arXiv:2305.05646v1 [math.PR] 9 May 2023

通过自由反卷积估计大协方差矩阵：计算与统计学方面

REDA CHHAIBI、FABRICE GAMBOA、SLIM KAMMOUN 和 MAURICIO VELASCO

摘要。大协方差矩阵的估计存在高维偏差。借助自由概率论这一工具，可以将偏差校正问题重新表述为自由反卷积。

本文旨在从计算和统计学角度解决这一问题。我们的方法基于复分析方法，用于求逆 S-变换。具体而言，我们需要从理论上理解多值 S 变换所处的黎曼曲面，并设计高效的计算方案。

目录

1. 引言 21.1. 借助 FPT 的重新表述 31.2. 文献综述 41.3. 论文结构 52. 主要结果 52.1. 统计学方面 52.2. 数值实验 82.3. 进一步评论 103. 斯蒂尔切斯变换的计算方面 103.1. 经验测度的斯蒂尔切斯变换计算 123.2. 谱密度估计 144. 克拉美-罗下界 144.1. 局部渐近正态性定理的证明 4.1 15附录 A. 黎曼曲面与马尔可夫-克莱因概论 18附录 B. 测度与矩 20 B.1. 正交多项式视角 20B.2. 三项递推 21B.3. 如何根据雅可比系数恢复测度？ 22参考文献 22

日期：2023年5月10日。2010 数学主题分类。主要分类 60F99；次要分类 60 G60、81P15。关键词和短语。自由性、自由卷积、大型随机矩阵、信号加噪声、变形矩阵模型。

1. 引言

大样本协方差矩阵估计： 一篇出色地解释该问题复杂性的论文是 [EK08]。回顾一下，给定一个独立同分布样本 X_1，....., X_n，其中的随机向量取自 $\mathbb{R}^d$，那么均值和协方差的经典估计量分别为：

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \in \mathbb{R}^d$$

$$\hat{\Sigma} = \frac{1}{n-d}\sum_{i=1}^{n} (X_i - \hat{\mu})(X_i - \hat{\mu})^* \in M_d(\mathbb{R}) \ .$$

当 d 固定且 $n \to \infty$ 趋于无穷时，这类估计量的性质完全良好。然而，如果 d 和 n 的规模相当，情况就会发生显著变化。事实上，现在假设 $d = \lfloor cn \rfloor$ 随 $n \to \infty$ 趋于无穷。为方便起见，我们假设 X_i 的各分量是独立同分布的约化且中心化的分量，因此 $\mu = \hat{}\mu = 0$，协方差估计量可以写为：

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^* = \frac{1}{n}\mathbb{X}\mathbb{X}^* \ ,$$

其中，$\mathbb{X} = (X_1, \ldots, X_n) \in M_{d,\ n}(\mathbb{R})$ 是以 X_i 为列构成的矩阵。

直观地看，人们可能会认为经验谱分布：

$$\mu_n(dx) := \frac{1}{d}\sum_{\lambda \in \mathrm{Spec}(\frac{1}{n}\mathbb{X}\mathbb{X}^*)} \delta_\lambda(dx)$$

收敛到 $\delta_1(dx)$——因为 $\frac{1}{n}\mathbb{X}\mathbb{X}^*$ 应当如固定 d 的情形那样收敛到单位矩阵。然而，实际情况是：

定理 1.1 （马尔琴科-帕斯图尔）。为简单起见，假设 $c < 1$。几乎必然地，当 $n \to \infty$ 时，我们有概率测度的弱收敛：

$$\lim_n \mu_n = MP_c$$

其中 $l = (1 - \sqrt{c})^2$, $r = (1 + \sqrt{c})^2$ 并且

$$MP_c(dx) := \mathbb{1}_{\{x \in [l,r]\}} \frac{\sqrt{(x-l)(r-x)}}{2\pi x} dx$$

是马尔琴科-帕斯图尔分布。

马尔琴科-帕斯图尔分布是一个以 1 为中心的宽分布，显然会在协方差矩阵（谱）的估计中引入高维偏差。

假设： 我们假设

$$\mathbb{X} = \Sigma^{\frac{1}{2}}\mathbb{Y} \ , \tag{1.1}$$

其中，$\Sigma \in M_d(\mathbb{R})$ 是真实协方差矩阵，而 $\mathbb{Y} \in M_{d,\ n}(\mathbb{R})$ 的系数独立同分布。在随机矩阵理论中，以下是一种自然的设定，尽管这一条件可以放宽。

假设 1.2. 我们假设谱测度收敛

$$\nu_d := \frac{1}{d} \sum_{\lambda \in \mathrm{Spec}(\Sigma)} \delta_\lambda$$

到概率测度 $\nu = \lim_{d\to\infty} \nu_d$。

这一假设可视为一种稀疏性假设，使我们能够在维度增长时估计目标对象。事实上，它基本上意味着，如果对真实的（通常未知的）协方差矩阵Σ 进行主成分分析并观察其特征模态，就会呈现出一种渐近形态。我们要探讨的主要问题是

问题 1.3。如何构造一个协方差矩阵 Σ，使得对于所有光滑的 f

$$\frac{1}{d} \operatorname{Tr} f\left(\widehat{\Sigma}\right) = \nu_d(f)$$

估计值为 $\nu(f)$，并在 $\frac{1}{n}$ 中波动。

速度 $\frac{1}{n}$ 是随机矩阵理论背景下线性统计量的自然收敛速度。

1.1. **借助 FPT 进行重述**。 自由概率论（FPT）源于 Voiculescu 的开创性研究，并已发展成为研究大型随机矩阵宏观行为的恰当框架。它提供了一个确定性模型，借助该模型，可以计算满足式（1.1）中乘法关系的大型矩阵的谱。

我们回顾各种变换的定义，这些变换可对测度进行重新编码 [Voi87]。给定一个测度 $\mu \in \mathcal{M}_1(\mathbb{R})$，我们将斯蒂尔切斯变换定义为

$$\forall z \in \mathbb{C}\backslash\mathbb{R},\ G_\mu(z) := \int_{\mathbb{R}} \frac{\mu(dt)}{z-t}\ .$$

一个众所周知的事实是，知道 μ 与知道 G_μ 是等价的。另一种重新编码方式是矩生成函数：

$$M_\mu(z) := zG_\mu(z) - 1 = \sum_{n=1}^{\infty} \frac{m_n(\mu)}{z^n}\ ,$$

若 $m_1(\mu) \neq 0$，则该函数在无穷远邻域内可逆。因此，函数逆 $M_\mu^{\langle -1\rangle}$在零点邻域内定义良好且为全纯函数。于是，Voiculescu S-变换为：

$$S_\mu(m) := \frac{1+m}{mM_\mu^{\langle -1\rangle}(m)}\ .$$

由于全纯延拓， μ、 G_μ、 M_μ 和 S_μ 都包含相同的信息，只是形式不同。因此，我们宽泛地将它们称为重新编码。

这使得我们能够简洁地重新表述 $\mathbb{X}$ 的观测谱与未知测度 ν 之间的渐近关系。

定理 1.4 (协方差矩阵估计的 FPT 结果总结). 在我们的假设下，$\mu_{\frac{1}{n}\mathbb{X}\mathbb{X}^*}$ 弱收敛于测度 μ，该测度是 ν 和 MP_c 的测度的确定性函数，而它们分别是 Σ 和 $\mathbb{Y}\mathbb{Y}^*$ 的极限谱测度。该测度称为乘法自由卷积：

$$\mu = \nu \boxtimes MP_c \ .$$

此外，上述关系可以改写为两个等价的方程。

- *(S-变换)* S 变换满足：

$$S_\mu(m) = S_\nu(m) S_{MP_c}(m) \ ,$$

对所有位于0某个邻域内的 m，均成立。

- *(Marchenko-Pastur 方程)*

$$-\frac{1}{G_\mu(z)} = z - c \int \frac{\lambda \nu(d\lambda)}{1 + \lambda G_\mu(z)} \ .$$

证明提示。收敛到乘法自由卷积是 Voiculescu 定理 [Voi87] 以及 Σ 和$\mathbb{Y}$ 渐近自由这一事实的一个特例。并且，$\mathbb{Y}\mathbb{Y}^*$ 的谱收敛到由经典定理给出的 MP_c测度 1.1。

自由卷积的两种刻画方式之间的等价性是经典结果。使用 S-变换是 Voiculescu 最初提出的观点。Marchenko-Pastur 方程可见于 [EK08, 定理 1]，它是更一般的从属现象的一个实例 [BB07]。 □

为了强调我们关注的是自由反卷积，我们引入一个顾名思义的符号，其隐式定义如下：

$$\mu_3 = \mu_1 \boxtimes \mu_2 \iff \mu_1 = \mu_3 \boxslash \mu_2 \ .$$

基本上，对于输入测度 μ 和 MP_c，目标是计算输出测度$\nu = \mu \boxslash MP_c$。借用信号处理领域的术语，这就是对测度 μ 进行 Marchenko-Pastur 分布 MP_c的自由反卷积。

