arXiv:2304.04074v1 [math.ST] 8 Apr 2023

排列上的一类指数族的推断

SumitMukherjee [1]，**以及 Daiki Tagami**[2]

[1] *Department of Statistics, Columbia University, New York, USA; e-mail:* sm3949@columbia.edu

[2] *Department of Statistics, University of Oxford, Oxford, UK; e-mail:* daiki.tagami@hertford.ox.ac.uk

摘要：本文研究排列上的一类指数族，其中包括一些常见的 Mallows 模型。我们证明，该指数族自然参数的伪似然估计量渐近服从正态分布，且其方差有显式表达式。基于此，我们能够构造渐近有效的置信区间。我们还证明，同一问题的最大似然估计量处处相合，并在原点处渐近服从正态分布。在这一特殊情形下，高性价比伪似然估计量的渐近方差与高成本最大似然估计量的渐近方差相同。据我们所知，这是关于排列模型（包括 Mallows 模型）的首个推断结果，但不包括非常特殊的 Kendall τ 型 Mallows 模型。

关键词和短语：Mallows 模型、排列、伪似然、置信区间。

1. 引言

当比较多个项目时，排序数据会自然地出现在各种应用中。排序／排列空间中研究最为广泛的一类模型，是著名的 Mallows 模型，最早由 [24] 提出。此后，Mallows 模型（及其变体）在统计学（[9, 10, 13, 14, 15, 25, 28]）、概率论（[3, 6, 12, 17, 19, 29, 30]）和机器学习（[2, 8, 21, 22, 23, 26, 27]）领域受到了广泛关注。在[28], 中，作者在排列空间上引入了一类指数族模型，其中包含一些常见的 Mallows 模型。在这些模型中，通过最大似然估计量（MLE）估计参数在计算上不可行，因为指数族的归一化常数极难分析，而现有技术只能得到对数归一化常数的粗略（首阶）渐近性。为绕过这一难题，他们研究了基于伪似然方法（[4, 5]）对指数族中的自然参数进行估计，并推导了其估计量的相合速率。然而，伪似然估计量（PLE）的渐近分布问题仍未解决。获得渐近分布很有意义，因为这将使我们能够开展更多推断任务，例如进行假设检验以及构造

置信区间。

一个有趣的观察是，[28, 公式(1.1)]（另见本文的公式（7））与熵正则最优传输（EOT）存在联系。特别是，该模型的对数归一化常数的渐近性（见[28, 定理 1.5(a)]）产生了与熵正则最优传输中所研究的相同优化问题，但在置换情形下增加了均匀边缘分布的限制（见 [18, 公式 5]或 [16, 公式 1.1]）。这令人惊讶，因为这两个问题之间似乎不存在直接联系。EOT 中的工具可能会在这里发挥作用，但这超出了本文的范围。

本文提出了 [28]中所提出的指数族的多元推广。在这一一般设定下，我们研究了从该模型中观测到一个随机排列后，如何估计多元自然参数的问题。我们推导出多元伪似然估计在真实参数的所有取值下的渐近极限分布。特别地，这一结果适用于 [28]中研究的单参数模型。我们还证明，最大似然估计量在所有参数值下都是相合的。聚焦于原点，我们计算了最大似然估计量的渐近分布，结果表明，两种估计量的渐近方差相同。因此，至少在原点处，与计算上低效的最大似然估计量相比，计算上易于处理的伪似然估计在渐近表现方面同样出色。我们还利用伪似然估计构造渐近有效的置信区间。

当前草稿尚未涉及的一个问题是，在真实参数范围内，最大似然估计量的渐近分布。尽管最大似然估计量无法计算，但比较最大似然估计量与伪似然估计所产生的误差似乎很有意义。未来潜在研究的另一个方向，是从半参数的角度研究本文的所提出的模型，此时底层充分统计量的形式并非先验已知。在这种情况下，我们需要能够获得来自底层模型的多个独立排列样本，因为显然，在半参数设置下，仅有一个来自该模型的样本（如当前草稿中那样）是不够的。

现在我们将介绍一些符号，以便陈述我们的主要结果。

1.1. 记号

对于任意正整数 n，令 $[n] := \{1, 2, \ldots, n\}$。令 S_n 表示 $[n]$的所有排列组成的集合。给定任意排列 $\pi \in S_n$，我们可以写成 $\pi = (\pi(1) \ldots, \pi(n))$，其中 $\pi(i) \in [n]$ 是 i 在排列 π 下的像。

通过以下概率质量函数在 S_n 上定义一个指数族。

$$\mathbb{P}_{n,\boldsymbol{\theta}}(\pi) = \exp\Big(\boldsymbol{\theta}^\top \mathbf{T}(\pi) - Z_n(\boldsymbol{\theta})\Big). \tag{1}$$

Here (i)$\mathbf{T}(\pi) := \sum_{i=1}^n \mathbf{f}\Big(\frac{i}{n}, \frac{\pi(i)}{n}\Big)$是充分统计量，其中 $\mathbf{f} = (f_1, \ldots, f_L)$: $[0,1]^2 \mapsto \mathbb{R}^L$ 是一个向量值连续函数；(ii) $\boldsymbol{\theta} \in \mathbb{R}^L$ 是一个向量值实参数；(iii) $Z_n(\boldsymbol{\theta})$ 是对数归一化常数，给出为

$$Z_n(\boldsymbol{\theta}) := \log\left[\sum_{\pi \in S_n} \exp\Big(\sum_{i=1}^n \boldsymbol{\theta}^\top \mathbf{f}\Big(\frac{i}{n}, \frac{\pi(i)}{n}\Big)\Big)\right]. \tag{2}$$

在本文中，我们始终假设$\mathbf{f}$ 已完全确定，并将重点研究未知参数 $\boldsymbol{\theta}$的推断。由于 $\mathbf{f}$ 在本文中始终固定，因此我们在（$\mathbf{f}$ 1）中省略了对 的依赖。为简化记号，本文其余部分也遵循这一约定。如果函数（f_1，……， f_L）线性相关，则无法估计 $\boldsymbol{\theta}$，因为此时族 $\mathbb{P}_{n,\,\boldsymbol{\theta}}$ 在 $\boldsymbol{\theta}$ 中不可识别。因此，以下均假设$\mathbf{f}$ 线性独立，即不存在向量 $\mathbf{d} \in \mathbb{R}^L$ 使得 $\mathbf{d}^\top \mathbf{f} \overset{a.s.}{=} 0$。

另一个重要观察是，如果我们将函数 $f_r(x,y)$替换为 $\tilde{f}_r(x,y) = f_r(x,y) + a_r(x) + b_r(y)$，其中 $a_r(.)$ $b_r(.)$ 为任意函数，那么底层模型保持不变，因为

$$\sum_{i=1}^n \tilde{f}_r\Big(\frac{i}{n}, \frac{\pi(i)}{n}\Big) = \sum_{i=1}^n f_r\Big(\frac{i}{n}, \frac{\pi(i)}{n}\Big) + \sum_{i=1}^n \Big[a_r\Big(\frac{i}{n}\Big) + b_r\Big(\frac{i}{n}\Big)\Big].$$

这使我们能够通过以下假设来限制我们的函数类。

定义 1.1。 令 $\mathcal{C}$ 表示所有连续函数 $F : [0,1]^2 \mapsto \mathbb{R}$ ，满足

$$\int_0^1 F(.,z)dz = 0 \ \textit{and} \ \int_0^1 F(z,.)dz = 0, \tag{3}$$

且 F几乎处处不为 0。如果 $\mathbf{f} = (f_1, \ldots, f_L)$ 是一个向量函数，我们称 $\mathbf{f} \in \mathcal{C}$，如果 $f_r \in \mathcal{C}$ 对所有 $r \in [L]$ 均成立。

现在我们可以陈述主要结果了。

1.2. 主要结果

如上所述，对数归一化常数 $Z_n(\boldsymbol{\theta})$ 在（2）中没有闭式表达，而使用数值技术/MCMC 方法近似归一化常数是一个众所周知的挑战性问题。利用[28, 定理 1。5]）我们现在将给出 $Z_n(\boldsymbol{\theta})$ 的渐近估计。

定义 1.2。令 $\mathcal{M}$ 表示所有在 $[0,1]^2$上具有均匀边缘分布的概率测度所组成的空间，并令 $u \in \mathcal{M}$ 为 $[0,1]^2$上的均匀分布。对于单位正方形上的任意测度 μ 和有界可测函数 $F : [0,1] \mapsto \mathbb{R}$，令 $\mu(F) := \int_{[0,1]^2} F d\mu$。最后，令 $D(.||.)$ 表示同一空间上两个概率测度之间的 ***Kullback-Leibler*** 散度（通常为 $[0,1]^2$ 或 S_n）。

命题 1.3。 ***(a)*** 假设 $Z_n(\boldsymbol{\theta})$ 如 (2)所定义，且 $\mathbf{f}$ 连续。则有

$$\lim_{n\to\infty} \frac{Z_n(\boldsymbol{\theta}) - \log n!}{n} = \sup_{\mu\in\mathcal{M}} \left\{ \boldsymbol{\theta}^\top \mu(\mathbf{f}) - D(\mu||u) \right\} =: Z(\boldsymbol{\theta}).$$

(b) ***(a)*** 部分中的上确界在唯一的测度 $\mu_{\boldsymbol{\theta}} \in \mathcal{M}$ 处取得，该测度相对于勒贝格测度具有严格正的密度 $\rho_{\boldsymbol{\theta}}$（记为）。***(c)*** 在 $\mathbb{P}_{n,\boldsymbol{\theta}}$ 条件下，经验测度序列 $\frac{1}{n}\sum_{i=1}^n \delta_{\frac{i}{n},\frac{\pi(i)}{n}}$ 依概率弱收敛到测度 $\mu_{\boldsymbol{\theta}}$。因此，

$$\frac{\mathbf{T}(\pi)}{n} \xrightarrow{P} \mu_{\boldsymbol{\theta}}(\mathbf{f}) =: \mathbf{z}(\boldsymbol{\theta}).$$

(d) ***(c)*** 部分定义的函数 $\mathbf{z}(.) : \mathbb{R}^L \mapsto \mathbb{R}^L$ 在 $\mathbb{R}^L$ 上连续。***(e)*** ***(a)*** 部分定义的函数 $Z(.)$ 在 $\mathbb{R}^L$ 中可微，且

$$\nabla Z(\boldsymbol{\theta}) = \mathbf{z}(\boldsymbol{\theta}) = \mu_{\boldsymbol{\theta}}(\mathbf{f}).$$

(f) 若 $\mathbf{f}$ 线性无关，且 $\mathbf{f} \in \mathcal{C}$，（见定义 *1.1*），则函数 $Z(.)$ 严格凸，即对于任意 $\boldsymbol{\theta}_0 \neq \boldsymbol{\theta}_1$，有

$$Z(\boldsymbol{\theta}_1) > Z(\boldsymbol{\theta}_0) + (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^\top \mathbf{z}(\boldsymbol{\theta}_0). \tag{4}$$

注1。值得注意的是，命题 *1.3* 上述 ***(f)***部分中所示的极限对数归一化常数的严格凸性，直接转化为该模型的渐近可识别性。在定理 *1.7* 的***(a)*** 部分的证明中，我们将利用这一点，在那里我们证明最大似然估计量的一致性。这也可以通过使用 (16)来观察到：

$$Z(\boldsymbol{\theta}_1) - Z(\boldsymbol{\theta}_0) - (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^\top \mathbf{z}(\boldsymbol{\theta}_0) = \lim_{n\to\infty} \frac{1}{n} D(\mathbb{P}_{n,\boldsymbol{\theta}_0} || \mathbb{P}_{n,\boldsymbol{\theta}_1}),$$

其中 $D(.||.)$ 是定义 *1.2*中所述的 ***Kullback-Leibler***散度。

现在我们介绍 $\boldsymbol{\theta}$ 的伪似然估计量（伪似然估计）。

定义 1.4。令 $\mathcal{E}_n := \{(i,j) : 1 \leq i < j \leq n\}$。对于任意 $(i,j) \in \mathcal{E}_n$，

$\pi(i)$ 和 $\pi(j)$ 在给定 $\{\pi(k)\ k \neq i, j\}$ 时的条件分布为

$$\begin{aligned}
&\mathbb{P}_{n,\boldsymbol{\theta}}\Big(\pi(i)=\sigma(i),\pi(j)=\sigma(j)|\pi(k)=\sigma(k),k\neq i,j\Big)\\
=&\frac{\exp\left(\boldsymbol{\theta}^\top\mathbf{f}\left(\frac{i}{n},\frac{\sigma(i)}{n}\right)+\boldsymbol{\theta}^\top\mathbf{f}\left(\frac{j}{n},\frac{\sigma(j)}{n}\right)\right)}{\exp\left(\boldsymbol{\theta}^\top\mathbf{f}\left(\frac{i}{n},\frac{\sigma(i)}{n}\right)+\boldsymbol{\theta}^\top\mathbf{f}\left(\frac{j}{n},\frac{\sigma(j)}{n}\right)\right)+\exp\left(\boldsymbol{\theta}^\top\mathbf{f}\left(\frac{i}{n},\frac{\sigma(j)}{n}\right)+\boldsymbol{\theta}^\top\mathbf{f}\left(\frac{j}{n},\frac{\sigma(i)}{n}\right)\right)}\\
=&\frac{\exp\left(\boldsymbol{\theta}^\top\mathbf{y}_\sigma(i,j)\right)}{1+\exp\left(\boldsymbol{\theta}^\top\mathbf{y}_\sigma(i,j)\right)},
\end{aligned}\tag{5}$$

