§2 随机矩阵初步 & 信息熵
第 2&2.5 次作业，2023.06.05，本次作业 ddl，2023.06.18

1 习题解答

1. GUE 矩阵 (Gaussian Unitary Ensemble) 指复矩阵 $\mathrm{X}=(x_{ij})_{i,j=1}^n$, 这里 $\{\mathrm{Re}x_{ij},\mathrm{Im}x_{ij}: i,j=1,\ldots,n\}$ 为独立同标准正态分布. 令

$$H=\frac{1}{2}(X+X^*), \tag{1}$$

试证

(i) H 矩阵元联合密度为

$$f(H)=2^{-\frac{n}{2}}\pi^{-\frac{n^2}{2}}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}H^2}\ . \tag{2}$$

(ii) (酉群不变性) 任给酉矩阵 $\mathrm{Q}\in\mathrm{U}(n)$ 有 QHQ^{-1} 与 H 同分布.

Proof. 作业中很多是通过证明 QX 和 X 同分布（左乘右乘一个正交矩阵 X 分布不变），进而得到 QHQ^{-1} 是同分布的。另一方面有一部分同学用 $trH^2=tr(QHQ^{-1})^2$ 来说明，这是有问题的，换元时同时需要考虑 Jacobi 矩阵。

另外，由于我们考虑的是多重正态分布，他的分布完全由协方差矩阵所确定了，所以实际上我们只需要计算换元之后的协方差。设 $QHQ^{-1}=(a_{ij})$，由于 $Q^{-1}=\bar{Q}^t$ 则

$$a_{ij}=\sum_{k_1,k_2}q_{ik_1}a_{k_1k_2}\bar{q}_{jk_2}. \tag{3}$$

考虑协方差

$$\begin{aligned}Cov(a_{ij},a_{st})&=\mathbb{E}(\sum_{k_1,k_2}q_{ik_1}h_{k_1k_2}\bar{q}_{jk_2})(\sum_{k_3,k_4}q_{sk_3}h_{k_3k_4}\bar{q}_{tk_4})\\&=\sum_{k_1,k_2,k_3,k_4}\mathbb{E}(q_{ik_1}\bar{q}_{jk_2}q_{sk_3}\bar{q}_{tk_4}h_{k_1k_2}h_{k_3k_4})\end{aligned} \tag{4}$$

和式里面的项只有在 $(k_1,k_2)=(k_4,k_3)$ 的时候才不为 0。所以上式等于

$$\begin{aligned}Cov(a_{ij},a_{st})&=\sum_{k_1,k_2,k_3,k_4}\mathbb{E}(q_{ik_1}\bar{q}_{jk_2}q_{sk_3}\bar{q}_{tk_4}h_{k_1k_2}h_{k_3k_4})\\&=\sum_{k_1\neq k_2}q_{ik_1}\bar{q}_{tk_1}\bar{q}_{jk_2}q_{sk_2}+\sum_{k_1}q_{ik_1}\bar{q}_{tk_1}\bar{q}_{jk_1}q_{sk_1}\\&=(\sum_{k_1}q_{ik_1}\bar{q}_{tk_1})(\sum_{k_2}\bar{q}_{jk_1}q_{sk_1})=\delta_{it}\delta_{js}.\end{aligned} \tag{5}$$

同样再考虑

$$Cov(a_{ij},\bar{a}_{st})=\delta_{is}\delta_{jt}. \tag{6}$$

即可证明不变。 □

2. 若 X 与 Y 独立，(X, Y) 旋转不变，即 $\forall\theta\in[0,2\pi)$，$\mathrm{e}^{i\theta}(\mathrm{X}+i\mathrm{Y})$ 与 $\mathrm{X}+i\mathrm{Y}$ 同分布，试证 X 与 Y 必为零均值且有相同方差正态分布。

   （提示：可先在 X 与 Y 高阶矩存在时验证此结论。）

   Proof. 我们考虑 (X,Y) 的特征函数

$$\begin{aligned}\phi_{(X,Y)}(s,t) &= \mathbb{E}_{(X,Y)}(e^{isX+itY})\\ &= \mathbb{E}_X(e^{isX})\mathbb{E}_Y(e^{itY})\\ &= \phi_X(s)\phi_Y(t). \quad (X,Y)\text{是独立随机变量}\end{aligned}\tag{7}$$

   由 (X,Y) 的旋转不变性，旋转 90 度，我们知道

$$(X,Y)\overset{d}{=}(-Y,X)\Rightarrow\phi_{(X,Y)}(s,t)=\phi_{(-Y,X)}(s,t).\tag{8}$$