更准确地说，这必须以经验方式完成。事实上，我们只能观测到 $\frac{1}{n}\mathbb{X}\mathbb{X}^*$ 的经验谱测度 μ_n。因此，我们旨在构造一个估计量 $\widehat{\nu}_n$，使得：

$$S_{\widehat{\nu}_n} \approx S_{\widehat{\mu}_n} / S_{MP_c} \ .$$

尽管此时 $\approx$ 的含义仍不明确，但我们可以这样理解：符号 $\approx$ 表示零点邻域内全纯函数的接近程度。反过来，这反映了底层测度的自由累积量之间的接近程度，因此也反映了测度上的通常弱拓扑。

1.2. **文献综述**。 探讨自由概率统计和计算方面的论文寥寥无几。有两篇值得关注的文献，本文需要将其与本研究进行比较。

关于这一主题的早期论文中，有[BGD08]。特别请注意第六节中关于使用基于 R变换的方法的评论：“遗憾的是，这种方法只在极少数情况下有用，因为这里所必需的运算（某些函数的求逆、解析函数的延拓）几乎总是无法在实践中实现。”

事实上，我们偏要反其道而行之，恰恰去做这件事。通过同伦方法和牛顿-拉夫森方案，可以非常有效地求解析函数的逆，前提是仔细控制覆盖和吸引域。

开创性论文 [EK08] 求解了一个凸优化问题，该问题源自定理 1.4中的 Marchenko-Pastur 方程。基本上，研究者要寻找使该方程误差最小的测度 ν 。显然，这里存在稳定性问题：方程的近似零点未必接近该方程的零点。

最近，在一系列论文 [ATV20、Tar20], 中，作者首先使用不动点算法计算所关注测度的斯蒂尔切斯变换。这是从属方法的一个副产品——参见 [BB07、BMS17] 及其中的参考文献。无论如何，尽管从属方法是一种出色且灵活的工具，但仍有两个主要难点：

- 从属方法是一种不动点方法。而在低维情形下，不动点方法总是不如著名的牛顿-拉夫森方案。
- 斯蒂尔切斯变换是在实轴之外计算的，因此自由反卷积的结果 $\mu \boxslash \nu$ 只能确定到与某个（可能方差很大的）柯西随机变量进行经典卷积的程度。这样一来，问题就转化为经典去卷积问题。需要注意的是，尽管这是一个经典的逆问题，但通常求解起来不稳定且困难。正如我们将看到的，若能在更接近实轴的位置计算斯蒂尔切斯变换，将大有裨益，而且这实际上是可行的。

1.3. **论文结构**。在第 2节中，我们从三个不同方向陈述论文的主要结果。

- 计算方面：我们介绍所用的方法，并陈述其依据的定理。特别是，需要构造适当的围道，并根据含噪矩重建测度。
- 统计方面：克拉美-罗下界表明，在一个有实际意义的模型中，收敛速度不可能超过 $\frac{1}{n}$ 。

论文其余部分将用于证明。

在第 3节中，我们讨论论文中与优化相关的方面。

在第 4节中，我们着手证明克拉美-罗下界。事实上，我们先讨论局部渐近正态性，再给出一般风险函数的界，从而证明一个更一般的结论。

2. 主要结果

2.1. 统计学方面。 本节旨在推导总体测度估计的克拉美-罗下界。由于克拉美-罗下界要求模型正则，因此我们将采用相当严格的设定——但仅限于本部分。为此，固定原子的数量 q ，并将总体测度写作：

$$\mu_\theta = \sum_{k=1}^{q} w_k \delta_{x_k},$$

其中 $\theta = (x, w) \in C_q \times \Delta_q^{(\mathbb{R})}$且

$$C_q := \{x \in \mathbb{R}_+^q \mid 0 < x_1 < x_2 < \cdots < x_q\} \ ,$$

$$\Delta_q^{(\mathbb{R})} := \left\{ w \in (\mathbb{R}_+^*)^q \mid \sum_{k=1}^q w_k = 1 \right\} \ ,$$

分别是 Weyl 室和概率测度单纯形。因此，编码测度 μ_θ 的参数空间是如下有限维空间：

$$\Theta := C_q \times \Delta_q^{(\mathbb{R})} \ .$$

注意， Θ 是一个与 $\mathbb{R}^{2q-1}$微分同胚的空间中的开集。

2.1.1. 模型。 对于 $\theta \in \Theta$，我们来描述实验的生成律 $\mathbb{P}_{n,\,\theta}$。给定 n 和 $p = p_n$，其中

$$p_n/n \to c > 0 \ .$$

观测到一个矩阵 $\mathbb{X} \in M_{p,\,n}(\mathbb{R})$，其生成方式如下：

$$\mathbb{X} = V^{\frac{1}{2}} \mathbb{Y}$$

其中 $\mathbb{Y}$ 矩阵的元素是独立同分布高斯元素。也就是说：

$$\mathbb{X}\mathbb{X}^* = V^{\frac{1}{2}} \mathbb{W} V^{\frac{1}{2}} \ ,$$

其中 $\mathbb{W}$ 是白色 Wishart 矩阵，而 $V = V_{n,\,\theta}$ 是总体矩阵。

此外，我们引入以下支集位于离散单纯形上的吉布斯测度

$$\Delta_{q,n}^{(\mathbb{N})} := \left\{ k \in \mathbb{N}^q \mid \sum_{i=1}^q k_i = n \right\} \ ,$$

其定义如下：

$$\begin{aligned} \mathbb{P}\left(N_k^{(n)} = k_i \ , \ 1 \le i \le q \right) =& \frac{1}{Z_{n,w}} \prod_{i=1}^q \left(\frac{k_i}{n} \right)^{\beta n^2 w_i} \\ =& \frac{1}{Z_{n,w}} \exp\left(\beta n^2 \sum_{i=1}^q w_i \log \frac{k_i}{n} \right) \ . \end{aligned}$$

注意，在将熵与

$$\beta n^2 \sum_{i=1}^q w_i \log \frac{k_i}{n}$$

多项分布的熵进行比较时：

$$n \sum_{i=1}^q \frac{k_i}{n} \log w_i \ ,$$

我们注意到，速度从 n 提升至 n^2。此外，参数 w 与估计量 k/n 之间也发生了反转。

假设 2.1（仅适用于克拉美-罗下界）。 总体矩阵具有均匀特征向量和独立谱，因此：

$$\mu_V = \sum_{i=1}^{q} \frac{N_k^{(n)}}{n} \delta_{x_k}$$

也就是说

$$V_{n,\theta} = U^* D_{n,\theta}^V U,$$

其中 U *Haar* 在正交群上服从 *Haar* 分布，$D_{n,\theta}^V$ 是对角矩阵，其相关系数按照欠离散测度 N 抽样得到。

注 2.2 (不恰当的选择**)**。 以下是两种自然但不恰当的选择。引入该测度是为了修正它们的缺陷。

随机矩阵理论缩放： 将元素 $D_{n,\theta}^V$ 选为按照测度 μ_θ 独立同分布抽样，可得到如下谱测度：

$$\mu_V = \sum_{k=1}^{q} \frac{N_k}{n} \delta_{x_k} \ ,$$

其中 $N^{(n)} = (N_k \ ; \quad 1 \le k \le q)$ 是一个多项分布。由于抽样相互独立，N_k 将在尺度 $n^{\frac{1}{2}}$ 上呈现高斯涨落——这对于随机矩阵理论区域来说大得多。

模型的正则性： 另一种常见的选择是令 V 具有谱测度

$$\mu_V = \sum_{k=1}^{q} \frac{\lfloor w_k p \rfloor + \mathcal{O}(1)}{p} \delta_{x_k}$$

使得 $V_{n,\theta}$ 的谱测度收敛到 μ_P。但这种选择行不通。取整函数给映射 $\theta \mapsto \mathbb{P}_{n,\theta}$带来了过多的不规则性。这种模型不是正则的，因此不适合使用似然进行分析。

众所周知，[And03，§7。2]的定律 $\mathbb{X}$ 具有如下概率密度：

$$\frac{d\mathbb{P}_{n,\theta}}{d\lambda_{M_p(\mathbb{R})}}(\mathbb{X}) = \frac{|\mathbb{X}|^{\frac{n-p-1}{2}} \exp\left(-\frac{1}{2} \operatorname{tr} V_\theta^{-1} \mathbb{X}\right)}{2^{\frac{np}{2}} |V|^{\frac{n}{2}} \Gamma_p(\frac{n}{2})} \ ,$$