其中

$$\mathbf{y}_\sigma(i,j):=\mathbf{f}\Big(\frac{i}{n},\frac{\sigma(i)}{n}\Big)+\mathbf{f}\Big(\frac{j}{n},\frac{\sigma(j)}{n}\Big)-\mathbf{f}\Big(\frac{i}{n},\frac{\sigma(j)}{n}\Big)-\mathbf{f}\Big(\frac{j}{n},\frac{\sigma(i)}{n}\Big)$$

对于 $\sigma\in S_n$。然后， $\boldsymbol{\theta}$ 的伪似然定义为对所有对 $(i,j)\in\mathcal{E}_n$ 的这些条件分布求乘积，即

$$PL_n(\pi,\boldsymbol{\theta})=\prod_{(i,j)\in\mathcal{E}_n}\frac{\exp\left(\boldsymbol{\theta}^\top\mathbf{y}_\pi(i,j)\right)}{1+\exp\left(\boldsymbol{\theta}^\top\mathbf{y}_\pi(i,j)\right)},$$

然后，伪似然估计量（伪似然估计）通过求解该方程得到。

$$\mathbf{L}_n(\pi,\boldsymbol{\theta})=\nabla\log PL_n(\pi,\boldsymbol{\theta})=0.$$

直接计算可得

$$\mathbf{L}_n(\pi,\boldsymbol{\theta})=\nabla\log PL_n(\pi,\boldsymbol{\theta})=\sum_{(i,j)\in\mathcal{E}_n}\frac{1}{1+e^{\boldsymbol{\theta}^\top\mathbf{y}_\pi(i,j)}}\mathbf{y}_\pi(i,j).\tag{6}$$

在 [28, 定理 1。11] 中证明了，如果 $L=1$ (即 $\boldsymbol{\theta}=\theta\in\mathbb{R}$)，表达式 (6) ^ ^ 以趋于 1 的概率有唯一零点 θ_{PL}，并且满足 $\sqrt{n}(\theta_{PL}-\hat{}\theta)=O_P(1)$。关于 $\overline{\boldsymbol{\theta}}_{PL}$ 的渐近分布问题一直悬而未决，即使在一维情形下也是如此。我们的第一个主要结果解决了这一问题，并通过求解该问题更一般的多变量类比问题，给出了答案。

定理 1.5。 设 π 是模型 (1)中的随机排列，其中$\mathbf{f}$ 是线性独立的，并且$\mathbf{f}\in\mathcal{C}$。那么，记真实参数为$\boldsymbol{\theta}_0\in\mathbb{R}^L$，则有以下结论：

（*a*）设 $\mathbf{L}_n$（π， $\boldsymbol{\theta}$） 如 （6）所定义，则有

$$n^{-3/2}\mathbf{L}_n(\pi,\boldsymbol{\theta}_0)\xrightarrow{D}N\Big(0,\Sigma(\boldsymbol{\theta}_0)\Big),$$

其中

$$\Sigma_{pq}(\boldsymbol{\theta}):=\int_{[0,1]^6}\frac{g_p(\mathbf{z}_1,\mathbf{z}_2)g_q(\mathbf{z}_1,\mathbf{z}_3)}{(1+e^{\boldsymbol{\theta}^\top\mathbf{g}(\mathbf{z}_1,\mathbf{z}_2)})(1+e^{\boldsymbol{\theta}^\top\mathbf{g}(\mathbf{z}_1,\mathbf{z}_3)})}\prod_{a=1}^{3}d\mu_{\boldsymbol{\theta}}(\mathbf{z}_a)$$

对于 p, $q \in [L]$。这里的测度 $\mu_{\boldsymbol{\theta}}$ 如命题 1.3，且 $\mathbf{g} : [0,1]^4 \mapsto \mathbb{R}^L$ 定义为

$$\mathbf{g}((x_1,y_1),(x_2,y_2)) := \mathbf{f}(x_1,y_1) + \mathbf{f}(x_2,y_2) - \mathbf{f}(x_1,y_2) - \mathbf{f}(x_2,y_1).$$

(b) 向量方程 $\mathbf{L}_n(\pi, \boldsymbol{\theta}) = \mathbf{0}$ 存在唯一根 $\hat{}\ \boldsymbol{\theta_{PL}}$，其概率趋于1，并且满足

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{PL} - \boldsymbol{\theta}_0) \xrightarrow{D} N\Big(0, [A(\boldsymbol{\theta}_0)]^{-1}\Sigma(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1}\Big).$$

这里，$A(\boldsymbol{\theta})$ 是一个正定的 $L \times L$ 矩阵，定义为

$$A_{pq}(\boldsymbol{\theta}) := \frac{1}{2}\int_{[0,1]^4} \frac{g_p(\mathbf{z}_1,\mathbf{z}_2)g_q(\mathbf{z}_1,\mathbf{z}_2)e^{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{z}_1,\mathbf{z}_2)}}{(1+e^{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{z}_1,\mathbf{z}_2)})^2} \prod_{a=1}^{2} d\mu_{\boldsymbol{\theta}}(\mathbf{z}_a).$$

作为定理 1.5 的一个特例，令 $L = 1$ 即可立即得到以下推论。

推论 1.6。设 π 是来自模型的一个随机排列

$$\mathbb{P}_{n,\theta}(\pi) = \exp\Big(\theta \sum_{i=1}^{n} F\Big(\frac{i}{n}, \frac{\pi(i)}{n}\Big) - Z_n(\theta)\Big), \tag{7}$$

其中 $F \in \mathcal{C}$ 且 $\theta \in \mathbb{R}$。然后，将真实参数记为 $\theta_0 \in \mathbb{R}$，伪似然估计 $\hat{}\theta_{PL}$ 以趋于1的概率存在，并且满足

$$\sqrt{n}(\hat{\theta}_{PL} - \theta_0) \xrightarrow{D} N\Big(0, \frac{\sigma^2(\theta_0)}{a(\theta_0)}\Big),$$

其中

$$\sigma^2(\theta) := \int_{[0,1]^6} \frac{G(\mathbf{z}_1,\mathbf{z}_2)G(\mathbf{z}_1,\mathbf{z}_3)}{(1+e^{\theta G(\mathbf{z}_1,\mathbf{z}_2)})(1+e^{\theta G(\mathbf{z}_1,\mathbf{z}_3)})} \prod_{a=1}^{3} d\mu_\theta(\mathbf{z}_a),$$

$$a(\theta) := \frac{1}{2}\int_{[0,1]^4} \frac{G^2(\mathbf{z}_1,\mathbf{z}_2)e^{\theta G(\mathbf{z}_1,\mathbf{z}_2)}}{(1+e^{\theta G(\mathbf{z}_1,\mathbf{z}_2)})^2} \prod_{a=1}^{2} d\mu_\theta(\mathbf{z}_a).$$

Here

$$G((x_1,y_1),(x_2,y_2)) = F(x_1,y_1) + F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1),$$

且 μ_θ 如命题 1.3 第 ***(c)*** 部分，其中 $L = 1$。

注 2。我们在此指出，取 $F(x, y) = -|x-y|$ 在 (7) 中，我们得到以斯皮尔曼足距为充分统计量的 *Mallows* 模型；取$F(x, y) = -(x-y)^2$（或 $F(x, y) = xy$）在 (7) 中，我们得到以斯皮尔曼秩相关为充分统计量的 *Mallows* 模型。关于这些 *Mallows* 模型以及其他 *Mallows* 模型的更多背景，我们请参阅 *[11, 24, 28]*。

文献中讨论的模型。据我们所知，定理 *1.5* 和推论 *1.6* 是首批能够对模型 （1）进行假设检验等推断程序的结果，该模型包含这些 *Mallows* 模型。在我们的工作之前，我们所知的唯一推断结果是在以 *Kendall* τ 为充分统计量的 *Mallows* 模型中得到的；由于该模型具有显式的归一化常数，因此具有很强的可处理性。

一个自然的问题是，PLE 是否在渐近意义下最优，也就是说，它是否具有最小的渐近方差。尽管 MLE 无法计算，但就统计效率而言，它仍被视为估计量的“黄金标准”，至少对于诸如（1）这样的良好指数族而言。因此，我们可以问，能否比较 MLE 与 PLE 的表现。为此，我们的下一个结果表明，对于所有 $\boldsymbol{\theta}_0 \in \mathbb{R}^L$，MLE 都是相合的。分析 MLE 的渐近分布更为微妙，需要对数归一化常数的精确渐近性质。我们能够在原点的一个邻域内完成这一分析，从而得到 $\boldsymbol{\theta}_0 = \mathbf{0}$ 处 MLE 的中心极限定理。

定理 1.7。 设 π 是来自模型的一个随机排列 (1)，其中$\mathbf{f} = (f_1, \dots, f_L)$ 线性独立，$\mathbf{f} \in \mathcal{C}$，且真实参数为$\boldsymbol{\theta}_0 \in \mathbb{R}^L$。则有以下结论：

(a) 方程 $\mathbf{T}(\pi) = \nabla Z_n(\boldsymbol{\theta})$ 存在唯一根 $\hat{} \ \boldsymbol{\theta_{ML}}$ ，且其成立的概率趋于1，并且满足

$$\hat{\boldsymbol{\theta}}_{ML} \xrightarrow{P} \boldsymbol{\theta}_0.$$

(b) 如果 $\boldsymbol{\theta}_0 = \mathbf{0}$，那么我们有

$$\frac{\mathbf{T}(\pi) - \nabla Z_n(\mathbf{0})}{\sqrt{n}} \xrightarrow{D} N\Big(\mathbf{0}, \Gamma\Big), \ \textit{where } \Gamma_{p,q} := \int_{[0,1]^2} f_p(x,y) f_q(x,y) dx dy.$$

(c) 如果 $\boldsymbol{\theta}_0 = \mathbf{0}$，那么我们有

$$\sqrt{n}\hat{\boldsymbol{\theta}}_{\boldsymbol{ML}} \xrightarrow{D} N\Big(\mathbf{0}, \Gamma^{-1}\Big).$$

比较定理 1.5 和 1.7可知，当 $\boldsymbol{\theta}_0 = \mathbf{0}$时，伪似然估计和最大似然估计量的渐近分布相同，如下述推论所示。

推论 1.8. 设 $A(\boldsymbol{\theta})$ $\Sigma(\boldsymbol{\theta})$ 和 Γ 如定理 *1.5* 和定理*1.7*中所定义，则有

$$A(\mathbf{0})^{-1}\Sigma(\mathbf{0})A(\mathbf{0})^{-1} = \Gamma^{-1}.$$

$\hat{}$ $\hat{}$因此，如果真实参数 $\boldsymbol{\theta}_0 = \mathbf{0}$，那么 $\boldsymbol{\theta}_{PL}$ 和 $\boldsymbol{\theta}_{ML}$ 具有相同的渐近分布。

注 3. 对于所有 $\boldsymbol{\theta}_0 \in \mathbb{R}^L$，*PLE* 的表现是否能与 *MLE* 相匹配，仍有待考察。主要的理论瓶颈在于，对于指数族 $\mathbb{P}_{n,\boldsymbol{\theta}}$，尚无关于 $\mathbf{T}(\pi)$ 的中心极限定理，而这本身就是一个有趣的

问题，且由于它与熵正则最优传输的联系（如引言所述），更值得研究。具体而言，*[18]* 推导出了 *EOT* 中出现的、定义在 $[0,1]^{2n}$ 上的相关指数族的中心极限定理；或许可以将其中使用的技术扩展到模型 $\mathbf{T}(\pi)$ 下，证明关于 (7) 的中心极限定理。

一个相关的推断问题是构造置信区间。接下来的结果为参数 $\mathbf{d}^\top\boldsymbol{\theta}$ 构造了一个渐近有效的置信区间，其中 $\mathbf{d} \neq \mathbf{0}$ 是 $\mathbb{R}^L$ 中的一个已知向量。

引理 1.9。对于 $1 \leq i < j \leq n$ 和 p、$q \in [L]$，令

$$\widehat{\Sigma}_{p,q}(\boldsymbol{\theta}) = \frac{1}{n^3} \sum_{(i,j),(k,\ell)\in\mathcal{E}_n:|(i,j)\cap(k,\ell)|=1} \frac{y_{\pi,p}(i,j)y_{\pi,q}(i,k)}{(1+e^{\boldsymbol{\theta}^\top\mathbf{y}_\pi(i,j)})(1+e^{\boldsymbol{\theta}^\top\mathbf{y}_\pi(i,k)})},$$

$$\widehat{A}_{p,q}(\boldsymbol{\theta}) = \frac{1}{n^2} \sum_{(i,j)\in\mathcal{E}_n} \frac{y_{\pi,p}(i,j)y_{\pi,q}(i,j)e^{\boldsymbol{\theta}^\top\mathbf{y}_\pi(i,j)}}{(1+e^{\boldsymbol{\theta}^\top\mathbf{y}_\pi(i,j)})^2},$$

其中 $\mathbf{y}_\pi(i,j)$ 如定义 *1.4*中所述。然后，对于任意 $\alpha \in (0,1)$ 和 $\mathbf{d} \in \mathbb{R}^L$，有

$$\lim_{n\to\infty} \mathbb{P}_{n,\boldsymbol{\theta}_0}\left(\left|\mathbf{d}^\top\hat{\boldsymbol{\theta}}_{PL} - \mathbf{d}^\top\boldsymbol{\theta}_0\right| \leq \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} \cdot \sqrt{\mathbf{d}^\top\left[[\widehat{A}(\hat{\boldsymbol{\theta}}_{PL})]^{-1}\widehat{\Sigma}(\hat{\boldsymbol{\theta}}_{PL})[\widehat{A}(\hat{\boldsymbol{\theta}}_{PL})]^{-1}\right]\mathbf{d}}\right) = 1-\alpha,$$