   也就是

$$\phi_X(s)\phi_Y(t)=\phi_Y(-s)\phi_X(t).\tag{9}$$

   取 $s=0$ 我们有 $\phi_Y(t)=\phi_X(t)$，取 $t=0$ 得到 $\phi_{-X}(t)=\phi_X(t)$。也就是说 X,Y 是同分布，对称（从而均值为 0）的随机变量。现在我们进一步运用旋转不变性，

$$\begin{aligned}\phi_X(s)\phi_X(t) &= \phi_{(X,Y)}(s,t)\\ &= \phi_{\left(\frac{sX+tY}{\sqrt{s^2+t^2}},\frac{tX-sY}{\sqrt{s^2+t^2}}\right)}(s,t)\\ &= \mathbb{E}\left(e^{\frac{i}{\sqrt{s^2+t^2}}(s(sX+tY)+t(tX-sY))}\right)\\ &= \phi_X(\sqrt{s^2+t^2}).\end{aligned}\tag{10}$$

   这实际上就是 Cauchy equation(Wiki)的变种。我们考虑变量大于 0 的情况，首先我们有

$$\phi_X(\sqrt{2}s)=\phi_X(\sqrt{s^2+(-s)^2})=\phi_X(s)\phi_X(-s)=|\phi_X(s)|^2\ge 0.\tag{11}$$

   当然不可能会等于 0，假设 $\phi_X(s)=0$，那么对任意的 $t>s$ 有 $\phi_X(t)=\phi_X(\sqrt{t^2-s^2})\phi_X(s)=0$. 再令 s 是使得 $\phi_X(s)=0$ 成立的最小 s，那么 $0=\phi_X(s)=\phi_X(s/\sqrt{2})^2=0$，矛盾。

   从而，我们可以取对数，令 $g(t)=\log\phi_X(\sqrt{t})$，我们有

$$g(t)+g(s)=g(t+s).\tag{12}$$

   从而 $g(t)=ct$，$\phi_X(s)=e^{-c^2s^2}$，也就是正态分布的特征函数。 □

3. Wishart 模型

   $X=(x_{ij})_{1\le i\le p,1\le j\le n}$，矩阵元 $\{x_{ij}\}$ 为独立同 N(0, 1)，对固定的 $\alpha=n-p\ge 0$，证明当 $p\to\infty$ 时

$$\frac{1}{p}\mathbb{E}\left[\operatorname{tr}\left(\frac{\mathrm{XX}^t}{p}\right)^m\right]\to C_m:=\frac{1}{m+1}\binom{2m}{m}.\tag{13}$$

   如果困难的话，只需验证 $m=2$ 特殊情形即可.

Proof. 这个我们之后证明。 □

4. 实对称 Wigner 矩阵 $A_n=(a_{ij})_{i,j=1}^n$ 满足条件:
(i)$\{a_{ij}:1\le i\le j\le n\}$ 相互独立，$\{a_{ij}:1\le i<j\le n\}$ 同分布，$\{a_{ii}:1\le i\le n\}$ 同分布; (ii)$\mathbb{E}(a_{11})=0,\mathbb{E}(a_{12})=0,\ \mathrm{Var}(a_{11})<\infty,\mathrm{Var}(a_{12})=1$; (iii) a_{11} 和 a_{12} 的各阶矩均存在. 令

$$\| A_n \|_2=\sup_{\|\vec{v}\|=1,\vec{v}\in\mathbb{R}^n} \| A_n\vec{v} \|, \tag{14}$$

试证明 $\forall\delta>0$ 有

$$\lim_{n\to\infty} \mathrm{P}\left(\| A_n \|_2\ge n^{\frac{1}{2}+\delta}\right)=0\ . \tag{15}$$

提示：利用 $\| A_n \|_2$ 与 A_n 特征值之间的关系.

Proof. 考虑到

$$\max_i \lambda_i^{2k}\le\sum_i\lambda_i^{2k}=trA_n^{2k}, \tag{16}$$

我们有:

$$\begin{aligned}\mathbb{P}\left(\| A_n \|_2\ge n^{\frac{1}{2}+\delta}\right)&=\mathbb{P}(\max_i|\lambda_i|^{2k}\ge n^{k(\frac{1}{2}+\delta)})\\&\le\mathbb{P}(trA_n^{2k}\ge n^{2k(\frac{1}{2}+\delta)})\\&\le\frac{\mathbb{E}(trA_n^{2k})}{n^{k+2k\delta}}.\end{aligned} \tag{17}$$

而我们知道 $trA_n^{2k}\sim C_kn^{k+1}$，故取 k 使得 $2k\delta>1$ 即可. □

5. (选做) 厄米 Wigner 矩阵 $A_n=(a_{ij})_{i,j=1}^n$ ，$A_n=A_n^*$.