其中 Γ_p 是多元伽马函数，$|\cdot|$ 是行列式，而$\lambda_{M_p(\mathbb{R})}$ 是对称矩阵上的自然勒贝格测度。

2.1.2. 陈述。 极限实验的设定要求计算对数似然的非平凡极限：

$$\log \frac{d\mathbb{P}_{n,\theta+\varepsilon_n v}}{d\mathbb{P}_{n,\theta}}(X)$$

其中 ε_n 是适当的重新缩放速度，v 是空间 Θ中的一种形变。

现在我们可以陈述本文的主要统计定理。

定理 2.3. 设 θ_n 为一个估计量，满足：

- 在 [VdV00 §8 的意义下正则。5]
- 渐近无偏。

那么，存在常数 $C > 0$ ，使我们得到克拉美-罗界：

$$\liminf_{n\to\infty} n \operatorname{Var} \widehat{\theta}_n \geq C \ .$$

证明概要。 这里我们采用以下证明策略。

- 我们在定理中证明了局部渐近正态性（局部渐近正态性）结果4.1。
- 我们援引局部渐近极小极大定理[VdV00，定理 8。11]——另见[IHM13]。

□

2.2. **数值实验**。 在本小节中，我们报告数值实验的结果。我们采用上述技术，并在 GitHub代码库中提供了实现。

https://github.com/mauricio-velasco/free_deconvolution

2.2.1. 其他技术的描述：我们将在此介绍上述提到的方法。回顾一下，我们观测到一个矩阵 $\mathbb{X}$ ，它满足$\mathbb{X} = V^{\frac{1}{2}}\mathbb{Y}$ ，其中 $\mathbb{Y}$ 矩阵的元素为独立同分布高斯元素。

El Karoui 的凸优化技术 [EK08]**：** 定义

$$\nu_n(z) = \frac{n-p}{nz} + \frac{1}{n}\operatorname{trace}((\mathbb{X}\mathbb{X}^* - z\,\mathrm{id}_p)^{-1}). \tag{2.1}$$

假设 V 的谱测度可以近似为

$$\sum_{k=1}^{K} w_k \delta_{x_k}$$

那么，当 n趋于无穷 ν_n 且满足合理条件时 [EK08、定理1和第3.2小节1]应满足以下近似关系

$$0 \simeq \frac{1}{\nu_n(z)} + z - \frac{p}{n}\sum_{k=1}^{K} \frac{w_k x_k}{1 + x_k \nu_n(z)}.$$

El Karoui 的思路是选择一些取值网格 $(z_j)_{1\leq j\leq J}$ 和 $(x_j)_{1\leq j\leq K}$，以最小化误差向量 $e = (e_j)_{1\leq j\leq J}$ 的范数，其中

$$e_j = \frac{1}{\nu_n(z_j)} + z_j - \frac{p}{n}\sum_{k=1}^{K} \frac{w_k x_k}{1 + x_k \nu_n(z_j)}$$

在满足约束条件 $\sum_k w_k = 1$ 的情况下，对于所有 k、$w_k \geq 0$ 均成立。El Karoui 考虑了 ℓ^p（$\mathbb{R}^J$）范数，其中 $p = 1$、2 和 ∞。在固定 x_j、将未知的 w_k 作为变量的情况下，我们得到一个带有约束条件的凸优化问题，这在理论上具有吸引力。

在实践中， z_j的选择至关重要。El Karoui 建议先选择$\nu_n(z_j)$，再对（2.1）求逆以求得 z_j。尽管 El Karoui 只详细讨论了测度为原子的情形，但他建议对具有分段仿射密度的测度也进行类似的研究。在实践中，改变测度字典似乎对结果影响不大。

该方法的缺点：

- 不存在规范的选择，并且将 z_j 确定为$\nu_n(z_j)$的函数时存在不稳定性。

- 稳定性问题： e_j中的微小误差并不能保证测度之间的接近性。

Arizmendi 等人的从属方法。 [ATV20]**：** 他们对加性去卷积和乘性去卷积都采用了从属方法。在估计协方差矩阵的背景下，我们只关注乘性情形。假设 $\mu_1 \boxtimes \mu_2 = \mu_3$ 并定义某个测度的 F–变换为 $F_\mu(z) = \frac{1}{G_\mu(z)}$。在合理条件下， [ATV20，定理 14] 证明了某个 a 的存在性，该 a σ 可以显式计算，并且存在一个函数 ω_3 定义在 $\mathbb{C}_\sigma$ 上：$= \{z \in \mathbb{C} : \Im(z) \geq \sigma\}$ 使得

$$F_{\mu_2}(z) = F_{\mu_3}[\omega_3(z)]z\omega_3(z)^{-1}.$$

此外，当 w 处于 z的某个邻域内时， T_z的迭代 $T_z^{\circ n}(w)$ 收敛于 $\omega_3(z)$。这里 $T_z(w)$：$= zh_1(h_3(w)^{-1}z^{-1})$ $h_1(w) = w - F_{\mu_1}(w)$，以及 $h_3(w) = w^{-2}(w - F_{\mu_3}(w))$。

既然 F_{μ_2} 可以在 $\mathbb{C}_\sigma$上计算，我们就从 F_{μ_2} 转到 μ_2。因此，需要通过柯西测度进行经典加性去卷积，因为

$$F_{\mu_2}(x + i\sigma) = 1/G_{\mu_2}(x + i\sigma) \ ,$$

and

$$-\Im G_{\mu_2}(x + i\sigma) = \int_{\mathbb{R}} \frac{\sigma}{\pi((x-y)^2 + \sigma^2)}\mu_2(dy) = \mu_2 * \mathcal{L}(\sigma\mathcal{C})(x) \ ,$$

其中 $\mathcal{C}$ 是标准柯西随机变量。

由此我们来到了[ATV20]方法的关键所在：非线性自由去卷积问题现在已转化为经典去卷积问题。当 μ 为原子测度，并进一步进行离散化（对于x的取值）， μ 可以通过求解如下形式的线性逆问题来近似 $KU = V$，其中 K 是经典卷积算子。在[ATV20], 中，他们建议采用 Tikhonov 正则化，即最小化 $\|KU - V\|_2 + \alpha^2\|U\|_2$，其中参数为 α。熟悉统计学的读者可以看出，Tikhonov 正则化不过就是常见的岭回归。

该方法的缺点：

- 当 σ取值相对较高时，次级化方法才有效。理论下界很不理想，需要手动调参。
- 经典反卷积是一个逆问题。尽管它是信号处理领域一个极为经典的问题，但最好避免采用这种方法。

2.2.2. 情景描述： 我们将在 5 种不同的情景下测试该算法。我们将参考 [EK08]中给出的三种情景。事实上，情景 1、2.1 和 3 与 [EK08] 中的相同，但我们对这些情景做了一些修改，以比较不同的方法。

- 情景 1：当 $\mathbb{X}\mathbb{X}^*$ 是 Wishart矩阵（V =id）时。
- 情景 2.1： Σ 的特征值一半为 1，另一半为 2，且（$p/n = 0.2$）
- 情景 2.2： Σ的特征值一半为 1，另一半为 1.2（$p/n = 1$）
- 情景 2.3： Σ 的谱测度为 $\frac{\delta_1+\delta_2+\delta_5+\delta_6+\delta_8}{5}$
- 情景 3： Σ 是一个 Toeplitz 矩阵，其元素为 $\Sigma_{i,\ j} = 0.3^{|i-j|}$。

2.2.3. 给出图表。

- 反卷积结果 我们绘制了不同 n取值下真实谱测度与估计谱测度之间的 Wasserstein 1距离。
- 性能速度

2.3. **进一步评论**。

更多应用： 我们从估计协方差矩阵这一问题入手，它是统计学领域的核心问题。事实上，协方差矩阵是许多方法的首要输入。因此，自由反卷积的相关性还不止于此，还可用于：

- 异常值检测、主成分分析。
- 另一个应用领域是无线电信号（多输入多输出，MIMO），如今的网络由大量异构天线和接收器构成。

与经典反卷积的比较： 经典反卷积是信号处理中的标准课题。在这种设定下，我们处理的是一个线性逆问题，由于高频成分而存在许多固有的不稳定性。此外，噪声水平通常未知。对于自由反卷积，已知测度与未知测度之间的依赖关系是非线性的，而且噪声水平实际上是已知的。因此，我们认为，经典反卷积与自由反卷积之间的比较，只有在类比层面上才有意义。

REE 类： 上述结果聚焦于谱，完全忽略了特征向量。然而，如果希望构造$\widehat{\ }$一个完整的矩阵 Σ，只需将范围限定在旋转等变估计量（REE）这一类中，便可轻松实现；在这一类估计量中，特征向量可直接根据观测结果确定。

