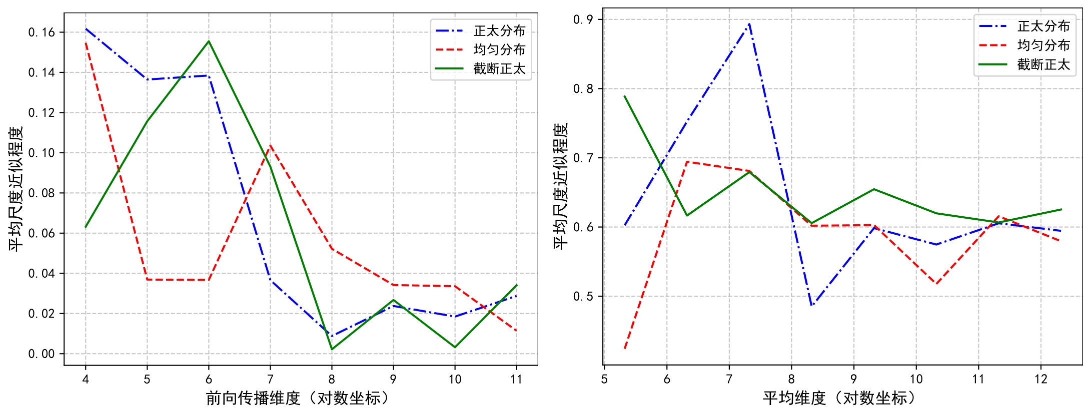
(a) 正太分布　　(b) 均匀分布　　(c) 截断正太

图 4　从不同随机分布采样所得到非方阵的奇异值分布



(a) 正太分布　　(b) 均匀分布　　(c) 截断正太

图 5　从不同随机分布采样所得到非方阵转置的奇异值分布



(a) 前向传播维度 *vs.* 平均尺度近似程度　　(b) 平均维度 *vs.* 平均尺度近似程度

图 6　非方阵维度（对数坐标）*vs.* 平均尺度近似程度