- ▷ 实 $\{a_{ii}\}$ 独立同，与 Y 同分布;
- ▷ $\{\mathrm{Re}a_{ij}\ ,\mathrm{Im}a_{ij}\}_{i<j}$ 独立同，与 Z 同分布;
- ▷ $\{a_{ii}:1\le i\le n\}\cup\{\mathrm{Re}a_{ij},\mathrm{Im}a_{ij}:\ 1\le i<j\le n\}$ 独立;
- ▷ $\mathbb{E}[\mathrm{Y}]=0$, $\mathbb{E}[\mathrm{Z}]=0$, $\mathrm{var(Y)}<\infty$, $\mathrm{var(Z)}=\frac{1}{2}$;
- ▷ $\forall k\ge3$, $\mathbb{E}[|\mathrm{Y}|^k]$,$\mathbb{E}[|\mathrm{Z}|^k]<\infty$.

试证明 Wigner 半圆律同样成立，即有

$$\frac{1}{n}\mathbb{E}\left[\mathrm{tr}\left(\frac{A_n}{\sqrt{n}}\right)^k\right]\to\gamma_k=\int x^k\frac{1}{2\pi}\sqrt{1-x^2}\mathrm{d}x\ . \tag{18}$$

Proof. 这个我们之后证明。 □

6. 掷 2 枚均匀骰子，X 表示点数之和，求熵 $H(X)$.

Proof. 直接写出骰子点数和的分布列6，从而可以直接计算熵 $H(X)$. □

Table 1: $X+Y$ 的分布列

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

7. 离散型随机变量 X，$g:\mathbb{R}\to\mathbb{R}$ 为 Borel 可测，试证明 $H(g(X))\le H(X)$.

Proof.

$$\begin{aligned}H(g(X)) &= -\sum_{i\in I}\mathbb{P}(g(X)=i)\log\mathbb{P}(g(X))\\ &= -\sum_{i}\sum_{X:g(X)=i}\mathbb{P}(X)\log(\sum_{X:g(X)=i}\mathbb{P}(X))\\ &\le -\sum_{i}\sum_{X:g(X)=i}\mathbb{P}(X)(\sum_{X:g(X)=i}\log\mathbb{P}(X))\\ &= \sum_{X}\mathbb{P}(X)\log\mathbb{P}(X).\end{aligned}\tag{19}$$

□

8. 验证 Thm 4 (2)(3) 结论。

(2). 设 X 是取值在 $[0,\infty)$ 上的随机变量，满足 $\mathbb{E}(X)=\frac{1}{\lambda}$，求证指数分布在这些限制条件下取得熵的最大值，为 $\ln\frac{e}{\lambda}$.

Proof. 用 Gibbs 不等式，若 $f(x),\ g(x)\ge 0,\ \int f(x)=1,\int g(x)=1$ 则

$$-\int f(x)\log f(x)\le -\int f(x)\log g(x).\tag{20}$$

取 $g(x)=\lambda e^{-\lambda x}$. 则有

$$-\int f(x)\log f(x)\le -\int f(x)\log(\lambda e^{-\lambda x})=\log\frac{e}{\lambda}.\tag{21}$$

□

(3). 设 X 是取值在 $[0,a]$ 上的随机变量，求证 X 为均匀分布时熵最大，为 $\ln a$.

Proof. 用 Gibbs 不等式，取 $g(x)=\frac{1}{a}1_{\{x\in[0,a]\}}$, 我们有：

$$-\int f(x)\log f(x)\le\int f(x)\log a=\log a.\tag{22}$$

□

9. 有限样本空间 Ω，$\mathcal{M}_1(\Omega)$ 表示 Ω 上概率分布构成集合，对 $\mu \in \mathcal{M}_1(\Omega)$，记熵

$$S(\mu) := -\sum_{\omega\in\Omega} \mu(\omega)\log\mu(\omega), \tag{23}$$

试证，$\forall \mu,\nu \in \mathcal{M}_1(\Omega)$, $\alpha \in [0,1]$ 有

$$S(\alpha\mu + (1-\alpha)\nu) \geq \alpha S(\mu) + (1-\alpha)S(\nu). \tag{24}$$

Proof. 这道题说的是熵 $H(\mu)$ 是一个凹的泛函，我们直接验证就好了。实际上可以逐点验证，易知 $-x\log x$ 是凹函数，从而我们有

$$-(\alpha\mu(\omega) + (1-\alpha)\nu(\omega)) \geq \alpha(-\mu(\omega)\log\omega) + (1-\alpha)(-\nu(\omega)\log\nu(\omega)). \tag{25}$$