$\widehat{\ }$ $\widehat{\ }$**定义 2.4** (REE)。 若估计量 $\Sigma = \Sigma(\mathbb{X})$ 具有以下性质：对于所有正交矩阵 $O \in O_n(\mathbb{R})$，则其属于 *REE* 类：

$$\widehat{\Sigma}(O\mathbb{X}) = O\widehat{\Sigma}(\mathbb{X})O^* . \tag{2.2}$$

特别地，将 $\mathbb{X}\mathbb{X}^* = ODO^*$ 对角化并考虑估计量是自然的

$$\widehat{\Sigma}(\mathbb{X}) = O\Lambda O^*$$

其中，Λ 是待确定的谱，它是 $\mathbb{X}$的谱的函数。

文献中的其他论文将这一性质称为旋转不变估计量 [BGBP19、 BG 22]。我们认为这一称谓并不准确，因为式（2.2）表达的并非不变性，而是等变性或协方差性质。此外，任何超出 REE 类的改进都需要给出动机说明，这不属于本文的讨论范围。

3. 关于斯蒂尔切斯变换的计算方面

测度 μ 的斯蒂尔切斯变换 $G(z)$ 的有用之处在于，它能够方便地编码其矩。如下面的引理所示，这些信息包含在 $G(z)$ 沿任意一条.....的取值中。

测度的支集 K 周围的围道。在本节中，我们将展示这种表示在计算上非常有用。

引理 3.1. 设 μ 是一个支集包含于紧集$K \subseteq \mathbb{R}$ 中的概率测度，并设 $\sigma(t)$ 是 $\mathbb{C} \setminus K$ 中与环绕 K的椭圆同伦等价的围道。若 $T(z)$ 是任意多项式，则有以下等式

$$\frac{1}{2\pi i} \oint_{\sigma} T(z)G(z)dz = \int_{K} T(t)d\mu(t).$$

证明. 由于斯蒂尔切斯变换在测度支集之外全纯， K 的紧性使其在无穷远处全纯，并且在那里具有以下幂级数展开

$$G(1/w) := \int_{\mathbb{R}} \frac{1}{1/w - t} d\mu(t) = \sum_{k=0}^{\infty} w^{k+1} \int_{\mathbb{R}} t^k d\mu(t)$$

若对于某个整数 j 有 $T(z) = z^j$，且 σ 是给定的围道，则有

$$\frac{1}{2\pi i} \oint_{\sigma} T(z)G(z)dz = \frac{1}{2\pi i} \oint_{\sigma} T(1/w)G(1/w)(-1/w^2)dw = \frac{1}{2\pi i} \oint_{-\sigma} \left(\sum_{k=0}^{\infty} w^{k-j-1} \int_{\mathbb{R}} t^k d\mu(t) \right) dw.$$

由于该函数在 $w = 0$ 处全纯，且 $-\sigma$ 是与围绕 $w = 0$ 的圆同伦的正向围道，根据柯西留数定理，该积分等于其留数 $\int_{\mathbb{R}} t^j d\mu(t)$，从而证明了结论。□

定义 3.2. </style}测度 的斯蒂尔切斯变换 (z)的围道表示是一个二元组 $(\sigma(t)\ v(t))$，其中 $0 \le t \le 2\pi$， $\sigma(t)$ 是对一条曲线的参数化，该曲线绕 K 一周，而 $v(t) := G(\sigma(t))$ 记录 G 在 σ各点处的取值。

注释 3.3. 在实现中，我们将用足够多组复数对来表示 $(\sigma(t)\ G(\sigma(t))$，即 $(\sigma(t_j)\ G(\sigma(t_j))$、$j = 1$，....., C，使得近似

$$\int_{K} T(t)d\mu(t) \sim \frac{1}{2\pi i} \sum_{j=1}^{C} T(\sigma(t_j))G(\sigma(t_j)) \left(\sigma(t_j) - \sigma(t_{j-1})\right)$$

在所需次数的多项式上足够准确。我们计算围道表示的方法使我们能够根据需要增加点的数量 C 以及数值 $G(\sigma(t_j))$ 的精度。

下面的引理给出了 $G(z)$ 的围道表示，其中已知函数 $S(m)$ 在原点的某个单连通邻域 U 上，从而得到一种实用的反演过程，

引理 3.4. 假设在原点的某个单连通开邻域 U 中已知 $S(m)$ 。若 $m(t)$ 是任意一条与绕原点圆周同伦且包含于 U 中的围道，则 $(z(t)\ v(t))$ 是 $G(z)$ 的围道表示，其中

$$z(t) := \frac{1 + m(t)}{m(t)S(m(t))} \ and\ v(t) := z(t)m(t) - 1.$$

证明。 根据 $S(m)$的定义，我们知道

$$z(t) = \frac{1 + m(t)}{m(t)S(m(t))} = \hat{M}^{-1}(m(t))$$

因此， $M(z(t)) = m(t)$。由此可知， G 沿轮廓 $z(t)$上各点的取值为 $G(z(t)) = z(t)M(z(t)) - 1 = z(t)m(t) - 1$，如所述。□

3.1. 计算经验测度的斯蒂尔切斯变换。本节讨论如何计算离散测度 μ 的逆变换 $S(m)$，其中其支集包含在紧集 $K \subseteq \mathbb{R}_{>0}$ 中。更具体地，对于 $j = 1$、.....、L，取正实数 λ_j 和正实权重 w_j，满足 $1 = \sum_{j=1}^{M} w_j$，并令 $\mu := \sum_{j=1}^{M} w_j \delta_{\lambda_j}$。在这种情况下，斯蒂尔切斯变换是关于 z 的有理函数，即

$$G(z) := \sum_{j=1}^{L} \frac{w_j}{z - \lambda_j} \text{ and } M(z) := -1 + \sum_{j=1}^{L} \frac{w_j z}{z - \lambda_j}.$$

我们的第一个引理给出了一个有效的判据，用于验证局部反函数 $\hat{}M^{-1}(m(t))$ 的存在性。

引理 3.5。设 U 是一个开连通集，其边界是一条简单闭曲线 $\sigma(t)$，其中 $0 \leq t \leq 2\pi$。设 $f : \overline{U} \to \mathbb{C}$ 是一个全纯函数，使得对于每个 $z \in U$，都有 $f'(z) \neq 0$。以下陈述等价：

(1) 函数 f 在 $\overline{U}$ 上是单射。*(2)* 函数 f 在边界曲线 $\sigma(t)$ 上是单射（等价地，曲线 $\tau(t) := f(\sigma(t))$ 对于 $0 \leq t \leq 2\pi$ 是一条简单闭曲线）。

证明。蕴含关系 (1) $\to$ (2) 是显然的，因为 $\mathrm{im}(\sigma) \subseteq \overline{U}$。(2) $\Longrightarrow$(1) 我们首先断言，不存在 $z_0 \in U$ 和 $z_1 \in \partial U$ 使得 $f(z_0) = f(z_1)$。否则，由于 $f'(z_0) \neq 0$，函数 f 在 z_0 附近是开映射，因此存在一点 $z^* \in U$，使得 $f(z^*)$ 严格位于若尔当曲线 $\tau(t)$ 的外部。由于 $f(z) - f(z^*)$ 在 z 中是亚纯函数，且在 σ 上没有零点或极点，因此辐角原理意味着，

$$1 \leq \frac{1}{2\pi i} \oint_{\sigma} \frac{[f(z) - f(z^*)]'}{f(z) - f(z^*)} dz = \frac{1}{2\pi i} \int_0^{2\pi} \frac{f'(\sigma(t))\sigma'(t)}{f(\sigma(t)) - f(z^*)} dt$$

作代换 $\tau(t) = f(\sigma(t))$ 后，最后一个积分等于

$$\frac{1}{2\pi i} \int_0^{2\pi} \frac{f'(\sigma(t))\sigma'(t)}{f(\sigma(t)) - f(z^*)} dt = \frac{1}{2\pi i} \int_0^{2\pi} \frac{\tau'(t)}{\tau(t) - f(z^*)} = \frac{1}{2\pi i} \oint_{\tau} \frac{dw}{w - f(z^*)} = 0$$

由于 $f(z^*)$ 严格位于 τ 的外部，该积分等于零。这一矛盾表明，不存在满足 $f(z_0) = f(z_1)$ 的 $z_0 \in U$ 和 $z_1 \in \partial U$，同样的论证也表明，不存在使 $f(z_0)$ 严格位于若尔当曲线 τ 外部的 $z_0 \in U$。因此，对于每个 $z_0 \in U$ $f(z_0)$，它都位于由 τ 围成的区域内部； f 唯一可能不具有单射性的点是 $z_0 \in U$。为证明单射性，我们将对每个 $z_0 \in U$ 计算函数 $f(z) - f(z_0)$ 的零点个数。根据辐角原理，零点个数为

$$\frac{1}{2\pi i} \oint_{\sigma} \frac{(f(z) - f(z_0))'}{f(z) - f(z_0)} dz = \frac{1}{2\pi i} \oint_{\tau} \frac{dw}{w - f(z_0)} = 1$$