其中 z_α 是标准正态分布的$(1-\alpha)$分位数。

1.3. 模拟结果

为说明我们的结果，我们聚焦于模型（7）的一个具体示例，其中$F(x,y) = xy$。这对应于具有斯皮尔曼秩相关的 Mallows 模型。在这种情况下，极限测度 μ_θ 相对于勒贝格测度在 $[0,1]^2$, 上的密度具有如下形式

$$e^{\theta xy + a_\theta(x) + a_\theta(y)},$$

额外的对称性源于 F 在（x，y）中具有对称性（参见[28, 第 2] 节）。函数 a_θ（.）由上述密度具有均匀边缘分布这一要求唯一确定。

为了展示本文的结果，我们需要能够高效地从该模型中进行模拟。在 [1] 中，作者推导出一种辅助变量命中与运行算法，用于从该模型中进行模拟，具体说明如下：

(i) 从 S_n 中均匀随机地模拟排列 π。(ii) 给定 π，模拟相互独立的随机变量 $\{U_i\}_{i=1}^n$，其中U_i 在 $[0, e^{(\theta/n^2)i\pi(i)}]$ 上服从均匀分布。

(iii) 给定 $\{U_i\}_{i=1}^n$，令 $b_j := \max\{(n^2/\theta j)\log U_j, 1\}$。从集合 $\{j \in [n] : b_j \le 1\}$ 中均匀随机地选择一个索引 i_1，并令 $\pi(i_1) = 1$。从 $[n]$ 中移除该索引，并从集合 $\{j \in [n]: b_j \le 2\} - \{i_1\}$ 中均匀随机地选择一个索引 i_2，并令 $\pi(i_2) = 2$。一般而言，在定义了索引 $\{i_1, \dots, i_{l-1}\}$ 后，将它们从 $[n]$ 中移除。从索引集合 $\{j \in [n]: b_j \le \ell\} - \{i_1, \dots, i_{l-1}\}$ 中均匀随机地选择 i_ℓ，并令 $\pi(i_\ell) = \ell$。(iv) 在步骤 2 和 3 之间迭代，直至收敛。

对于我们的模拟，我们将步骤 2 和 3 总共迭代 10 次，以获得单个排列 π。给定排列 π，我们使用二分法计算伪似然估计。为了考察伪似然估计的分布，我们将上述过程重复 2000 次，置换规模分别为 $n = 500$、2000、8000，真实参数为 $\theta = 2$。由此得到的三个直方图叠加显示于下方图 1 中，分别以粉色、绿色和蓝色表示。从图中，我们

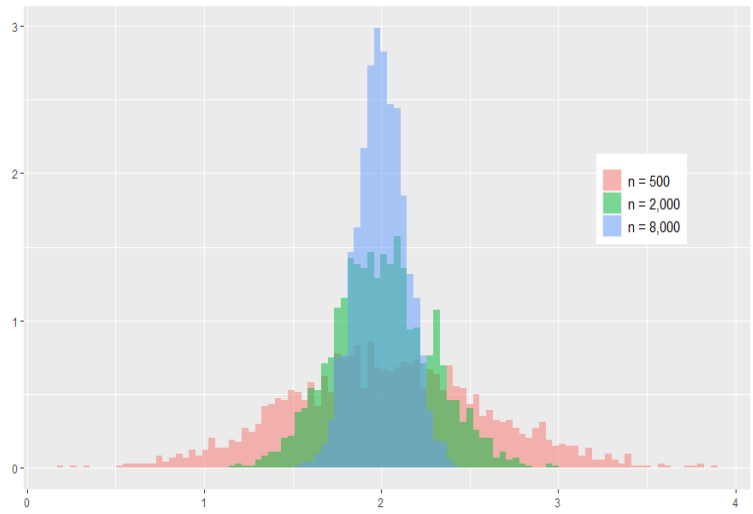

FIG 1. *Histogram of the PLE in the Spearman's rank correlation model, with permutation size* $n = 500$ 2000 8000 *and* $\theta_0 = 2$ *based on* 2000 *samples.*

可以看到，三个直方图都呈钟形，中心约为2。这与定理 1.5一致，该定理表明伪似然估计渐近服从正态分布。此外，随着 n 的增加，直方图的宽度减小，这与伪似然估计的方差随 n增加而减小这一事实一致。

为了比较伪似然估计与最大似然估计量的性能，我们关注相同的模型，但改为集合 $\theta_0 = 0$。请注意，计算最大似然估计量颇具挑战，因此我们采用以下启发式近似：

在定理1.7的 (c) 部分的证明中使用 (21)，可得若 $\theta_0 = 0$,

对于任意 $u \in \mathbb{R}$，我们有

$$Z'_n\left(\frac{u}{\sqrt{n}}\right) \approx Z'_n(0) + u\gamma\sqrt{n}, \quad \gamma = \int_{[0,1]^2} F^2(x,y)dxdy.$$

令 $u = U_n = \sqrt{n}\theta_{ML}$ 利用 $\sqrt{n}$ 在 $\theta_0 = 0$ 下最大似然估计量的一致性，我们有 $U_n = O_P(1)$，因此上述公式给出

$$T(\pi) = Z'_n\Big(\hat{\theta}_{ML}\Big) = Z'_n\Big(\frac{U_n}{\sqrt{n}}\Big) \approx Z'_n(0) + n\gamma\hat{\theta}_{ML},$$

其中 $T(\pi) = \sum_{i=1}^n F\left(\frac{i}{n}, \frac{\pi(i)}{n}\right)$. 这进而给出简单近似

$$\hat{\theta}_{ML} \approx \frac{T(\pi) - Z'_n(0)}{n\gamma}, \tag{8}$$

这很容易计算。在我们的例子中，我们有

$$Z'_n(0) = \mathbb{E}_{n,\theta=0}T(\pi) = \sum_{i=1}^n \frac{i}{n^2}\mathbb{E}_{n,\theta=0}[\pi(i)] = \frac{(n+1)^2}{4n}, \qquad \gamma = \int_{[0,1]^2} x^2y^2dxdy = \frac{1}{9}.$$

代入这些值，可以使用（8）计算最大似然估计量的近似值。我们注意到，从技术上讲，这并不是一个估计量，因为我们利用了对 $\theta_0 = 0$ 的了解来证明该启发式方法的合理性。

在图 2中，我们比较了该近似最大似然估计量和伪似然估计的直方图，其中底层真实参数为 $\theta_0 = 0$ （即 π 是一个均匀随机排列）。我们模拟了 2000 个排列 π，其大小分别为 $n = 1000$、2000、4000、8000，从而得到四组这样的对比直方图（每个 n值对应一组）。在上述每幅图中，我们都能看到两个直方图相互重叠，这表明当底层真实参数设为 $\boldsymbol{\theta_0 = 0}$时，最大似然估计量和伪似然估计的渐近分布相同。此外，正如预期的那样，我们看到随着排列n 规模增大，直方图的宽度会缩小。

最后，我们为参数构造一个 95% 置信区间 θ_0，其中真实参数为 $\theta_0 = 2$，排列的大小为 $n = 1000$。每个置信区间的构造都基于一个排列，因此本质上，我们是基于一个样本（当然，样本大小为 $n!$ ）进行推断。我们重复构造该置信区间 100 次，并在图3中绘出这些置信区间。结果表明，在这样构造的 100 个置信区间中，有 95 个包含真实值 θ_0 =2。这验证了引理1.9的结论：当 n 较大时，该置信区间的覆盖率约为 95%。

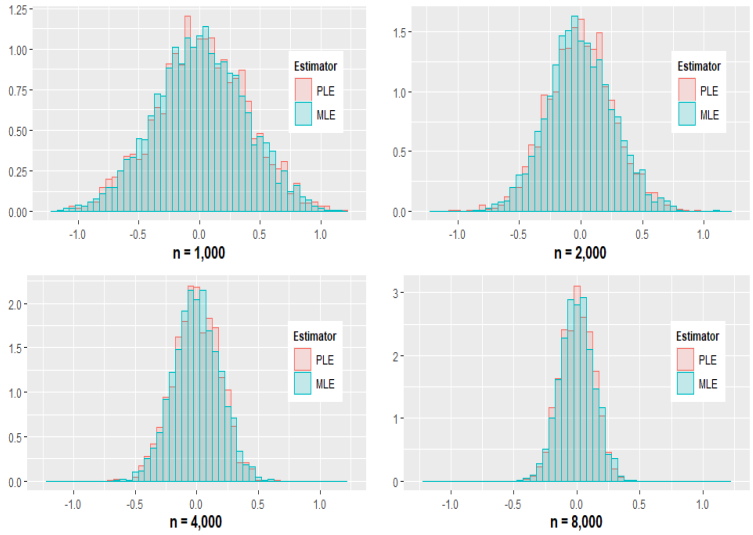

FIG 2. *Histogram of MLE and PLE in the Spearman's rank correlation model, with permutation size* $n = 1000\ 2000\ 4000\ 8000$ *and* $\theta_0 = 0$ *based on* 2000 *samples.*

1.4. 文章概述

在第2 节中，我们证明本文引言中陈述的主要结果。这些结果的证明借助若干辅助引理完成，其证明将在第3节中给出。

2. 主要结果的证明

我们首先介绍以下三个引理，其证明推迟至第 3 节。前两个引理计算 $\mathbf{L}_n(\pi, \boldsymbol{\theta})$ 的线性函数的矩。这些结果将用于定理 1.5 的证明。

引理 2.1. 对于 $(i,j) \in \mathcal{E}_n$ 以及任意 $\mathbf{d} \in \mathbb{R}^L$，令

$$\mathbf{C}_{ij}(\pi) := \frac{1}{1+\exp\left(\boldsymbol{\theta}_0^\top \mathbf{y}_\pi(i,j)\right)}\mathbf{y}_\pi(i,j) \in \mathbb{R}^L, \text{ and } C_{ij}(\pi,\mathbf{d}) = \mathbf{d}^\top \mathbf{C}_{ij}(\pi),$$

我们有

$$\mathbb{E}_{n,\boldsymbol{\theta}_0}\Big(C_{ij}(\pi,\mathbf{d})\Big|\pi(\ell), \ell \neq i,j\Big) = 0.$$

引理2.2. 令 $\{D_{ij},(i,j) \in \mathcal{E}_n\}$ 为联合中心化多元高斯，且 $Cov(D_{ij}, D_{k\ell}) = \mathbb{E}_{n,\boldsymbol{\theta}_0} C_{ij}(\pi,\mathbf{d})C_{k\ell}(\pi,\mathbf{d})$。因此，令 $D_n := \sum_{(i,j)\in\mathcal{E}_n} D_{ij}$，对于任意 $k \in \mathbb{N}$，我们有

$$\lim_{n\to\infty} \frac{\mathbb{E}_{n,\boldsymbol{\theta}_0}\Big(\sum_{(i,j)\in\mathcal{E}_n} C_{ij}(\pi,\mathbf{d})\Big)^k - \mathbb{E}D_n^k}{n^{\frac{3k}{2}}} = 0.$$

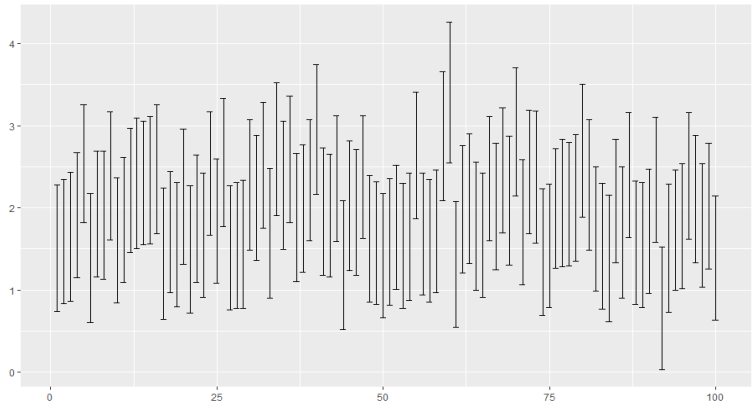

FIG 3. *Confidence Interval based on the PLE in the Spearman's rank correlation model, with permutation size* $n = 1000$ *and* $\theta_0 = 2$ *based on* 100 *separate samples.*

第三个引理证明了 M估计量的一般渐近正态性（也适用于依赖情形），并将用于证明 PLE（定理 1.5 的 (b) 部分）和 MLE（定理 1.7 的 (c) 部分）的渐近正态性。尽管该引理是针对定义在 S_n上的函数陈述的，以便于应用，但当 S_n 替换为更一般的样本空间时，证明仍可原样适用。

引理 2.3。 设 $\mathcal{S}$： $S_n \times \mathbb{R}^L \mapsto \mathbb{R}$ 是一个函数，使得对每个 $\pi \in S_n$，映射 $\mathcal{S}(\pi, .)$ 均为 $\mathcal{C}^2$，且海森矩阵 $H(\pi, \boldsymbol{\theta}) \in \mathbb{R}^{L \times L}$ 是半正定的。固定 $\boldsymbol{\theta}_0 \in \mathbb{R}^L$，进一步假设在 $\mathbb{P}_{n,\ \boldsymbol{\theta}_0}$ 下满足以下条件：*(i)* 梯度 $\nabla \mathcal{S}(\pi, \boldsymbol{\theta})$ 满足

$$n^{-1/2} \nabla \mathcal{S}(\pi, \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, B_1).$$