两边对 ω 求和就得到

$$S(\alpha\mu + (1-\alpha)\nu) \geq \alpha S(\mu) + (1-\alpha)S(\nu). \tag{26}$$

$\square$

2 随机矩阵的 k 阶矩

这一节我们考虑作业题中的 3 和 5，分别是 Wishart 矩阵和 Wigner 矩阵，首先我们还是先浅浅介绍一下这两个矩阵模型的来源和背景。

定义 2.1. 定义 X_n 为 $p(n)\times n(p(n)\leq n)$ 阶矩阵，满足 X_{ij} 为 *i.i.d.* 随机变量，$E(X_{ij})=0$，$E(X_{ij}^2)=1$，$E(|X_{ij}|^k)<\infty$ 对 $k\geq 3$. 定义

$$S_n = \frac{1}{n}XX^T \in \mathbb{R}^{p\times p}. \tag{27}$$

Wishart 矩阵定义来自于统计，当我们考虑高维数据（n 维）时每一分量的数据都会和真实数据相比有一定的误差和扰动，假设我们从中抽样了 $p(n)$ 组数据出来，我们能多大程度还原出真实的数据?

答案当然永远是真实数据比较“显著”之时我们能够恢复它，但是多显著才是“显著”就需要对扰动的大小进行估计。也就是对 Wishart 矩阵的特征值分布进行估计，我们有如下结果。

令

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \tag{28}$$

为 S_n 的特征值，我们定义他的谱测度为

$$\mu_n = \frac{1}{p(n)}\sum_{i=1}^{n}\delta_{\lambda_i}. \tag{29}$$

则我们有它的经验谱测度收敛到 Marchenko-Pastur law.

定理 2.1 (Marchenko-Pastur law)**.** S_n, μ_n 如上定义，设 $p/n \overset{n\to\infty}{\to} \alpha \in (0,1]$. 则我们有

$$\mu_n(\cdot,\omega) \Rightarrow \mu \quad a.s. \tag{30}$$

其中，$\Rightarrow$ 指弱收敛，μ 如下给出：

$$\frac{d\mu}{dx} = \frac{1}{2\pi\alpha x}\sqrt{(\lambda_+ - x)(x-\lambda_-)}1_{\lambda_- \le x \le \lambda_+} \tag{31}$$

并且

$$\lambda_+ = (1+\sqrt{\alpha})^2,\ \ \lambda_- = (1-\sqrt{\alpha})^2. \tag{32}$$

一些参数时的曲线形状可参看图2。我们作业题的第三题实际上就是 $p/n \to 1$ 的情况。

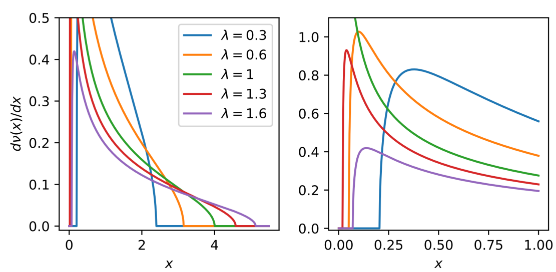

Figure 1: Marchenko-Pastur law 不同参数时的取值，它的 λ 就是我们的 α，当 $\lambda > 1$ 时会有很多特征值是 0，图片来源:wiki

接下来我们介绍一下 Wigner 矩阵的来源于背景，虽然都同样是“随机” + “矩阵”但是它和 Wishart 矩阵来源完全不同。Wigner 矩阵来源于量子多体系统，如果大家对量子力学有些了解的话，其实会知道量子力学中最基本的 Schrödinger 方程可以写成如下形式

$$i\hbar\frac{d}{dt}|\Psi(t)>= H|\Psi(t)> . \tag{33}$$

其中 $|\Psi>$ 为 Hilbert 空间的一个态矢量，我们可以理解成一个有限高维空间里的向量，H 是系统的 Hamiltonian，一个线性算子，我们可以理解成一个高维的矩阵。对于复杂系统我们无法写出 Hamiltonian 的具体形式，常用的办法便是取个随机的矩阵进行模拟发现和实际物理观测的一些现象拟合地很好，这就是随机矩阵的来源之一。物理中的 Hamiltonian 都是一个 Hermite 算子，对应到随机矩阵就是对称或者 Hermite 矩阵，实数时就是实 Wigner 矩阵。