其中，最后两个等式由变量代换 $\tau(t) := f(\sigma(t))$ 得出；此外，我们已证明 $f(z_0)$ 位于

由 τ 界定的区域内部。该等式证明了 f 在 $\overline{U}$ 上的单射性，正如所述。□

下面的引理总结了这类测度的 $M(z)$ 的基本性质。我们用字母 $\mathcal{S}$ 表示黎曼球面：它是在拓扑球面 $\mathbb{C}\cup\{\infty\}$ 上赋予唯一复结构后得到的，该复结构将 $\mathbb{C}$ 上通常的复结构延拓到整个球面。

引理 3.6. 除去一个勒贝格测度为零的集合，以下结论对所有 $(\lambda_1,\ldots,\lambda_L,w_1,\ldots,w_L)\in K^L\times\Delta_L$ 均成立。

(1) 映射 $M:\ \mathbb{C}\setminus\bigcup_{j=1}^{L}\{\lambda_j\}\to\mathbb{C}$ 可延拓为唯一的全纯映射 $M:\ \mathcal{S}\to\mathcal{S}$。*(2)* 映射 M 的次数为 L，并有 $2(L-1)$ 个分歧点（即满足 $M'(z)=0$ 的点 z），且这些分歧点的像互不相同。这些点在 M 下的像称为 M 的分支点，它们由 $\mathbb{C}$ 中 $L-1$ 对互为共轭且各不相同的点组成。*(3)* 记 $p_1,\ldots,p_{L-1}$ 为每对共轭分支点中虚部为正的点，并令 U 为去掉连接每个 p_j 与无穷远点的竖直直线及其共轭直线后所得的集合，则 U 是单连通的，且存在唯一的全纯函数 $\hat{}M^{-1}:\ U\to\mathcal{S}$，使得 $\hat{}M^{-1}(0)=\infty$，并且对每个 $u\in U$，均有 $M(\hat{}M^{-1}(u))=u$。

证明。 □

前述引理给出了在给定复数 m 处构造 $S(m)$ 的算法，可分为以下两个步骤：

(1) 构造所有分支点 $p_1,\overline{p_1},\ldots,p_{(L-1)},\overline{p_{(L-1)}}$，它们属于 M，并定义单连通域 U。首先通过求解方程 $M'(z)=0$，找出所有分歧点 $q_1,\ldots,q_{2(L-1)}$；再令 p_i 为它们在 M 下的像，并将其分成共轭对。按照上文的方式定义单连通区域 U：令 U 为去掉连接每个 p_j 与无穷远处的竖直直线及其共轭线后所得的补集。

(2) 通过路径提升 计算 $\hat{}M^{-1}(m)$。任取路径 $m(t)$，使得 $m(0)=0$，且 $m(1)=m$ 完全包含于 U。沿着满足方程 $M(w(t))=m(t)$ 的解提升路径 $m(t)$，并施加初始条件 $w(0)=\infty$，即可唯一确定 $\hat{}M^{-1}(m)$ 的值。由于 M 是一个在分支点原像之外的覆盖映射，因此该提升的存在性和唯一性显而易见。

(3) 通过 $S(m)$ 计算 $S(m):=(1+m)/m\ \hat{M}^{-1}(m)$

执行上述前两个步骤需要专门的算法工具，值得更精确地描述如下：

(1) 构造所有分支点需要求出方程 $M'(z)=0$ 的所有解。难点在于确保我们找到了所有 解。(2) 要准确提升路径，需要结合两种方法，即同伦方法和牛顿迭代。这一思路源自数值代数几何领域的大量文献 [?]

并且能够极其准确地估算 S 的值。具体而言，给定路径 $m(t)$，提升路径$w(t)$ 的计算分为两个步骤：

(a) 初始近似。 给定步长 h ，我们希望计算一个初始猜测值 $w(t+h)$。为此，我们将方程$M(w(t)) = m(t)$展开至一阶并离散化，得到公式

$$\widetilde{w(t+h)} = w(t) + \frac{m(t+h) - m(t)}{M'(w(t))}$$

(b) 牛顿法细化。 我们对初始近似进行细化 $\widetilde{w(t+h)}$，即 $w(t+h)$，对方程 $M(w(t+h)) = m(t+h)$应用牛顿法。由此得到迭代格式：

$$\begin{cases} w_0 = \widetilde{w(t+h)} \\ w_{n+1} = w_n + a_n \text{ where } a_n := \frac{m(t+h) - M(w_n)}{M'(w_n)} \end{cases}$$

We define $w(t+h) := w_n$ for sufficiently large n.

注 3.7。 请回忆，若初值 w_0 位于真实解的吸引域内，则牛顿法具有二次收敛性；当 w_0 在第一步中计算得到且 h充分小时，这一条件成立。

3.2. **谱密度估计**。 结合前面的工作，我们现在可以构造谱密度的估计量。请回忆，我们只能观测到$\frac{1}{n}\mathbb{X}\mathbb{X}^*$ 的经验谱测度 μ_n ，并希望构造总体谱测度 ν的估计量 $\widehat{\nu}_n$ 。我们通过间接估计 S_ν来实现这一目标。本节的主要结果如下

定理 3.8。 如果 $T_n(m) := S_{\mu_n}(m)/S_{MP}(m)$ 那么以下陈述成立：

(1) $T_n(m)$ 在 $\mathbb{C}$中的原点的紧子集上一致收敛于 $S_\nu(m)$ ，且 *(2)* 收敛速率满足 $\|T_n - S_\nu\| \sim O(1/n)$

上述定理提出了一种估计谱测度 ν的方法。

(1) 计算 $T_n(m) := S_{\mu_n}(m)/S_{MP}(m)$ 其中，分子按照第 3.1 节中的方法进行近似。(2) 使用 $T_n(m)$ 获得 $G_\nu(m)$ 的近似的围道表示，如引理 3.4所示。(3) 通过从其围道表示中恢复测度，构造估计量测度 $\widehat{\nu}$ ，即利用通过引理 3.1从其矩的近似知识中恢复测度 ν 。

根据谱测度 ν 的性质，我们将采用不同的恢复机制，这些机制将在本节余下部分进行讨论。

4. Cramér-Rao lower bound

下面介绍一些记号。记

$$(x_1(\theta), \ldots, x_q(\theta))$$

$$(w_1(\theta), \ldots, w_q(\theta))$$

分别表示与参数 θ 相关的支集和权重。

此外，将形变 v as $v = (g, h) = v_g + v_h$ 拆分为沿支集的形变 $h \in \mathbb{R}^q$ 和沿权重的形变 $g \in \mathbb{R}^q$，我们有：

$$x_k(\theta + \varepsilon_n v) = x_k(\theta) + \varepsilon_n h_k \ ,$$
$$w_k(\theta + \varepsilon_n v) = w_k(\theta) + \varepsilon_n g_k \ .$$

自然，为了保持与单纯形相切，我们需要取 $\sum g_k = 0$，而 h 是自由的。

现在，我们可以陈述本节的主要结果。

定理 4.1 (局部渐近正态性（LAN）)。对于速度 $\varepsilon_n = \frac{1}{n}$，我们有以下依分布收敛：

$$\lim_{n\to\infty} \log \frac{d\mathbb{P}_{n,\theta+\varepsilon_n v}}{d\mathbb{P}_{n,\theta}}(\mathbb{X})$$
$$= -\frac{1}{4}c\sum_{k=1}^{q} w_k(\theta)\frac{h_k^2}{x_k(\theta)^2} + \frac{1}{2}c^{\frac{1}{2}}\sum_{j=1}^{q}\mathcal{N}_k w_k(\theta)\frac{h_k}{x_k(\theta)} \ ,$$

在参考分布 $d\mathbb{P}_{n,\ \theta}$ 下。

4.1. **局部渐近正态性定理的证明 4.1.** 先介绍一些记号。独立同分布随机变量 $m_i \in \{1, 2, \ldots, q\}$ 是以下 n种模态的结果：

$$[D^V]_{i,i} = x_{m_i} \ .$$

与之对应的 Gibbs 型测度为 $N = (N_1, \ldots, N_q)$，其定义如下：

$$N_k := \mathrm{Card}\,\{1 \le i \le n \mid m_i = k\} \ .$$

对数似然的（离散）部分为：

$$\beta n^2 \sum_{k=1}^{q} w_k \log \frac{N_k}{n} \ .$$

根据密度公式

$$\log \frac{d\mathbb{P}_{n,\theta+\varepsilon_n v}}{d\mathbb{P}_{n,\theta}}(\mathbb{X})$$
$$= \beta n^2 \sum_{k=1}^{q} \left(w_k(\theta + \varepsilon_n v) - w_k(\theta)\right) \log \frac{N_k}{n}$$
$$- \frac{n}{2}\log\frac{|V_{n,\theta+\varepsilon_n v}|}{|V_{n,\theta}|} - \frac{1}{2}\,\mathrm{tr}\left((V_{n,\theta+\varepsilon_n v}^{-1} - V_{n,\theta}^{-1})\mathbb{X}\right) \ .$$