(ii) 对任意收敛至 $\mathbf{h}$ 的序列 $\mathbf{h}_n \in \mathbb{R}^L$，海森矩阵 $H(\pi, .)$ 满足

$$\frac{1}{n} H(\pi, \boldsymbol{\theta}_0 + n^{-1/2} \mathbf{h}_n) \to B_2,$$

其中 B_2 正定，且与$\mathbf{h}$无关。于是，随着概率趋于1，存在唯一的 $\boldsymbol{\theta}$，满足 $\nabla \mathcal{S}(\pi, \boldsymbol{\theta}) = \mathbf{0}$。此外，我们有

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, B_2^{-1} B_1 B_2^{-1}).$$

类似地，第四个引理给出了关于M估计量的一般一致性结果，该结果同样适用于依赖情形。我们将利用它证明1.7部分(a)中最大似然估计量的一致性。与上文一样，该引理可以很容易地扩展，以涵盖样本空间不是 S_n的情形。

引理2.4。 设 $\mathcal{S}$： $S_n \times \mathbb{R}^L \mapsto \mathbb{R}$ 是一个关于第二个参数严格凸的函数。设 $\boldsymbol{\theta}_0 \in \mathbb{R}^L$ 满足在$\mathbb{P}_{n,\ \boldsymbol{\theta}_0}$下的以下条件：

(i) 对每个 $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$，存在 $\varepsilon := \varepsilon_{\theta_1} > 0$ 使得

$$\mathbb{P}_{n,\boldsymbol{\theta}_0}\Big(\mathcal{S}(\pi,\boldsymbol{\theta}_0) + \varepsilon < \mathcal{S}(\pi,\boldsymbol{\theta}_1)\Big) = 1.$$

(ii) 存在一个随机变量 $M(\pi)$，它是 $O_P(1)$，使得对于所有$\boldsymbol{\theta}_1$、$\boldsymbol{\theta}_2 \in \mathbb{R}^L$，我们都有

$$|\mathcal{S}(\pi,\boldsymbol{\theta}_1) - \mathcal{S}(\pi,\boldsymbol{\theta}_2)| \leq M(\pi)\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2.$$

然后，存在唯一的极小化元 $\hat{}\ \boldsymbol{\theta}_n$，它是函数 $\mathcal{S}(\pi,.)$ 在 $\mathbb{R}^L$ 中的极小化元，其概率论趋近于 1。进一步地，我们有

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0.$$

利用这些引理，我们现在完成定理 1.5、定理 1.7，以及引理1.9。

2.1. 定理的证明 *1.5*

(a)对于第 (a) 部分，只需证明对于任意 $\mathbf{d} \in \mathbb{R}^L$ 我们有

$$\frac{1}{n\sqrt{n}}\mathbf{d}^\top \mathbf{L}_n(\pi,\boldsymbol{\theta}_0) \xrightarrow{D} N\left(0, \boldsymbol{d}^\top \Sigma(\boldsymbol{\theta}_0)\boldsymbol{d}\right).$$

根据 $C_{ij}(\pi,\mathbf{d}$，如引理2.1中所定义，我们有

$$\mathbf{d}^\top \mathbf{L}_n(\pi,\boldsymbol{\theta}_0) = \sum_{(i,j)\in\mathcal{E}_n} C_{ij}(\pi,\mathbf{d}).$$

应用引理 2.2，只需证明 $\frac{D_n}{n\sqrt{n}} \xrightarrow{D} N(0,\ \boldsymbol{d}^\top\Sigma\boldsymbol{d})$。~~由于~~$D_n$ 按构造是一个中心高斯随机变量，因此只需证明

$$\frac{Var_{n,\boldsymbol{\theta}_0}(D_n)}{n^3} = \frac{1}{n^3}\sum_{(i,j),(k,\ell)\in\mathcal{E}_n} \mathbb{E}_{n,\boldsymbol{\theta}_0} C_{ij}(\pi,\mathbf{d})C_{k\ell}(\pi,\mathbf{d}) \to \boldsymbol{d}^\top\Sigma(\boldsymbol{\theta}_0)\boldsymbol{d}. \tag{9}$$

继续验证上述公式，结合 [29, 推论 1.12] 和 [29,定理 1.4]，我们得出对于任意 $\ell \in \mathbb{N}$，都有

$$\lim_{n\to\infty} \sup_{\mathbf{p},\mathbf{q}\in S(n,\ell)} \left| \frac{n^\ell \mathbb{P}_{n,\boldsymbol{\theta}_0}\Big(\pi(p_a) = q_a, a \in [\ell]\Big)}{\prod_{a=1}^{\ell} \rho_{\boldsymbol{\theta}}\left(\frac{p_a}{n}, \frac{q_a}{n}\right)} - 1 \right| = 0. \tag{10}$$

在上面的显示中，$S(n,\ \ell) \subset [n]^\ell$ 是所有 ℓ 由不同条目组成的元组的集合，并且 $\rho_{\boldsymbol{\theta}}$ 如命题1.3中所定义。（10）实质上意味着，对于任意有限的指标集合 $\{p_1,\ldots,p_\ell\}$，随机

变量 $\{\pi(p_1)\ldots,\pi(p_\ell)\}$ 近似独立，并确定了它们的联合极限分布。利用引理 2.1，可知如果 $|(i,j)\cap(k,\ell)|=0$，那么

$$\mathbb{E}_{n,\boldsymbol{\theta}_0}C_{ij}(\pi,\mathbf{d})C_{k\ell}(\pi,\mathbf{d})=0. \tag{11}$$

此外，显然

$$\Big|\sum_{(i,j),(k,\ell)\in\mathcal{E}_n:|(i,j)\cap(k,\ell)|=2}\mathbb{E}_{n,\boldsymbol{\theta}_0}C_{ij}(\pi,\mathbf{d})C_{k\ell}(\pi,\mathbf{d})\Big|=O(n^2). \tag{12}$$

因此，式（9）中的主导项来自以下形式的项$\mathbb{E}_{n,\boldsymbol{\theta}_0}C_{ij}(\pi,\mathbf{d})C_{k\ell}(\pi,\mathbf{d})$，这些项满足 $|(i,j)\cap(k,\ell)|=1$。利用式（10），得到

$$\begin{aligned}&\sum_{(i,j),(k,\ell)\in\mathcal{E}_n:|(i,j)\cap(k,\ell)|=1}\mathbb{E}_{n,\boldsymbol{\theta}_0}C_{ij}(\pi,\mathbf{d})C_{k\ell}(\pi,\mathbf{d})\\ \to&\sum_{p,q=1}^{L}d_pd_q\int_{[0,1]^6}\frac{g_p(\mathbf{z}_1,\mathbf{z}_2)}{1+e^{\boldsymbol{\theta}_0^\top\mathbf{g}(\mathbf{z}_1,\mathbf{z}_2)}}\frac{g_q(\mathbf{z}_1,\mathbf{z}_3)}{1+e^{\boldsymbol{\theta}_0^\top\mathbf{g}(\mathbf{z}_1,\mathbf{z}_3)}}\prod_{a=1}^{3}d\mu_{\boldsymbol{\theta}_0}(\mathbf{z}_a)=\boldsymbol{d}^\top\Sigma(\boldsymbol{\theta}_0)\boldsymbol{d}.\end{aligned} \tag{13}$$

其中 $\mathbf{g}(.,.)$ 与定理 1.5的陈述中相同。结合式（11）和式（13），即可验证式（12）式（9），从而完成了 (a) 部分的证明。

(b) 对于 (b) 部分，我们将采用引理 2.3 ，并作如下选择:

$$\mathcal{S}(\pi,\boldsymbol{\theta})=-n^{-1}\log PL_n(\pi,\boldsymbol{\theta}),\quad B_1=\Sigma(\boldsymbol{\theta}_0),\quad B_2=A(\boldsymbol{\theta}_0).$$

在这种情况下，我们有

$$n^{-1/2}\nabla\mathcal{S}(\pi,\boldsymbol{\theta}_0)=-n^{-3/2}\mathbf{L}_n(\pi,\boldsymbol{\theta}_0)\xrightarrow{D}N\Big(\mathbf{0},\Sigma(\boldsymbol{\theta}_0)\Big)$$

根据 (a) 部分，假设 (i) 成立，因此引理 2.3 的假设 (i) 成立。求二阶导数可得

$$H_{pq}(\pi,\boldsymbol{\theta})=\frac{\partial^2\mathcal{S}(\pi,\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q}=n^{-1}\sum_{(i,j)\in\mathcal{E}_n}y_{\pi,p}(i,j)y_{\pi,q}(i,j)\frac{e^{\boldsymbol{\theta}^\top\mathbf{y}_\pi(i,j)}}{(1+e^{\boldsymbol{\theta}^\top\mathbf{y}_\pi(i,j)})^2}.$$

因此，海森矩阵 $H(\pi,\boldsymbol{\theta})$ 是半正定的。此外，若$\{\mathbf{h}_n\}_{n\geq1}$ 是 $\mathbb{R}^L$ 中收敛于$\mathbf{h}$的一个序列，则利用（10）可得

$$\begin{aligned}&\frac{1}{n}H_{pq}(\pi,\boldsymbol{\theta}_0+n^{-1/2}\mathbf{h}_n)\\ \xrightarrow{P}&\frac{1}{2}\int_{[0,1]^4}g_p(\mathbf{z}_1,\mathbf{z}_2)g_q(\mathbf{z}_1,\mathbf{z}_2)\frac{e^{\boldsymbol{\theta}_0^\top\mathbf{g}(\mathbf{z}_1,\mathbf{z}_2)}}{(1+e^{\boldsymbol{\theta}_0^\top\mathbf{g}(\mathbf{z}_1,\mathbf{z}_2)})^2}\prod_{a=1}^{2}d\mu_{\boldsymbol{\theta}_0}(\mathbf{z}_a)=A_{pq}(\boldsymbol{\theta}_0),\end{aligned} \tag{14}$$

其中 $A(.)$ 的定义如定理 1.5所述。为验证假设(ii)，还需证明 $A(\boldsymbol{\theta})$ 是正定的。为此，对任意向量 $\mathbf{b}$ $\boldsymbol{\theta} \in \mathbb{R}^L$ ，有

$$\begin{aligned}\boldsymbol{b}^\top A(\boldsymbol{\theta})\boldsymbol{b} &= \frac{1}{2}\sum_{p,q=1}^{L} b_p b_q \int_{[0,1]^4} \frac{g_p(\mathbf{z}_1,\mathbf{z}_2)g_q(\mathbf{z}_1,\mathbf{z}_2)e^{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{z}_1,\mathbf{z}_2)}}{(1+e^{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{z}_1,\mathbf{z}_2)})^2} \prod_{a=1}^{2} d\mu_{\boldsymbol{\theta}}(\mathbf{z}_a) \\ &= \frac{1}{2}\int_{[0,1]^4} \frac{\left(\sum_{r=1}^{L} b_r g_r(\mathbf{z}_1,\mathbf{z}_2)\right)^2 e^{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{z}_1,\mathbf{z}_2)}}{(1+e^{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{z}_1,\mathbf{z}_2)})^2} \prod_{a=1}^{2} d\mu_{\boldsymbol{\theta}}(\mathbf{z}_a)\end{aligned}$$

显然 $A(\boldsymbol{\theta})$是非负定的，并且要验证正定性，只需证明上面的右端不为 0。如果该值为 0，那么我们必有

$$\sum_{r=1}^{L} b_r g_r(\mathbf{z}_1,\mathbf{z}_2) \overset{a.s.}{=} 0 \tag{15}$$

关于 $\mu_{\boldsymbol{\theta}^{(0)}}^{\otimes 2}$。应用命题 1.3 的第 (b) 部分可得， $\mu_{\boldsymbol{\theta}^{(0)}}$相对于 勒贝格测度在 $[0,1]^2$,上具有严格正的密度，因此，关系式 (15) 相对于 勒贝格测度在 $[0,1]^4$ 上也成立。对 $\mathbf{z}_2$ 积分，并利用假设 $f_r \in \mathcal{C}$ 可得

$$\sum_{r=1}^{L} b_r f_r(\mathbf{z}_1) \overset{a.s.}{=} 0$$

关于 $[0,1]^2$上的勒贝格测度。但这违背了假设，即 $(f_1,\ldots,f_r)$ 线性独立。这个矛盾表明 $A(\boldsymbol{\theta}_0)$ 是正定的，因此验证了引理2.3的假设 (ii)。然后调用该引理即可证明 (b) 部分。

2.2. 定理*1.7* 的证明

(a) 为证明第 (a) 部分，我们将运用引理2.4 ，其

$$\mathcal{S}(\pi,\boldsymbol{\theta}) = \frac{1}{n}\Big[Z_n(\boldsymbol{\theta}) - \boldsymbol{\theta}^\top \mathbf{T}(\pi)\Big].$$