这两种随机矩阵整体特征值的谱分布都可以用矩方法 (moment method) 来进行，也就是课堂上讲的考虑

$$\frac{1}{n}\mathrm{Tr}(A^k) = \frac{1}{n}\sum_{i=1}^n \lambda_i^k = \mu_n(x^k). \tag{34}$$

来进行证明，不过我们只说明 $\mathbb{E}\mu_n(x^k) = \frac{1}{n}\mathbb{E}(\mathrm{Tr}(A^k))$ 会收敛过去，要证明 $a.s.$ 收敛还需要估计 $\mathrm{Var}(\mathrm{Tr}(A^k))$. 不过这些内容都可以参考这篇本科毕业论文和这个lecture notes.

故事讲完了，我们开始介绍严格的数学证明。先考虑 5 中的厄米 Wigner 矩阵

proof of 5. 我们考虑 $\mathbb{E}\mathrm{Tr}A_n^k$，首先我们展开

$$\mathrm{Tr}A_n^k = \sum_{i_1,i_2,\dots,i_k,i_{k+1}=i_1} a_{i_1i_2}a_{i_2i_3}\dots a_{i_ki_1}. \tag{35}$$

如果我们把指标集 $i_1,i_2,\dots,i_k,i_{k+1}$ 放到 n 个节点的完全图上去看，那么上面的求和实际上就是对这个完全图上长度为 k 的回路进行求和 a_{ij} 表示一条从 i 到 j 的边。考虑其的期望，因为 $\mathbb{E}(a_{ij})=0$ 所以起到非 0 贡献的项每条边至少会出现两次。

如果 $k=2m+1$，那么起贡献的项一共出现最多 m 个不同边，这时因为这个图是连通图，它最多会有 $m+1$ 个顶点。我们给这 $m+1$ 个顶点进行标号有 $O(n^{m+1})$ 种选择，考虑归一化 $\frac{1}{n}\mathbb{E}(A_n^k)$ 后，我们知道其为

$$\frac{1}{n}\mathbb{E}(\frac{A_n}{\sqrt{n}})^{2m+1} = n^{-m-3/2} * O(n^{m+1}) = O(n^{-1/2}) \to 0. \tag{36}$$

我们对 $k=2m$ 的情况进行讨论，此时起主要贡献的项会出现 m 个不同边，以及 $m+1$ 个顶点，实际上就是一颗 $m+1$ 结点的树，并且带上一种走法使得每条边经过两次。考虑这样一个映射 $\phi:\{1,2,\dots,2m\}\to -1,1$ 的映射，如果第 i 步走过树上一条新的边 $\phi(i)=1$，否则吞掉过去走过的边，$\phi(i)=-1$.

令 $S_t=\sum_{i=1}\phi(i)$，那么 S_t 实际上是一个不穿过实轴的随机游走，可以证明这个随机游走和一颗带走法的树一一对应，而 S_t 的个数又是经典的 Catalan 数 $C_m=\frac{1}{m+1}\binom{2m}{m}$. 我们再把标号放回去，一共有 $n(n-1)\dots(n-m)\sim n^{m+1}$ 中标记方式，从而

$$\frac{1}{n}\mathbb{E}(\frac{A_n}{\sqrt{n}})^{2m} \sim n^{-m-1}C_m n^{m+1} \to C_m. \tag{37}$$

也就是说总会有

$$\frac{1}{n}\mathbb{E}(\frac{A_n}{\sqrt{n}})^{k} = \mu_n(x^k) \to \int_{-2}^{2} x^k \frac{1}{2\pi}\sqrt{4-x^2}. \tag{38}$$

也就是说 $\mathbb{E}\mu_n(x^k)\to\mu_{\text{semicircle}}(x^k)$，进一步如果证明这些测度都在一个紧集 $[-a,a]$ 中，由 weierstrass 逼近定理知道对于任意连续函数 f 会有 $\mathbb{E}\mu_n(f)\to\mu_{\text{semicircle}}(f)$。通过考虑方差我们可以将 $\mathbb{E}$ 去掉，而 $\mu_n(f)\to\mu_{\text{semicircle}}(f)$ 正是一列测度弱收敛的定义（ref to 泛函分析）。 □

proof of 3. 他实际上也是考虑图上回路的个数，不过此时是二部图上的闭合回路个数，我们只验证 $m=2$ 的情形，一般情况可以参考 2022spring 概率论进阶的第一次讲义。 □