现在我们来分析这三项。

步骤 1：参数化 Θ 以及前两项。

得益于定理陈述前引入的便捷记号，我们可以写成：

$$\log|V_{n,\theta}| = \log\left|D_{n,\theta}^V\right|$$
$$= \sum_{j=1}^{p} \log\left(D_{n,\theta}^V\right)_{j,j}$$

$$= \sum_{k=1}^{q} p_k \log x_k(\theta) \ .$$

因此，我们有：

$$\begin{aligned} & -\frac{n}{2} \log \frac{|V_{n,\theta+\varepsilon_n v}|}{|V_{n,\theta}|} \\ =& -\frac{n}{2} \left(\sum_{k=1}^{q} p_k \log x_k(\theta+\varepsilon_n v) - p_k \log x_k(\theta) \right) \\ .=& -\frac{n}{2} \sum_{k=1}^{q} p_k \log \left(1 + \frac{\varepsilon_n h_k}{x_k(\theta)} \right) \ . \end{aligned}$$

因此，前两项为：

$$\begin{aligned} & \sum_{k=1}^{q} p_k \log \frac{w_k(\theta+\varepsilon_n v)}{w_k(\theta)} - \frac{n}{2} \log \frac{|V_{n,\theta+\varepsilon_n v}|}{|V_{n,\theta}|} \\ =& \sum_{k=1}^{q} p_k \log \left(1 + \frac{\varepsilon_n g_k}{w_k(\theta)} \right) - \frac{n}{2} \sum_{k=1}^{q} p_k \log \left(1 + \frac{\varepsilon_n h_k}{x_k(\theta)} \right) \ . \end{aligned}$$

步骤 2：用总体矩阵表示现在回顾两个事实。首先， X 服从分布 $\mathbb{P}_{n,\theta}$，因此

$$X = V_{n,\theta}^{\frac{1}{2}} \mathbb{W} V_{n,\theta}^{\frac{1}{2}} \ .$$

其次，参数为（p， n）的白色 Wishart 矩阵可以写成秩1投影算子的和：

$$\mathbb{W} = \sum_{i=1}^{n} \xi_i \xi_i^* \ ,$$

其中， ξ_i 是 $\mathbb{R}^p$ 中具有标准高斯元素的独立同分布向量。因此，从迹的循环性质出发，第三项为：

$$\begin{aligned} & -\frac{1}{2} \operatorname{tr} \left((V_{n,\theta}^{\frac{1}{2}} V_{n,\theta+\varepsilon_n v}^{-1} V_{n,\theta}^{\frac{1}{2}} - \mathrm{id}) \mathbb{W} \right) \\ =& -\frac{1}{2} \sum_{i=1}^{n} \operatorname{tr} \left(\xi_i^* (V_{n,\theta}^{\frac{1}{2}} V_{n,\theta+\varepsilon_n v}^{-1} V_{n,\theta}^{\frac{1}{2}} - \mathrm{id}) \xi_i \right) \\ =& -\frac{1}{2} \sum_{i=1}^{n} \xi_i^* (V_{n,\theta}^{\frac{1}{2}} V_{n,\theta+\varepsilon_n v}^{-1} V_{n,\theta}^{\frac{1}{2}} - \mathrm{id}) \xi_i \ . \end{aligned}$$

由于总体矩阵 $V_{n,\theta}$ 可以共对角化，我们可以进一步简化上述表达式中的二次型：

$$\begin{aligned} & -\frac{1}{2} \operatorname{tr} \left((V_{n,\theta}^{\frac{1}{2}} V_{n,\theta+\varepsilon_n v}^{-1} V_{n,\theta}^{\frac{1}{2}} - \mathrm{id}) \mathbb{W} \right) \\ =& -\frac{1}{2} \sum_{i=1}^{n} (U\xi_i)^* \left(\frac{D_{n,\theta}^V}{D_{n,\theta+\varepsilon_n v}^V} - \mathrm{id} \right) U\xi_i \\ =& -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{p} \left[(U\xi_i)^j \right]^2 \left(\frac{(D_{n,\theta}^V)_{j,j}}{(D_{n,\theta+\varepsilon_n v}^V)_{j,j}} - 1 \right) \end{aligned}$$

$$= -\frac{1}{2}\sum_{j=1}^{p}\chi^2_{n,j}\left(\frac{(D^V_{n,\theta})_{j,j}}{(D^V_{n,\theta+\varepsilon_n v})_{j,j}}-1\right),$$

其中 $\chi^2_{n,j}$ 是参数为 n 的独立同分布（iid） χ^2 分布。

现在，我们需要按照多项分布（$p_1,\ldots,p_k$）对各项进行分组：

$$\begin{aligned}
&-\frac{1}{2}\operatorname{tr}\left((V^{\frac{1}{2}}_{n,\theta}V^{-1}_{n,\theta+\varepsilon_n v}V^{\frac{1}{2}}_{n,\theta}-\mathrm{id})\mathbb{W}\right)\\
=&-\frac{1}{2}\sum_{j=1}^{q}\left(\sum_{j:m_j=k}\chi^2_{n,j}\right)\left(\frac{x_k(\theta)}{x_k(\theta+\varepsilon_n v)}-1\right)\\
=&-\frac{1}{2}\sum_{j=1}^{q}\chi^2_{np_k,k}\left(\frac{x_k(\theta)}{x_k(\theta+\varepsilon_n v)}-1\right)\\
=&\frac{1}{2}\sum_{j=1}^{q}\chi^2_{np_k,k}\frac{\varepsilon_n h_k}{x_k(\theta+\varepsilon_n v)},
\end{aligned}$$

其中 $\chi^2_{\cdot,k}$ 仍然是相互独立的 χ^2 随机变量，具有指定的参数。

步骤3：分组。最终，我们得到：

$$\begin{aligned}
&\log\frac{d\mathbb{P}_{n,\theta+\varepsilon_n v}}{d\mathbb{P}_{n,\theta}}(\mathbb{X})\\
=&\beta n^2\sum_{k=1}^{q}\varepsilon_n g_k\log\frac{N_k}{n}-\frac{n}{2}\sum_{k=1}^{q}p_k\log\left(1+\frac{\varepsilon_n h_k}{x_k(\theta)}\right)\\
&+\frac{1}{2}\sum_{j=1}^{q}\chi^2_{np_k,k}\frac{\varepsilon_n h_k}{x_k(\theta+\varepsilon_n v)}\\
=&\beta n^2\sum_{k=1}^{q}\varepsilon_n g_k\log\frac{N_k}{n}-\frac{n}{2}\sum_{k=1}^{q}p_k\left(\log\left(1+\frac{\varepsilon_n h_k}{x_k(\theta)}\right)-\frac{\varepsilon_n h_k}{x_k(\theta+\varepsilon_n v)}\right)\\
&+\frac{1}{2}\sum_{j=1}^{q}\left(\chi^2_{np_k,k}-np_k\right)\frac{\varepsilon_n h_k}{x_k(\theta+\varepsilon_n v)}.
\end{aligned}$$

因为

$$\begin{aligned}
&\log\left(1+\frac{\varepsilon_n h_k}{x_k(\theta)}\right)-\frac{\varepsilon_n h_k}{x_k(\theta)+\varepsilon_n h_k}\\
=&\mathcal{O}(\varepsilon_n^3)+\frac{\varepsilon_n h_k}{x_k(\theta)}-\frac{\varepsilon_n^2 h_k^2}{2x_k(\theta)^2}-\frac{\varepsilon_n h_k}{x_k(\theta)}\left(1-\frac{\varepsilon_n h_k}{x_k(\theta)}+\mathcal{O}(\varepsilon_n^2)\right)\\
=&\mathcal{O}(\varepsilon_n^3)+\frac{\varepsilon_n^2 h_k^2}{2x_k(\theta)^2},
\end{aligned}$$

and:

$$\log\left(1+\frac{\varepsilon_n g_k}{w_k(\theta)}\right)=\mathcal{O}(\varepsilon_n^2)+\frac{\varepsilon_n g_k}{w_k(\theta)},$$