在这种情况下， $\mathcal{S}(\pi,.)$ 是凸函数，因为其海森矩阵等于 $\frac{1}{n}Var_{n,\boldsymbol{\theta}}(\mathbf{T}(\pi))$。现在我们断言，对于所有足够大的 n， $Var_{n,\boldsymbol{\theta}}(\mathbf{T}(\pi))$ 都不是奇异的。反证法：假设对于某个 n， $Var_{n,\boldsymbol{\theta}}(\mathbf{T}(\pi))$ 是奇异的。由于两个概率律 $\mathbb{P}_{n,\boldsymbol{\theta}}$ 和 $\mathbb{P}_{n,\mathbf{0}}$ 彼此绝对连续，因此 $Var_{n,\mathbf{0}}(\mathbf{T}(\pi))$ 也同样是奇异的。但我们在 (b) 部分直接论证了（见 (18)）：

$$\frac{1}{n}Var_{n,\mathbf{0}}(\mathbf{T}(\pi)) \to \Gamma,$$

其中 Γ 是正定的。因此，当 n 足够大时，$Var_{n,\boldsymbol{\theta}}(\mathbf{T}(\pi))$ 不可能是奇异的，因此当 n 足够大时，$\mathcal{S}(\pi,.)$ 是严格凸的。

为验证条件 (i)，注意到

$$\begin{aligned}\mathcal{S}(\pi,\boldsymbol{\theta}_1)-\mathcal{S}(\pi,\boldsymbol{\theta}_0) =& \frac{1}{n}\Big[Z_n(\boldsymbol{\theta}_1)-Z_n(\boldsymbol{\theta}_0)\Big]-\frac{1}{n}\Big[\boldsymbol{\theta}_1-\boldsymbol{\theta}_0\Big]^\top\mathbf{T}(\pi)\\ \xrightarrow{P} & Z(\boldsymbol{\theta}_1)-Z(\boldsymbol{\theta}_0)-(\boldsymbol{\theta}_1-\boldsymbol{\theta}_0)^\top\mathbf{z}(\boldsymbol{\theta}) \qquad (16)\end{aligned}$$

其中我们使用了命题 1.3 的 (a) 和 (c) 两部分。不过，(16) 右端由命题1.3 的 (f) 部分可知为正，因此 (i) 成立。

接下来证明 (ii)，我们有

$$\begin{aligned}\Big|\mathcal{S}(\pi,\boldsymbol{\theta}_1)-\mathcal{S}(\pi,\boldsymbol{\theta}_2)\Big| \le & \frac{1}{n}\Big|Z_n(\boldsymbol{\theta}_1)-Z_n(\boldsymbol{\theta}_2)\Big|+\frac{1}{n}|(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^\top\mathbf{T}(\pi)|\\ \le & 2\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_2K\sqrt{L},\end{aligned}$$

其中 $K:=$最大值$_{r\in[L]}\|f_r\|_\infty$。因此，(ii) 成立，其中 $M(\pi)=2K\sqrt{L}$ 是一个与 π 无关的常数（因此显然 $O_P(1)$）。由此，我们验证了引理 2.4 的所有条件。注意到 $\boldsymbol{\theta}\mapsto\mathcal{S}(\pi,\boldsymbol{\theta})=\frac{1}{n}[Z_n(\boldsymbol{\theta})-\boldsymbol{\theta}^\top\mathbf{T}(\pi)]$ 的唯一极小化元就是最大似然估计量，因此最大似然估计量的一致性得证。

(b) 为证明 (b) 部分，任取 $\mathbf{d}=(d_1,\dots,d_L)$，只需证明

$$\begin{aligned}\mathbf{d}^\top\mathbf{T}(\pi)-\mathbf{d}^\top\nabla Z_n(\mathbf{0}) =& \sum_{i=1}^n\mathbf{d}^\top\mathbf{f}\left(\frac{i}{n},\frac{\pi(i)}{n}\right)-\mathbb{E}_{\mathbf{0}}\sum_{i=1}^n\mathbf{d}^\top\mathbf{f}\left(\frac{i}{n},\frac{\pi(i)}{n}\right)\\ \to & N(0,\boldsymbol{d}^\top\Gamma\boldsymbol{d}).\end{aligned}$$

通过设定来定义函数 $G(.,.):[0,1]^2\mapsto\mathbb{R}$

$$G_{\mathbf{d}}(x,y):=\mathbf{d}^\top\mathbf{f}(x,y),$$

使得 $\mathbf{d}^\top\mathbf{T}(\pi)=\sum_{i=1}^n G_{\mathbf{d}}\left(\frac{i}{n},\frac{\pi(i)}{n}\right)$。若 $\boldsymbol{\theta_0}=\mathbf{0}$，则 π 在 S_n 上服从均匀分布。应用 Hoeffding 的组合中心极限定理（[20, Thm3]），令

$$c_n(i,j):=G_{\mathbf{d}}\Big(\frac{i}{n},\frac{j}{n}\Big)-\frac{1}{n}\sum_{k=1}^n\Big[G_{\mathbf{d}}\Big(\frac{i}{n},\frac{k}{n}\Big)+G_{\mathbf{d}}\Big(\frac{k}{n},\frac{j}{n}\Big)\Big]+\frac{1}{n^2}\sum_{k,\ell=1}^n G_{\mathbf{d}}\Big(\frac{k}{n},\frac{\ell}{n}\Big)$$

我们得出结论：

$$\frac{\mathbf{d}^\top\mathbf{T}(\pi)-\mathbf{d}^\top\nabla Z_n(\mathbf{0})}{\sqrt{Var_{\mathbf{0}}(\mathbf{d}^\top\mathbf{T}(\pi))}}\xrightarrow{D}N(0,1), \qquad (17)$$

一旦

$$\frac{n \max_{i,j\in[n]} c_n^2(i,j)}{\sum_{i,j=1}^n c_n^2(i,j)} \to 0.$$

但注意到以下事实，最后一个公式便显而易见：

$$\frac{n^2 \max_{i,j\in[n]} c_n^2(i,j)}{\sum_{i,j=1}^n c_n^2(i,j)} \to \frac{\sup_{x,y\in[0,1]} G_{\mathbf{d}}^2(x,y)}{\int_{[0,1]^2} G_{\mathbf{d}}^2(x,y)dxdy} < \infty.$$

为完成 (b) 部分的证明，利用 (17)，只需证明：

$$\frac{1}{n} Var_{n,\mathbf{0}}\Big(\mathbf{d}^\top \mathbf{T}(\pi)\Big) \to \boldsymbol{d}^\top \Gamma \boldsymbol{d}. \tag{18}$$

但利用 [20, (Eq 10)] 可注意到以下事实，从而得到这一结论：

$$\frac{1}{n} Var_{n,\mathbf{0}}\Big(\mathbf{d}^\top \mathbf{T}(\pi)\Big) = \frac{1}{n} Var_{n,\mathbf{0}}\Big(\sum_{i=1}^n G_{\mathbf{d}}\Big(\frac{i}{n}, \frac{\pi(i)}{n}\Big)\Big) = \frac{1}{n(n-1)} \sum_{i,j=1}^n c_n^2(i,j),$$

它收敛于

$$\sum_{p,q=1}^L d_p d_q \int_{[0,1]^2} f_p(x,y) f_q(x,y) dxdy = \boldsymbol{d}^\top \Gamma \boldsymbol{d},$$

正如所需。

(c) 部分的证明将通过引理 2.3 并作如下选择来完成：

$$\mathcal{S}(\pi, \boldsymbol{\theta}) = Z_n(\boldsymbol{\theta}) - \boldsymbol{\theta}^\top \mathbf{T}(\pi), \quad B_1 = \Gamma, \quad B_2 = \Gamma.$$

在这种情况下，对于 $\mathbf{z}_n(\boldsymbol{\theta}) := \nabla Z_n(\boldsymbol{\theta})$，我们有

$$\nabla \mathcal{S}(\pi, \boldsymbol{\theta}) = \nabla Z_n(\boldsymbol{\theta}) - \mathbf{T}(\pi) = \mathbf{z}_n(\boldsymbol{\theta}) - \mathbf{T}(\pi), \quad H(\pi, \boldsymbol{\theta}) = Var_{\boldsymbol{\theta}}(\mathbf{T}(\pi)),$$

因此 $\mathcal{S}(\pi,.)$ 是半正定的。注意，引理2.3 的假设 (i) 由 (b) 部分可得。此外，由 $\{f_1, \ldots, f_r\}$的线性无关性可立即推出 Γ是正定的。为验证引理 2.3的条件 (ii)，从而完成证明，只需证明对于任意收敛到 $\mathbf{h}$ $\{\mathbf{h}_n\}_{n\geq 1} \in \mathbb{R}^L$ 的序列，我们有

$$\frac{1}{n} Var_{n, \frac{\mathbf{h}_n}{\sqrt{n}}}(\mathbf{T}(\pi)) \to \Gamma. \tag{19}$$

为此，令 $\{\mathbf{u}_n\}_{n\geq 1}$ 为 $\mathbb{R}^L$ 中收敛于$\mathbf{u}$的一个序列。于是由(b)可得

$$\frac{\mathbf{u}_n^\top \mathbf{T}(\pi) - \mathbf{u}_n^\top \mathbf{z}_n(\mathbf{0})}{\sqrt{n}} \xrightarrow{D} N(0, \mathbf{u}^\top \Sigma \mathbf{u}).$$

此外，利用 [7, 命题 3.10] 可得 $\left\{\exp\left(\frac{\mathbf{u}_n^\top \mathbf{T}(\pi) - \mathbf{u}_n^\top \mathbf{z}_n(\mathbf{0})}{\sqrt{n}}\right)\right\}_{n\geq 1}$ $\mathbf{uz}()$是一致可积的，因此利用上面的最后一个显示式可得

$$Z_n\left(\frac{\mathbf{u}_n}{\sqrt{n}}\right) - Z_n(\mathbf{0}) - \frac{1}{\sqrt{n}}{\mathbf{u}_n}^\top \mathbf{z}_n(\mathbf{0}) = \log \mathbb{E}_{n,\mathbf{0}} e^{\frac{\mathbf{u}_n^\top \mathbf{T}(\pi) - \mathbf{u}_n^\top \mathbf{z}_n(\mathbf{0})}{\sqrt{n}}}$$
$$\to \log \mathbb{E} e^{N(0,\mathbf{u}^\top \Sigma \mathbf{u})} = \frac{1}{2}\mathbf{u}^\top \Gamma \mathbf{u}. \quad (20)$$

特别地，对所有 n 取$\mathbf{u}_n = \mathbf{u}$，我们得到

$$Z_n\left(\frac{\mathbf{u}}{\sqrt{n}}\right) - Z_n(\mathbf{0}) - \frac{1}{\sqrt{n}}\mathbf{u}^\top \mathbf{z}_n(\mathbf{0}) \to \frac{1}{2}\mathbf{u}^\top \Gamma \mathbf{u}.$$

由于函数$\mathbf{u} \mapsto Z_n\left(\frac{\mathbf{u}}{\sqrt{n}}\right)$ 是凸的，且上面的右端可微，因此梯度收敛，即

$$\frac{\mathbf{z}_n\left(\frac{\mathbf{u}}{\sqrt{n}}\right) - \mathbf{z}_n(\mathbf{0})}{\sqrt{n}} \to \Gamma \mathbf{u}.$$

此外，由于上面右端的极限函数连续，因此上述收敛是一致的；也就是说，对于任何收敛到$\mathbf{u}$的序列 $\{\mathbf{u}_n\}_{n\geq 1}$，都有

$$\frac{\mathbf{z}_n\left(\frac{\mathbf{u}_n}{\sqrt{n}}\right) - \mathbf{z}_n(\mathbf{0})}{\sqrt{n}} \to \Gamma \mathbf{u}. \quad (21)$$

因此，固定 $\mathbf{d} \in \mathbb{R}^L$ 并在 $\mathbf{u}_n = \mathbf{h}_n + t\mathbf{d}$ 时应用（20），并在$\mathbf{u}_n = \mathbf{h}_n$时应用（21），对于任意 $t \in \mathbb{R}$，我们有

$$\log \mathbb{E}_{n,n^{-1/2}\mathbf{h}_n} \exp\left[\frac{t}{\sqrt{n}}\left(\mathbf{d}^\top \mathbf{T}(\pi) - \mathbf{d}^\top \mathbf{z}_n\left(\frac{\mathbf{h}_n}{\sqrt{n}}\right)\right)\right]$$
$$= Z_n\left(\frac{\mathbf{h}_n + t\mathbf{d}}{\sqrt{n}}\right) - Z_n\left(\frac{\mathbf{h}_n}{\sqrt{n}}\right) - \left(\frac{t}{\sqrt{n}}\right)\mathbf{d}^\top \mathbf{z}_n\left(\frac{\mathbf{h}_n}{\sqrt{n}}\right) \to \frac{t^2}{2}\mathbf{d}^\top \Gamma \mathbf{d}.$$

上面的公式表明，在 $\mathbb{P}_{n,\frac{\mathbf{h}_n}{\sqrt{n}}}$ 条件下，

$$\frac{\mathbf{d}^\top \mathbf{T}(\pi) - \mathbf{d}^\top \mathbf{z}_n\left(\frac{\mathbf{h}_n}{\sqrt{n}}\right)}{\sqrt{n}} \xrightarrow{D} N(0, \mathbf{d}^\top \Gamma \mathbf{d}),$$

而且上述收敛在矩意义下也成立。因此，我们有

$$\frac{1}{n} Var_{n,\frac{\mathbf{h}_n}{\sqrt{n}}}(\mathbf{d}^\top \mathbf{T}(\pi)) \to \mathbf{d}^\top \Gamma \mathbf{d}.$$

由于这对所有 $\mathbf{d} \in \mathbb{R}^L$ 都成立，因此我们得出$\frac{1}{n} Var_{n,\frac{\mathbf{h}_n}{\sqrt{n}}}(\mathbf{T}(\pi)) \to \Gamma$。这验证了（19），从而完成了该定理的证明。

2.3. 证明引理 1.9

我们声称，只需证明

$$\widehat{A}(\hat{\boldsymbol{\theta}}_{PL}) \xrightarrow{P} A(\boldsymbol{\theta}_0), \quad \widehat{\Sigma}(\hat{\boldsymbol{\theta}}_{PL}) \xrightarrow{P} \Sigma(\boldsymbol{\theta}_0), \tag{22}$$

我们首先完成该引理的证明，证明该断言的部分将推迟进行。利用定理1.5 （b）部分，我们得到

$$\frac{\sqrt{n}\mathbf{d}^\top(\hat{\boldsymbol{\theta}}_{PL} - \boldsymbol{\theta}_0)}{\sqrt{\mathbf{d}^\top[A(\boldsymbol{\theta}_0)]^{-1}\Sigma(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1}\mathbf{d}}} \xrightarrow{D} N\Big(0,1\Big).$$

此外，（22）给出

$$\mathbf{d}^\top[\widehat{A}(\hat{\boldsymbol{\theta}}_{PL})]^{-1}\widehat{\Sigma}(\hat{\boldsymbol{\theta}}_{PL})[\widehat{A}(\hat{\boldsymbol{\theta}}_{PL})]^{-1}\mathbf{d} \xrightarrow{P} \mathbf{d}^\top[A(\boldsymbol{\theta}_0)]^{-1}\Sigma(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1}\mathbf{d}.$$

结合上述两个公式以及斯勒茨基定理可得

$$\frac{\sqrt{n}\mathbf{d}^\top(\hat{\boldsymbol{\theta}}_{PL} - \boldsymbol{\theta}_0)}{\sqrt{\mathbf{d}^\top\Big[[\widehat{A}(\hat{\boldsymbol{\theta}}_{PL})]^{-1}\widehat{\Sigma}(\hat{\boldsymbol{\theta}}_{PL})[\widehat{A}(\hat{\boldsymbol{\theta}}_{PL})]^{-1}\Big]\mathbf{d}}} \xrightarrow{D} N(0,1),$$

由此即可直接构造置信区间。

要完成证明，只需验证（22）。为此，令 $M := \max_{1\le r\le L}\|f_r\|_\infty < \infty$ 我们有$\max_{i,\ j\in[n],\ \pi\in S_n,\ r\in[L]}|y_{\pi,\ r}(i,\ j)| \le 4M$。由于 $\hat{\boldsymbol{\theta}}_{PL} \xrightarrow{P} \boldsymbol{\theta}_0$ （根据定理 1.5 第 (b) 部分），由所涉及函数在紧致域上的一致连续性可得

$$\widehat{A}(\hat{\boldsymbol{\theta}}_{PL}) - \widehat{A}(\boldsymbol{\theta}_0) \xrightarrow{P} 0, \quad \widehat{\Sigma}(\hat{\boldsymbol{\theta}}_{PL}) - \widehat{\Sigma}(\boldsymbol{\theta}_0) \xrightarrow{P} 0.$$

因此，只需证明 Danmarkimi

$$\widehat{A}(\boldsymbol{\theta}_0) \xrightarrow{P} A(\boldsymbol{\theta}_0), \quad \widehat{\Sigma}(\boldsymbol{\theta}_0) \xrightarrow{P} \Sigma(\boldsymbol{\theta}_0).$$

其中，第一个结论由使用（14）得到。对于第二个结论，使用（13）可注意到

$$\mathbb{E}_{n,\boldsymbol{\theta}_0}\widehat{\Sigma}_{pq}(\boldsymbol{\theta}_0) \to \Sigma_{pq}(\boldsymbol{\theta}_0).$$

要完成证明，只需证明 that

$$Var_{n,\boldsymbol{\theta}_0}(\widehat{\Sigma}_{pq}(\boldsymbol{\theta}_0)) \to 0.$$

但利用（10）可知，对于任意有限个指标$\{p_1,\dots,p_\ell\}$，随机变量 $\{\pi(p_1)\dots,\pi(p_\ell)\}$ 近似独立。

3. 辅助引理的证明

3.1. 命题的证明 1.3

(a)这可通过调用 [28, 定理 1.5 (a)]，使用 $f = \boldsymbol{\theta}^\top \mathbf{f}$ 和 $\theta = 1$得到。
(b) (a) 部分中的优化问题的唯一性由 [28, 定理 1.5 (b)] 得出。密度 $\rho_{\boldsymbol{\theta}}$ 处处严格为正这一事实，由 [28,定理 1.5 (c)] 中最优密度的形式得出。

(c) 经验测度 ν_π 的收敛性由 [28, 定理 1.5 (b)]得出。第二个结论通过注意到

$$\frac{1}{n}\mathbb{E}_{n,\boldsymbol{\theta}}[\mathbf{T}(\pi)] = \mathbb{E}_{n,\boldsymbol{\theta}}[\nu_\pi(\mathbf{f})] \xrightarrow{P} \mu_\theta(\mathbf{f}) = \mathbf{z}(\boldsymbol{\theta}),$$

其中我们使用控制收敛定理。
(d) 令 $\{\boldsymbol{\theta}_k\}_{k\geq 1}$ 为 $\mathbb{R}^L$ 中收敛于 $\boldsymbol{\theta}_\infty$ 的一个序列。我们声称

$$\mu_{\boldsymbol{\theta}_k} \xrightarrow{w} \mu_{\boldsymbol{\theta}_\infty}. \tag{23}$$

由此可得

$$\mathbf{z}(\boldsymbol{\theta}_k) = \mu_{\boldsymbol{\theta}_k}(\mathbf{f}) \to \mu_{\boldsymbol{\theta}_\infty}(\mathbf{f}) = \mathbf{z}(\boldsymbol{\theta}_\infty),$$

从而证明连续性。继续证明（23），令 μ_∞ 为测度序列 $\{\mu_{\boldsymbol{\theta}_k}\}_{k\geq 1}$的任意极限点，该极限点由$\mathcal{M}$的紧性保证存在。利用 Z（。）的连续性以及 D（.||u）的下半连续性，我们有

$$Z(\boldsymbol{\theta}_\infty) = \lim_{k\to\infty} Z(\boldsymbol{\theta}_k) = \lim_{k\to\infty} \left\{\boldsymbol{\theta}_k \mu_{\boldsymbol{\theta}_k}(\mathbf{f}) - D(\mu_{\boldsymbol{\theta}_k}||u)\right\} \leq \boldsymbol{\theta}_\infty \mu(\mathbf{f}) - D(\mu_\infty||u).$$

但由部分 (a) 和 (b) 可知

$$Z(\boldsymbol{\theta}_\infty) = \sup_{\mu\in\mathcal{M}} \left\{\boldsymbol{\theta}_\infty \mu(\mathbf{f}) - D(\mu||u)\right\} = \boldsymbol{\theta}_\infty \mu_{\boldsymbol{\theta}_\infty}(\mathbf{f}) - D(\mu_{\boldsymbol{\theta}_\infty}||u).$$

比较最后两个展示式，并结合 (b) 部分优化器的唯一性，可得 $\mu_{\boldsymbol{\theta}_\infty} = \mu_\infty$。因此我们证明了（23），从而$\mathbf{z}$（。）的连续性得证。

(e) 利用(c)部分我们得到

$$\frac{\nabla Z_n(\boldsymbol{\theta})}{n} = \frac{1}{n}\mathbb{E}_{n,\boldsymbol{\theta}}[\mathbf{T}(\pi)] \xrightarrow{P} \mathbf{z}(\boldsymbol{\theta}).$$

由于函数 Z(.) 是一列凸函数的极限，因此它是凸函数，进而几乎处处可微。利用 (a) 部分以及上面的公式，我们得到

$$\nabla Z(\boldsymbol{\theta}) \overset{a.s.}{=} \mathbf{z}(\boldsymbol{\theta}).$$

最后，由于 $\mathbf{z}$(.) 是连续的（由上面的(d)部分可知），因此 Z(.) 处处可微，并且上述等式处处成立。

(f) 为证明严格凸性，反设存在$\theta_0 \neq \theta_1$ 使得

$$Z(\boldsymbol{\theta}_1) = Z(\boldsymbol{\theta}_0) + (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^\top \mathbf{z}(\boldsymbol{\theta}_0). \tag{24}$$

定义函数 $\phi:\ [0,1] \mapsto \mathbb{R}$ 通过设置

$$\phi(t) := Z\Big((1-t)\boldsymbol{\theta}_0 + t\boldsymbol{\theta}_1\Big) = Z\Big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)\Big).$$

那么 ϕ 是凸的，并且

$$\phi(0) = Z(\boldsymbol{\theta}_0), \quad \phi(1) = Z(\boldsymbol{\theta}_1), \quad \phi'(t) = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^\top \mathbf{z}\Big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)\Big).$$

因此，(24) 转化为 $\phi(1) = \phi(0) + \phi'(0)$。凸性表明，$\phi$ 于是 $\phi(t) = \phi(0) + t\phi'(0)$ 对所有 $t \in [0,1]$均成立。求导可得，$\phi'(t) = \phi'(0)$ 对所有 $t \in [0,1]$均成立，因此

$$(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^\top \mathbf{z}(\boldsymbol{\theta}_1) = \phi'(1) = \phi'(0) = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^\top \mathbf{z}(\boldsymbol{\theta}_0). \tag{25}$$

现在，$\mu_{\boldsymbol{\theta}_1}$ 的最优性给出

$$\boldsymbol{\theta}_1^\top \mu_{\boldsymbol{\theta}_1}(\mathbf{f}) - D(\mu_{\boldsymbol{\theta}_1}||u) \geq \boldsymbol{\theta}_1^\top \mu_{\boldsymbol{\theta}_0}(\mathbf{f}) - D(\mu_{\boldsymbol{\theta}_0}||u),$$

这意味着

$$\boldsymbol{\theta}_1^\top \Big(\mathbf{z}(\boldsymbol{\theta}_1) - \mathbf{z}(\boldsymbol{\theta}_0)\Big) \geq D(\mu_{\boldsymbol{\theta}_1}||u) - D(\mu_{\boldsymbol{\theta}_0}||u). \tag{26}$$

同样，$\mu_{\boldsymbol{\theta}_0}$ 的最优性给出。

$$\boldsymbol{\theta}_0^\top \Big(\mathbf{z}(\boldsymbol{\theta}_1) - \mathbf{z}(\boldsymbol{\theta}_0)\Big) \leq D(\mu_{\boldsymbol{\theta}_1}||u) - D(\mu_{\boldsymbol{\theta}_0}||u). \tag{27}$$

结合(26)和(27)得到

$$\boldsymbol{\theta}_1^\top \Big(\mathbf{z}(\boldsymbol{\theta}_1) - \mathbf{z}(\boldsymbol{\theta}_0)\Big) \geq D(\mu_{\boldsymbol{\theta}_1}||u) - D(\mu_{\boldsymbol{\theta}_0}||u) \geq \boldsymbol{\theta}_0^\top \Big(\mathbf{z}(\boldsymbol{\theta}_1) - \mathbf{z}(\boldsymbol{\theta}_0)\Big).$$

但根据（25），上述公式中的左端与右端相同，因此

$$\boldsymbol{\theta}_1^\top \Big(\mathbf{z}(\boldsymbol{\theta}_1) - \mathbf{z}(\boldsymbol{\theta}_0)\Big) = D(\mu_{\boldsymbol{\theta}_1}||u) - D(\mu_{\boldsymbol{\theta}_0}||u) = \boldsymbol{\theta}_0^\top \Big(\mathbf{z}(\boldsymbol{\theta}_1) - \mathbf{z}(\boldsymbol{\theta}_0)\Big).$$

这意味着等式必须在(26)和(27)中都成立，因此优化器的唯一性（由(b)部分得出）意味着 $\mu_{\boldsymbol{\theta}_1} = \mu_{\boldsymbol{\theta}_0}$。结合 [28, 定理 1.5 (c)] 中的密度形式可得

$$\begin{aligned}\rho_{\boldsymbol{\theta}_1}(x,y) &= \exp\Big(\boldsymbol{\theta}_1^\top \mathbf{f}(x,y) + a_{\boldsymbol{\theta}_1}(x) + b_{\boldsymbol{\theta}_1}(y)\Big) \\ &= \exp\Big(\boldsymbol{\theta}_0^\top \mathbf{f}(x,y) + a_{\boldsymbol{\theta}_0}(x) + b_{\boldsymbol{\theta}_0}(y)\Big) = \rho_{\boldsymbol{\theta}_0}(x,y),\end{aligned}$$

对于某些函数 $a_{\boldsymbol{\theta}}$, $b_{\boldsymbol{\theta}}$ 从 $[0,1]$ 到 $\mathbb{R}$。取对数并化简得

$$(\boldsymbol{\theta}_1-\boldsymbol{\theta}_0)^\top \mathbf{f}(x,y)=a_{\boldsymbol{\theta}_0}(x)+b_{\boldsymbol{\theta}_0}(y)-a_{\boldsymbol{\theta}_1}(x)-b_{\boldsymbol{\theta}_1}(y). \tag{28}$$

对 $y\in[0,1]$ 积分，并利用对所有 $r\in[L]$ 都有 $f_r\in\mathcal{C}$ 这一事实，我们得到

$$a_{\boldsymbol{\theta}_1}(x)-a_{\boldsymbol{\theta}_0}(x)=\int_0^1 [b_{\boldsymbol{\theta}_0}(y)-b_{\boldsymbol{\theta}_1}(y)]dy.$$