我们得到：

$$\begin{aligned}
&\log \frac{d\mathbb{P}_{n,\theta+\varepsilon_n v}}{d\mathbb{P}_{n,\theta}}(\mathbb{X}) \\
&= \mathcal{O}\left(p\varepsilon_n^2 + np\varepsilon_n^3\right) + \sum_{k=1}^{q} p_k \frac{\varepsilon_n g_k}{w_k(\theta)} - \frac{n}{2}\sum_{k=1}^{q} p_k \frac{\varepsilon_n^2 h_k^2}{2x_k(\theta)^2} \\
&\quad + \frac{1}{2}\sum_{j=1}^{q}\left(\chi^2_{np_k,k} - np_k\right)\frac{\varepsilon_n h_k}{x_k(\theta+\varepsilon_n v)} \\
&= \mathcal{O}\left(p\varepsilon_n^2 + np\varepsilon_n^3\right) + \varepsilon_n p \sum_{k=1}^{q} \frac{p_k}{p}\frac{g_k}{w_k(\theta)} - \frac{np}{4}\varepsilon_n^2 \sum_{k=1}^{q}\frac{p_k}{p}\frac{h_k^2}{x_k(\theta)^2} \\
&\quad + \frac{1}{2}\varepsilon_n n^{\frac{1}{2}} p^{\frac{1}{2}} \sum_{j=1}^{q} \frac{\chi^2_{np_k,k} - np_k}{\sqrt{np_k}}\frac{p_k}{p}\frac{h_k}{x_k(\theta+\varepsilon_n v)} \ .
\end{aligned}$$

步骤 4：极限。回顾这些极限：

$$\begin{aligned}
\lim_{n\to\infty} p_n/n &= c \ , \\
\lim_{n\to\infty} p_k/p &= w_k(\theta) \ , \\
\lim_{n\to\infty} \frac{\chi^2_\alpha - \alpha}{\alpha^{\frac{1}{2}}} &= \mathcal{N}_k.
\end{aligned}$$

第一个是一个假设，第二个是大数定律，第三个是标准的依分布收敛。

根据步骤 3 的最终表达式，我们看到，对于 $\varepsilon_n = \frac{1}{n}$，我们有收敛。极限为：

$$\begin{aligned}
&\lim_{n\to\infty} \log \frac{d\mathbb{P}_{n,\theta+\varepsilon_n v}}{d\mathbb{P}_{n,\theta}}(\mathbb{X}) \\
&= c\sum_{k=1}^{q} w_k(\theta)\frac{g_k}{w_k(\theta)} - \frac{1}{4}c\sum_{k=1}^{q} w_k(\theta)\frac{h_k^2}{x_k(\theta)^2} + \frac{1}{2}c^{\frac{1}{2}}\sum_{j=1}^{q}\mathcal{N}_k w_k(\theta)\frac{h_k}{x_k(\theta)} \\
&= \ -\frac{1}{4}c\sum_{k=1}^{q} w_k(\theta)\frac{h_k^2}{x_k(\theta)^2} + \frac{1}{2}c^{\frac{1}{2}}\sum_{j=1}^{q}\mathcal{N}_k w_k(\theta)\frac{h_k}{x_k(\theta)} \ .
\end{aligned}$$

这就是所述结果。

附录 A. 黎曼曲面与马尔可夫-克莱因的一般理论

由于我们关注的是对所发生情况的计算理解，因此将测度具体化为以下形式：

$$\mu_d := \sum_{j=1}^{d} w_j \delta_{x_j}(dx)$$

其中 w_j 是单纯形上的非负权重， x_j 给出支集。在这个特定情形下， $M : \mathbb{C}\backslash\mathbb{R} \to \mathbb{C}$ 是一个有理函数。自然地，它扩展为一个从黎曼球面 $\mathbb{S}$ 到其自身的映射。

经典地，这会产生如下的 d-重分歧覆盖 $\mathbb{S}$。临界点是 $z \in \mathbb{S}$ 这些点，在这些点处 M 不再局部可逆

$$Z := \{z \in \mathbb{S} \mid M'(z) = 0\} \ ,$$

而分歧点/分支点是像点 $m \in \mathbb{S}$:

$$M(Z) := \{m \in \mathbb{S} \mid \exists z \in Z,\ m = M(z)\} \ .$$

次数为 d，因为一般的 $m \in \mathbb{S}$ 通过 n 有 d 个原像。此外，$\mathrm{Card}\, Z = 2(d-1)$，事实上，临界点作为一个次数为 $2(d-1)$ 的多项式的根，互为复共轭。

第一类和第二类零点。 在本段中，我们借用一些实轴正交多项式（OPRL）的记号和术语。在实践中以及 FPT 理论中，我们所用的测度 μ_d 是参考测度 μ_∞ 的近似。在求积近似的背景下，请回忆:

$$G_{\mu_d}(z) = \int_{\mathbb{R}} \frac{\mu_d(dx)}{z-x} = \frac{Q_d(z)}{P_d(z)} \ ,$$

$$M_{\mu_d}(z) = \frac{zQ_d(z)}{P_d(z)} - 1 \ .$$

P_d 和 Q_d 的根，分别记为

$$(x_j \ ; \ 1 \le j \le d) \ \text{ and } \ (y_j \ ; \ 1 \le j \le d-1) \ ,$$

称为第一类和第二类零点。

马尔可夫–克雷因变换的定义是测度:

$$\nu_d = \delta_0 + \sum_{j=1}^{d-1} \delta_{y_j} - \sum_{j=1}^{d} \delta_{x_j} \ ,$$

并写作

$$M_{\mu_d}(z) = \exp\left(\int_{\mathbb{R}} \nu_d(dx) \log(z-x) \right) - 1 \ .$$

下面的引理表明，临界点是 ν_d 的柯西-斯蒂尔切斯变换的零点，而该变换是 μ_d（加上一个 δ_0）的马尔可夫–克雷因变换。

引理 A.1. 以下陈述等价:

$$M'(z) = -\int_{\mathbb{R}} \frac{\mu_d(dx)}{(z-x)^2} = -\sum_{j=1}^{d} \frac{w_j}{(z-x_j)^2} = 0$$

$$\int_{\mathbb{R}} \frac{\nu_d(dx)}{z-x} = \frac{1}{z} + \sum_{j=1}^{d-1} \frac{1}{z-y_j} - \sum_{j=1}^{d} \frac{1}{z-x_j} = 0 \ .$$

证明。写作 $M(z) = e^{F(z)}$，其中 $F(z) = \int_{\mathbb{R}} \nu_d(dx) \log(z-x)$。因此，$M'(z) = F'(z)e^{F(z)}$当且仅当为 0。注意到 F' 是所需的柯西–斯蒂尔切斯变换，即得证。□

注 A.2。 这不仅仅是一个有趣的现象，而且很有用，因为所有数值过程（辐角原理、牛顿-拉夫森法等）都必须专门针对柯西-斯蒂尔切斯变换进行设计。

此外，通过二分法也可以极快地计算第二类零点。

附录 B. 测度与矩

本节将讨论如何根据测度的矩对其进行数值近似。为此，我们先回顾一些概念。若存在某个实测度 μ，使得 $\forall n \geq 0\, m_n = \int x^n d\mu(x)$，则称序列 $(m_n)_{n\geq 0}$ 是一个（Hamburger）矩序列。特别地，以下等价刻画是众所周知的（需添加参考文献？）。

定理 B.1. 以下断言等价：

(1) $(m_n)^{n\geq 0}$ 是一个矩序列 *(2)*

Hankel 核

$$H = \begin{pmatrix} m_0 & m_1 & m_2 & \cdots \\ m_1 & m_2 & m_3 & \cdots \\ m_2 & m_3 & m_4 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

是半正定的。

（*3*）在连分式中，

$$\sum_{n=0}^{\infty} m_n z^n = \cfrac{1}{1 - \alpha_0 z - \cfrac{\beta_1 z^2}{1 - \alpha_1 z - \cfrac{\beta_2 z^2}{\ddots}}},$$

系数 β_n 对每个 $n \geq 1$ 都是非负的。

此外，

- $\mathrm{card}(\mathrm{supp}(\mu)) = \infty$ 当且仅当 H 是正定的
- $\mathrm{card}(\mathrm{supp}(\mu)) = n_0$ 当且仅当 $\beta_n > 0$ 对于任意 $n < n_0$ 和 $\beta_{n_0} = 0$。

B.1. 正交多项式的视角。设 μ 为一个测度（其所有矩均有限）。令 $m_n(\mu) = \int x^n d_\mu$ 并通过 $\langle g, f\rangle_\mu = \int f g d\mu$定义内积$\langle .\rangle_\mu$。可以定义与 μ相关的单位正交多项式集合（p_i）$_{i\geq 0}$，即：

- $\forall i \in \mathbb{N}, deg(p_i) = i$
- $\forall i \in \mathbb{N}, [x^i]p_i(x) = 1$
- $\forall i \neq j, \langle p_i, p_j\rangle_\mu = 0$。

不是令

$$H_n = \begin{pmatrix} m_0(\mu) & m_1(\mu) & m_2(\mu) & \cdots & m_{n-1}(\mu) \\ m_1(\mu) & m_2(\mu) & m_3(\mu) & \cdots & m_n(\mu) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1}(\mu) & m_n(\mu) & m_{n+1}(\mu) & \cdots & m_{2n-1}(\mu) \end{pmatrix}$$

命题 B.2。 乔列斯基分解 $H_n = LL^T$ 存在，且 L 的各项为 $L_{i,j} = \frac{\langle p_{j-1}}{\sqrt{\langle p_{j-1}}}$, $x^{i-1}\rangle_\mu$ $,p_{j-1}\rangle_\mu$.