因此，函数 $a_{\boldsymbol{\theta}_1}(.)-a_{\boldsymbol{\theta}_0}(.)$ 几乎处处为常数。对 x 积分进行类似计算可得 $b_{\boldsymbol{\theta}_1}(.)-b_{\boldsymbol{\theta}_0}(.)$ 几乎处处为常数，因此利用 (28) 可知 $(\boldsymbol{\theta}_1-\boldsymbol{\theta}_0)^\top\mathbf{f}(x,y)$ 几乎处处为常数。但假设$f_r\in\mathcal{C}$ 随即迫使该常数为 0，这与 $\mathbf{f}$ 线性独立这一假设相矛盾。因此 (24) 不成立，从而$Z(.)$ 严格凸，证毕。

*3.2.*引理 的证明 *2.1*

利用给定 $\pi(i)$ $\pi(j)$) 的条件分布 $\{\pi(\ell)\ \ell\neq i,\ j\}$ 如（5）所示，我们得到

$$\begin{aligned}
&\mathbb{E}_{n,\boldsymbol{\theta}}\Big(\mathbf{C}_{ij}(\pi)\Big|\pi(\ell)=\sigma(\ell),\ell\neq i,j\Big)\\
=&\frac{e^{\boldsymbol{\theta}^\top\mathbf{y}_\sigma(i,j)}}{1+e^{\boldsymbol{\theta}^\top\mathbf{y}_\sigma(i,j)}}\frac{1}{1+e^{\boldsymbol{\theta}^\top\mathbf{y}_\sigma(i,j)}}\mathbf{y}_\sigma(i,j)+\frac{1}{1+e^{\boldsymbol{\theta}^\top\mathbf{y}_\sigma(i,j)}}\frac{1}{1+e^{-\boldsymbol{\theta}^\top\mathbf{y}_\sigma(i,j)}}[-\mathbf{y}_\sigma(i,j)]\\
=&\frac{e^{\boldsymbol{\theta}^\top\mathbf{y}_\sigma(i,j)}}{\Big(1+e^{\boldsymbol{\theta}^\top\mathbf{y}_\sigma(i,j)}\Big)^2}\Big[\mathbf{y}_\sigma(i,j)-\mathbf{y}_\sigma(i,j)\Big]=\mathbf{0}.
\end{aligned}$$

注意到 $C_{ij}(\pi,\ \mathbf{d})\ =\mathbf{d}^\top\mathbf{C}ij(\pi)$，即可得到所需结论。

3.3. 证明 引理 *2.2*

首先，直接展开可得

$$\begin{aligned}
&\Big|\mathbb{E}_{n,\boldsymbol{\theta}_0}\Big(\sum_{(i,j)\in\mathcal{E}_n}C_{ij}(\pi,\mathbf{d})\Big)^k-\mathbb{E}D_n^k\Big|\\
&\leq\sum_{(i_1,j_1),\ldots,(i_k,j_k)\in\mathcal{E}_n}\Big|\mathbb{E}_{n,\boldsymbol{\theta}_0}\Big[\prod_{a=1}^k C_{i_aj_a}(\pi)\Big]-\mathbb{E}\Big[\prod_{a=1}^k D_{i_aj_a}\Big]\Big|.
\end{aligned}\tag{29}$$

给定 k 对 $(i_1,j_1)\ldots,(i_k,j_k)$ in $\mathcal{E}_n$，令 $H(\mathbf{i},\mathbf{j})$ 表示由边$(i_1,j_1)\ldots,(i_k,j_k)$ 形成的带标记多重图。图 $H(\mathbf{i},\mathbf{j})$ 含有多重边，但没有环。根据图 $H(\mathbf{i},\mathbf{j})$ 的不同，我们将论证分为以下两种情况:

• **情况 1：** $H = H(\mathbf{i}, \mathbf{j})$ 有一条孤立边，即它有一条边(i_1, j_1)，该边不与其他任何边 $\{(i_a, j_a)\ 2 \le a \le k\}$ 相交。

在这种情况下，我们将证明

$$\mathbb{E}_{n,\boldsymbol{\theta}_0}\Big[\prod_{a=1}^{k} C_{i_a j_a}(\pi)\Big] = 0, \quad \mathbb{E}\Big[\prod_{a=1}^{k} D_{i_a j_a}\Big] = 0. \tag{30}$$

确实，调用引理 2.1 可得

$$\mathbb{E}_{n,\boldsymbol{\theta}_0}\Big[\prod_{a=1}^{k} C_{i_a j_a}(\pi)\Big] = \mathbb{E}_{n,\boldsymbol{\theta}_0}\Big[\mathbb{E}_{n,\boldsymbol{\theta}_0}\Big(C_{i_1 j_1}(\pi)\Big|\pi(\ell), \ell \neq i_1, j_1\Big)\prod_{a=2}^{k} C_{i_a j_a}(\pi)\Big] = 0,$$

从而验证了（中的第一个等式30）。接下来证明（30），再次使用引理 2.1 以及 $\{D_{ij}, (i,\ j)\ \in \mathcal{E}_n\}$ 的构造可得

$$\mathbb{E}\Big[D_{i_1 j_1} D_{\ell r}\Big] = \mathbb{E}_{n,\boldsymbol{\theta}_0}\Big[C_{i_1 j_1}(\pi) C_{\ell r}(\pi)\Big] = 0 \text{ whenever } (i_1, j_1) \cap (\ell, r) = \phi.$$

由于 $\{D_{ij}, (i,j)\ \in \mathcal{E}_n\}$ 是多元高斯，我们得出 $D_{i_1 j_1}$独立于 $\{D_{\ell r}, (\ell, r)\ \cap (i_1, j_1)\ = \phi\}$。因此，我们有

$$\mathbb{E}\Big(D_{i_1 j_1} | D_{\ell r}, (\ell, r) \cap (i_1, j_1) = \phi\Big) = \mathbb{E} D_{i_1 j_1} = 0,$$

这给出的是

$$\mathbb{E}\Big[\prod_{a=1}^{k} D_{i_a j_a}\Big] = \mathbb{E}\Big[\mathbb{E}\Big(D_{i_1 j_1}\Big|D_{\ell r}, (\ell, r) \cap (i_1, j_1) = \phi\Big)\prod_{a=2}^{k} D_{i_a j_a}\Big] = 0.$$

这验证了（30）中的第二个等式，从而完成了情况 1 的证明。

• **情况 2:** $H = H(\mathbf{i},\mathbf{j})$ 有超过 $\binom{3k}{2}$ 个顶点。

在这种情况下，我们将证明存在一条孤立边。因此，利用上述情况 1，我们将得到

$$\mathbb{E}_{n,\boldsymbol{\theta}_0}\Big[\prod_{a=1}^{k} C_{i_a j_a}(\pi)\Big] = 0, \quad \mathbb{E}\Big[\prod_{a=1}^{k} D_{i_a j_a}\Big] = 0.$$

接下来寻找一条孤立边，令 $\cup_{b=1}^{s} H_b$ 表示图 H 分解为 s 个连通分量的过程。反设不存在孤立边。那么每个分量 H_b 都有$|V(H_b)| \ge 3$，$|E(H_b)| \ge 2$，由此得到

$$k = |E(H)| = \sum_{b=1}^{s} |E(H_b)| \ge 2s. \tag{31}$$

此外，由于每个 H_b 都是连通的，我们有 $|E(H_b)| \geq |V(H_b)|-1$。对 b求和可得

$$k = |E(H)| \geq |V(H)| - s > \frac{3k}{2} - s. \tag{32}$$

结合（31）和（32）可得 $s \leq \frac{k}{2}$ 以及 $s > \frac{k}{2}$，这产生了矛盾。情况 2 的证明至此完成。

- **情况 3：** $H(\mathbf{i},\mathbf{j})$ 恰好有 $\frac{3k}{2}$ 个顶点，且 H 中的边均不是孤立边。

  在这种情况下，我们将证明 H 是两个星的无交并，即 H 的每个分量都是一个具有两条边和三个顶点的连通图。因此，调用（10）可得

$$\begin{aligned}
&\mathbb{E}_{n,\boldsymbol{\theta}_0}\Big[\prod_{a=1}^{k} C_{i_a j_a}(\pi)\Big] \\
=&\prod_{b=1}^{\frac{k}{2}} \mathbb{E}_{n,\boldsymbol{\theta}_0}\Big[\prod_{(i_a,j_a)\in E(H_b)} C_{i_a j_a}(\pi)\Big] + o(1) \\
=&\prod_{b=1}^{\frac{k}{2}} \mathbb{E}\Big[\prod_{(i_a,j_a)\in E(H_b)} D_{i_a j_a}\Big] + o(1) \text{ [Since } H_b \text{ has exactly two edges]} \\
=&\mathbb{E}\Big[\prod_{a=1}^{k} D_{i_a j_a}\Big] + o(1) \text{ [By independence of uncorrelated Gaussians].}
\end{aligned}$$

  接下来验证 H 确实是两个星的无交并。首先注意到，由于 H 有 $\frac{3k}{2}$ 个顶点，k 必为偶数。此外，与情况 2 中相同的证明表明，H 的连通分量数必须恰好为 $s = \frac{k}{2}$。另外，H 没有孤立边这一假设给出$|V(H_b)| \geq 3$, $|E(H_b)| \geq 2$。于是得到

$$\frac{3k}{2} = |V(H)| = \sum_{b=1}^{\frac{k}{2}} |V(H_b)| \geq \frac{3k}{2}, \quad k = |E(H)| = \sum_{b=1}^{\frac{k}{2}} |E(H_b)| \geq k,$$

  因此，H 的每个分量恰好有 3 个顶点和 2 条边，正如所需。情况 3 的证明至此完成。

综合以上三种情况，并注意到选择$\mathbf{i}$，$\mathbf{j}$ 使得 H（$\mathbf{i}$，$\mathbf{j}$）具有 r 个顶点的方式数为 O（n^r）、29）可得

$$\Big|\mathbb{E}_{n,\boldsymbol{\theta}_0}\Big(\sum_{(i,j)\in\mathcal{E}_n} C_{ij}(\pi,\mathbf{d})\Big)^k - \mathbb{E}D_n^k\Big| \leq \sum_{1\leq r<\frac{3k}{2}} O(n^r) + o(n^{\frac{3k}{2}})1\{k \text{ is even}\},$$

由此可得引理 2.2 的结论。

3.4. 引理的证明 2.3

首先，注意到函数 $\mathcal{S}(\pi,.)$ 是凸的。在整个证明过程中，对于任意 $\delta > 0$ 和 $\boldsymbol{\theta}_1 \in \mathbb{R}^L$ 集合

$$B(\boldsymbol{\theta}_1, \delta) := \{\boldsymbol{\theta} \in \mathbb{R}^L : \|\boldsymbol{\theta} - \boldsymbol{\theta}_1\|_2 \leq \delta\}, \quad \partial B(\boldsymbol{\theta}_1, \delta) := \{\boldsymbol{\theta} \in \mathbb{R}^L : \|\boldsymbol{\theta} - \boldsymbol{\theta}_1\|_2 = \delta\}.$$

现在我们将证明分解为以下步骤:

- 对于每个 $K > 0$ ，我们都有

$$\lim_{n\to\infty} \mathbb{P}_{n,\boldsymbol{\theta}_0}\left(\text{There exists a unique minimizer in } B\Big(\boldsymbol{\theta}_0, \frac{K}{\sqrt{n}}\Big)\right) = 1.$$

证明。固定 $K > 0$，函数 $\mathcal{S}(\pi,.)$是连续的，因此在紧集 $B\Big(\boldsymbol{\theta}_0, \frac{K}{\sqrt{n}}\Big)$上取得其最小值。现在我们论证极小化元的唯一性。事实上，如果$\theta_{n,K},1$ 和 $\theta_{n,K},2$是 $\mathcal{S}(\pi, .)$ 在 $B\Big(\boldsymbol{\theta}_0, \frac{K}{\sqrt{n}}\Big)$中的两个极小化元，那么利用 $\mathcal{S}(\pi, .)$ 的凸性，它必在线段$\theta_{n,K},1$和$\theta_{n,K},2$ 上为常数，因此梯度必在线段上处处为零。于是存在随机变量 $\boldsymbol{\xi}_{n,K} \in B\Big(\boldsymbol{\theta}_0, \frac{K}{\sqrt{n}}\Big)$ ，使得矩阵 $H(\pi,\boldsymbol{\xi}_{n,K})$是奇异的。但由假设 (ii) 可知 $n^{-1}H(\pi,\boldsymbol{\xi}_{n,K}) \xrightarrow{P} B_2$ ，其中 B_2 非奇异，这就产生了矛盾。因此，以 1 为极限的概率下，在集合 $B\Big(\boldsymbol{\theta}_0, \frac{K}{\sqrt{n}}\Big)$ 上存在唯一的极小化元 $\theta_{n,K}$ 。

□

- $$\limsup_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}_{n,\boldsymbol{\theta}_0}\Big(\mathcal{S}(\pi, \boldsymbol{\theta}_0) \geq \inf_{\boldsymbol{\theta}\in\partial B\left(\boldsymbol{\theta}_0, \frac{K}{\sqrt{n}}\right)} \mathcal{S}(\pi, \boldsymbol{\theta})\Big) = 0.$$