证明。 根据构造， H_n 是对称的，而定理 B.1 保证其为半正定矩阵。因此，存在唯一的下三角矩阵 L_1 使得 $H_n = L_1 L_1^T$。我们只需验证命题中定义的矩阵 L 是下三角矩阵，并且满足 $H_n = LL^T$。

由于 $x^{i-1} \in \mathbb{R}_{i-1}[X]$ 由 $(p_\ell)_{\ell<i}$ 生成的向量空间，因此 $\langle p_{j-1}, x^{i-1}\rangle_\mu$ =只要 $j > i$ 就为 0，进而 L 是下三角矩阵。

此外，由于 $\{p_k\}_{0\le k\le n-1}$ 是 $\mathbb{R}_{n-1}[X]$ 的正交基，因此 H_n 的 (i, j) 项为

$$\begin{aligned} m_{i+j-2}(\mu) = \langle x^{i-1}, x^{j-1}\rangle_\mu &= \sum_{k=0}^{n-1} \frac{\langle x^{i-1}, p_k\rangle_\mu \langle x^{j-1}, p_k\rangle_\mu}{\langle p_k, p_k\rangle_\mu} \\ &= \sum_{k=0}^{n-1} L_{i,k} L_{j,k} \end{aligned}$$

这恰好是矩阵 LL^T 的 (i, j) 项，由此完成证明。

□

注意，将 H_n 替换为任意格拉姆矩阵，上述论证仍然成立。

B.2. 三项递推。 上述定义的正交多项式满足一个三项递推方程，

$$XP_n = P_{n+1} + a_n p_n + b_n p_{n-1}.$$

O一种方式 t要看 it is that $XP_n - P_{n+1}$ i作为一个 po lynomial of d次的 a t most n. 并且对于任意 $i \le n-2$、 $Xp_i \in \mathbb{R}_{n-1}[X]$ 和

$$\langle XP_n - P_{n+1}, p_i\rangle_\mu = \langle p_n - Xp_i\rangle_\mu + \langle p_{n+1}, p_i\rangle_\mu = 0.$$

在这种情况下，

$$\begin{aligned} \langle p_n, p_n\rangle_\mu = \langle p_n, Xp_{n-1}\rangle_\mu + \langle p_n, p_n - Xp_{n-1}\rangle_\mu &= \langle Xp_n, p_{n-1}\rangle_\mu \\ &= b_n \langle p_{n-1}, p_{n-1}\rangle. \end{aligned}$$

特别地， $b_n \ge 0$，由一个简单的递推关系可知， $\langle p_n$， $p_n\rangle = \mu$（$\mathbb{R}$）$^2 \prod_{i=1}^n b_i$.此外

$$L_{n,n} = \frac{\langle p_{n-1}, x^{n-1}\rangle_\mu}{\sqrt{\langle p_{n-1}, p_{n-1}\rangle_\mu}} = \frac{\langle p_{n-1}, p_{n-1}\rangle_\mu}{\sqrt{\langle p_{n-1}, p_{n-1}\rangle_\mu}} = \sqrt{\langle p_{n-1}, p_{n-1}\rangle_\mu} \text{ and}$$

$$b_n = \frac{\langle p_n, p_n\rangle_\mu}{\langle p_{n-1}, p_{n-1}\rangle_\mu} = \frac{L_{n+1,n+1}^2}{L_{n,n}^2}.$$

为恢复 a_n，我们有

$$\begin{aligned} L_{n+2,n+1} &= \frac{\langle X^{n+1}, P_n\rangle_\mu}{||p_n||_\mu} \\ &= \frac{\langle X^n, XP_n\rangle_\mu}{||p_n||_\mu} \\ &= \frac{\langle X^n, P_{n+1} + a_n P_n + b_n P_{n-1}\rangle_\mu}{||p_n||_\mu} \\ &= \frac{\langle X^n, P_{n+1}\rangle_\mu + a_n \langle X^n, P_n\rangle_\mu + b_n \langle X^n, P_{n-1}\rangle_\mu}{||p_n||_\mu} \end{aligned}$$

$$= a_n L_{n+1,n+1} + b_n L_{n+1,n} \frac{||p_{n-1}||_\mu}{||p_n||_\mu} = a_n L_{n+1,n+1} + \frac{L_{n+1,n}}{b_n}.$$

B.3.如何从雅可比系数恢复测度? 令$J_{(n)}=[J^{(n)}_{i,j}]_{i,j\leq n}$为雅可比矩阵，它是一个大小为 n的三对角矩阵，对角元素等于 $J^{(n)}_{i,i} = a_i$，非对角元素为 $J_{(n)i+1,i} = J_{(n)i,i+1} =b_i$。

我们将矩阵对角化为

$$J^{(n)} = V^* \operatorname{diag}\left(\lambda_i^{(n)}\right) V \ ,$$

其中， $\lambda_1^{(n)} \leq \lambda_2^{(n)} \leq \cdots \leq \lambda_n^{(n)}$是 $J^{(n)}$ 的特征值集合， $v_i^{(n)}$ 是与 $\lambda_i^{(n)}$ 对应的单位特征向量。于是我们得到以下结果。

命题 B.3. 如果 $card(supp(\mu)) \leq n$ 那么

$$\mu(dx) = \sum_{i=1}^{n} \langle v_i^{(n)}, e_1\rangle^2 \delta_{\lambda_i^{(n)}}(dx) \ .$$

参考文献

[和03] T.W. Anderson, 《多元统计分析导论》, Wiley 概率与统计丛书，Wiley, 2003年。[ATV20] Octavio Arizmendi、Pierre Tarrago 和 Carlos Vargas, 《自由反卷积的从属方法》, 亨利·庞加莱研究所年刊：概率与统计 56（2020）, 第4期，2565–2594。[BB07] Serban T Belinschi 和 Hari Bercovici, 《自由概率中从属结果的新方法》, 《数学分析杂志》101（2007）, 第1期，357–365。[BG22] Florent Benaych-Georges, 《Ledoit-P\'ech\'e 协方差矩阵公式的简短证明》, arXiv 预印本 arXiv:2201.05690（2022）。[BGBP19] Florent Benaych-Georges、Jean-Philippe Bouchaud 和 Marc Potters, 《交叉协方差矩阵奇异值的最优清理》, arXiv 预印本 arXiv:1901.05543（2019）。[BGD08] Florent Benaych-Georges 和 Mérouane Debbah, 《自由反卷积：从理论到实践》, 《生物启发式自主网络与服务范式》（2008）, 201–224。[BMS17] Serban T Belinschi、Tobias Mai 和 Roland Speicher, 《算子值自由加法卷积的解析从属理论及一般随机矩阵问题的解》, 《纯粹与应用数学杂志》2017（2017）, 第732卷，21–53。[EK08] Noureddine El Karoui, 《利用随机矩阵理论估计大维协方差矩阵的谱》, 《统计年刊》36（2008）, 第6期，2757–2790。[IHM13] Ildar Abdulovich Ibragimov 和 Rafail Zalmanovich Has' Minskii, 《统计估计：渐近理论》, 第16卷，Springer Science & Business Media，2013年。[Tar20] Pierre Tarrago, 《酉不变矩阵模型的谱反卷积》, 2020年。[VdV00] Aad W Van der Vaart, 《渐近统计学》, 第3卷，剑桥大学出版社，2000年。[Voi87] Dan Voiculescu, 《某些非交换随机变量的乘法》, 《算子理论杂志》（1987）, 223–235。

保罗·萨巴捷大学，图卢兹第三大学 – 图卢兹数学研究所（IMT）– 118, route de Narbonne, 31400,图卢兹,法国电子邮件地址: reda.chhaibi@math.univ-toulouse.fr

保罗·萨巴捷大学，图卢兹第三大学 – 图卢兹数学研究所（IMT）– 纳博讷路118号，31400，图卢兹，法国电子邮件地址：fabrice.gamboa@math.univ-toulouse.fr

保罗·萨巴捷大学，图卢兹第三大学 – 图卢兹数学研究所（IMT）– 纳博讷路118号，31400，图卢兹，法国电子邮件地址：slim.kammoun@math.univ-toulouse.fr

数学系，H-304，安第斯大学，波哥大，哥伦比亚电子邮件地址：mvelasco@uniandes.edu.co