证明。 令$\boldsymbol{\theta}_{n,K}$ 是 $\mathcal{S}(\pi,.)$ 在紧集$\partial B\Big(\boldsymbol{\theta}_0, \frac{K}{\sqrt{n}}\Big)$上的全局极小值点。令 $\mathbf{u}_{n,K} := \frac{\sqrt{n}}{K}(\boldsymbol{\theta}_{n,K} - \boldsymbol{\theta}_0)$，有 $\|\mathbf{u}_{n,K}\| = 1$。二阶泰勒展开给出

$$\begin{aligned}
\mathcal{S}(\pi, \boldsymbol{\theta}_{n,K}) - \mathcal{S}(\pi, \boldsymbol{\theta}_0) =& (\boldsymbol{\theta}_{n,K} - \boldsymbol{\theta}_0)^\top \nabla\mathcal{S}(\pi, \boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}_{n,K} - \boldsymbol{\theta}_0)^\top H(\pi, \boldsymbol{\xi}_{n,K})(\boldsymbol{\theta}_{n,K} - \boldsymbol{\theta}_0) \\
=& \frac{K}{\sqrt{n}}\mathbf{u}_{n,K}^\top \nabla\mathcal{S}(\pi, \boldsymbol{\theta}_0) + \frac{K^2}{2n}\mathbf{u}_{n,K}^\top H(\pi, \boldsymbol{\xi}_{n,K})\mathbf{u}_{n,K} \\
\geq& -K\frac{\|\nabla\mathcal{S}(\pi, \boldsymbol{\theta}_0)\|_2}{\sqrt{n}} + \frac{K^2}{2n}\lambda_1(H(\pi, \boldsymbol{\xi}_{n,K})).
\end{aligned}$$

这给出

$$\begin{aligned}
&\mathbb{P}_{n,\boldsymbol{\theta}_0}\Big(\mathcal{S}(\pi,\boldsymbol{\theta}_0) \geq \inf_{\boldsymbol{\theta}\in\partial B\left(\boldsymbol{\theta}_0,\frac{K}{\sqrt{n}}\right)} \mathcal{S}(\pi,\boldsymbol{\theta})\Big)\\
=&\mathbb{P}_{n,\boldsymbol{\theta}_0}\Big(\mathcal{S}(\pi,\boldsymbol{\theta}_0) \geq \mathcal{S}(\pi,\boldsymbol{\theta}_{n,K})\Big)\\
\leq&\mathbb{P}_{n,\boldsymbol{\theta}_0}\left(\frac{\|\nabla\mathcal{S}(\pi,\boldsymbol{\theta}_0)\|_2}{\sqrt{n}} \geq \frac{K}{2n}\lambda_1\Big(H(\pi,\boldsymbol{\xi}_{n,K})\Big)\right), \qquad (33)
\end{aligned}$$

其中，λ_1 表示矩阵的最小特征值。利用 (ii)，我们有

$$n^{-1}H(\pi,\boldsymbol{\xi}_{n,K}) \xrightarrow{P} B_2 \;\Rightarrow\; n^{-1}\lambda_1(H(\pi,\boldsymbol{\xi}_{n,K})) \xrightarrow{P} \lambda_1(B_2) > 0.$$

另一方面，利用(i) 我们有 $\|\nabla\mathcal{S}(\pi,\ \boldsymbol{\theta}_0)\|_2 = O_P(\sqrt{n})$。所需的结论由这两个观察以及（33）得出。

□

- 极限$_{n\to\infty}\mathbb{P}_{n,\ \boldsymbol{\theta}_0}$（在 $\mathbb{R}^L$ 中存在唯一全局极小值点 $\hat{\ }\boldsymbol{\theta}_n$ （记作））$= 1$. 此外，$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = O_P(1)$。

证明。利用前一步，存在一个发散到 ∞ 的正实数序列$\{K_n\}_{n\geq 1}$，使得以下结论成立:

$$\begin{aligned}
&\lim_{n\to\infty}\mathbb{P}_{n,\boldsymbol{\theta}_0}\left(\text{There exists a unique minimizer in } B\Big(\boldsymbol{\theta}_0,\frac{K_n}{\sqrt{n}}\Big)\right) = 1,\\
&\lim_{n\to\infty}\mathbb{P}_{n,\boldsymbol{\theta}_0}\Big(\mathcal{S}(\pi,\boldsymbol{\theta}_0) < \inf_{\boldsymbol{\theta}\in\partial B\left(\boldsymbol{\theta}_0,\frac{K_n}{\sqrt{n}}\right)} \mathcal{S}(\pi,\boldsymbol{\theta})\Big) = 1.
\end{aligned}$$

不失一般性，我们在上述两个集合的交集上进行讨论。令$\hat{\ }\boldsymbol{\theta}_n$ 为 $\mathcal{S}(\pi,.)$ 在 $B\Big(\boldsymbol{\theta}_0,\frac{K_n}{\sqrt{n}}\Big)$ 上的唯一全局极小值点。于是，$\hat{\ }\boldsymbol{\theta}_n$ 是 $B\Big(\boldsymbol{\theta}_0,\frac{K_n}{\sqrt{n}}\Big)$ 的内部点，因此 $\nabla\mathcal{S}(\pi,\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$。由于$\hat{\mathcal{S}}(\pi,.)$是凸的，$\boldsymbol{\theta}_n$ 是全局极小值点。接下来只需证明不存在其他全局极小值点。但利用凸性可知，若存在其他全局极小值点，则必将在集合 $B\Big(\boldsymbol{\theta}_0,\frac{K}{\sqrt{n}}\Big)$ 中存在其他全局极小值点，这就产生了矛盾。

为验证第二个结论，令 $\{\alpha_n\}_{n\geq 1}$ 为一个趋于 ∞ 的正实数序列。注意到，将上述证明中的 K_n 替换为 $K'_n := \min(K_n, \sqrt{\alpha_n})$ 后，证明过程完全相同。因此，重复上述证明可得 $\sqrt{n}\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\| < K'_n = o(\alpha_n)$。由于这对任意趋于 ∞ 的正实数序列 $\hat{\ }\{\alpha_n\}_{n\geq 1}$ 都成立，因此可得 $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = O_P(1)$，证毕。

□

- $\hat{\boldsymbol{\theta}}_n$ 是 $\nabla\mathcal{S}(\pi,\boldsymbol{\theta}) = \mathbf{0}$的唯一解。

  ˆ证明。 由于 $\boldsymbol{\theta}_n$ 是一个二次可微函数的全局极小值点，ˆ 因此 $\nabla\mathcal{S}(\pi,\boldsymbol{\theta}_n) = \mathbf{0}$。此外，注意到 ˆ$\mathcal{S}(\pi,.)$ 是凸的，且 $\boldsymbol{\theta}_n$ 是唯一的全局极小值点，即可得唯一性。 □

- $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, B_2^{-1}B_1B_2^{-1})$.

  证明。 泰勒级数展开得到

  $$\mathbf{0} = \nabla\mathcal{S}(\pi,\hat{\boldsymbol{\theta}}_n) = \nabla\mathcal{S}(\pi,\boldsymbol{\theta}_0) + \int_0^1 H\Big(\pi,\boldsymbol{\theta}_0 + t(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\Big)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)dt,$$

  这给出了

  $$-\frac{1}{\sqrt{n}}\nabla\mathcal{S}(\pi,\boldsymbol{\theta}_0) = \int_0^1 \left[\frac{1}{n}H\Big(\pi,\boldsymbol{\theta}_0 + t(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\Big)\right]\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)dt.$$

  根据 (i)，上面的左侧收敛于 $N(\mathbf{0}, B_1)$。对于 ˆ 右侧，利用 (ii) 以及 $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = O_P(1)$ 这一事实，对每个$t \in [0,\bar{1}]$，我们得到

  $$\frac{1}{n}H\Big(\pi,\boldsymbol{\theta}_0 + t(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\Big) \xrightarrow{P} B_2.$$

  利用 DCT，令 $n \to \infty$ 得到 $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{-} - \boldsymbol{\theta}_0) \xrightarrow{D} \mathbf{W}$，其中 $B_2\mathbf{W} \sim N(\mathbf{0}, B_1)$，或等价地，$\mathbf{W} \sim N(\mathbf{0}, B_2^{-1}B_1B_2^{-1})$。这就完成了该引理的证明。

  □

3.5. 引理的证明 *2.4*

如引理的证明中所示 2.3，对于每个 $\delta > 0$，令

$$B(\boldsymbol{\theta}_0,\delta) := \{\boldsymbol{\theta} \in \mathbb{R}^L : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \le \delta\}, \quad \partial B(\boldsymbol{\theta}_0,\delta) := \{\boldsymbol{\theta} \in \mathbb{R}^L : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 = \delta\}.$$

我们将该引理的证明分为几个步骤。

- 对于每个 $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$ 存在 $\eta := \eta_{\boldsymbol{\theta}_1} > 0$ 使得

  $$\mathbb{P}_{n,\boldsymbol{\theta}_0}\Big(\mathcal{S}(\pi,\boldsymbol{\theta}_0) < \inf_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}_1,\eta)} \mathcal{S}(\pi,\boldsymbol{\theta})\Big) \to 1.$$

  证明。 反设不存在这样的 $\eta > 0$。那么存在一个正实数序列 $\{\eta_n\}_{n\ge1}$，使得沿着 n 中的一个子序列，我们有

  $$\limsup_{n\to\infty} \mathbb{P}_{n,\boldsymbol{\theta}_0}\Big(\mathcal{S}(\pi,\boldsymbol{\theta}_0) \ge \inf_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}_1,\eta_n)} \mathcal{S}(\pi,\boldsymbol{\theta})\Big) > 0. \tag{34}$$

利用假设 (ii)，我们得到

$$\left|\mathcal{S}(\pi,\boldsymbol{\theta}_1)-\inf_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}_1,\eta_n)}\mathcal{S}(\pi,\boldsymbol{\theta})\right|\leq M(\pi)\eta_n.$$

此外，利用假设 (i)，不失一般性，我们可以在集合上进行讨论

$$\mathcal{S}(\pi,\boldsymbol{\theta}_0)+\varepsilon<\mathcal{S}(\pi,\boldsymbol{\theta}_1).$$

在上述两个集合的交集上，我们有

$$\mathcal{S}(\pi,\boldsymbol{\theta}_0)\geq\mathcal{S}(\pi,\boldsymbol{\theta}_1)-M(\pi)\eta_n>\mathcal{S}(\pi,\boldsymbol{\theta}_0)+\varepsilon-M(\pi)\eta_n\Rightarrow M(\pi)>\frac{\varepsilon}{\eta_n}.$$

由此可得

$$\mathbb{P}_{n,\boldsymbol{\theta}_0}\left(\mathcal{S}(\pi,\boldsymbol{\theta}_0)\geq\inf_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}_1,\eta_n)}\mathcal{S}(\pi,\boldsymbol{\theta})\right)\leq o(1)+\mathbb{P}_{n,\boldsymbol{\theta}_0}\left(M(\pi)>\frac{\varepsilon}{\eta_n}\right)=o(1),$$

其中，最后一个等式利用了以下事实： $M(\pi)=O_P(1)$。但这与 (34) 矛盾，因此证明完成。 □

- $\mathbb{P}_{n,\boldsymbol{\theta}_0}(E_{n,\delta})\to 1$, where $E_{n,\delta}:=\left\{\pi\in S_n:\mathcal{S}(\pi,\boldsymbol{\theta}_0)<\inf_{\boldsymbol{\theta}\in\partial B(\boldsymbol{\theta}_0,\delta)}\mathcal{S}(\pi,\boldsymbol{\theta})\right\}$.

证明。对于每个 $\boldsymbol{\theta}_1\neq\boldsymbol{\theta}_0$，令 $\eta_{\boldsymbol{\theta}_1}$ 为上一步构造的对象。这些球的集合

$$\left\{B\left(\boldsymbol{\theta}_1,\eta_{\boldsymbol{\theta}_1}\right),\boldsymbol{\theta}_1\in\partial B(\boldsymbol{\theta}_0,\delta)\right\}$$

是紧集 $\partial B(\boldsymbol{\theta}_0,\ \delta)$ 的一个覆盖，因此存在一个正整数 $N<\infty$ 和一些点 $\{\boldsymbol{\theta}_a\}_{1\leq a\leq N}\in\partial B(\boldsymbol{\theta}_0,\ \delta)$，使得

$$\partial B(\boldsymbol{\theta}_0,\delta)\subseteq\cup_{a=1}^{N}B\left(\boldsymbol{\theta}_a,\eta_{\boldsymbol{\theta}_a}\right).$$

因此，我们有

$$\inf_{\boldsymbol{\theta}\in\partial B(\boldsymbol{\theta}_0,\delta)}\mathcal{S}(\pi,\boldsymbol{\theta})\geq\min_{1\leq a\leq N}\inf_{\boldsymbol{\theta}\in B\left(\boldsymbol{\theta}_a,\eta_{\boldsymbol{\theta}_a}\right)}\mathcal{S}(\pi,\boldsymbol{\theta}).$$

但根据上一步（以及 $\eta_{\boldsymbol{\theta}_a}$ 的选取），我们有

$$\mathbb{P}_{n,\boldsymbol{\theta}_0}\left(\mathcal{S}(\pi,\boldsymbol{\theta}_0)<\min_{1\leq a\leq N}\inf_{\boldsymbol{\theta}\in B\left(\boldsymbol{\theta}_a,\eta_{\boldsymbol{\theta}_a}\right)}\mathcal{S}(\pi,\boldsymbol{\theta})\right)\to 1.$$

结合上述两个公式，我们得到

$$\mathbb{P}_{n,\boldsymbol{\theta}_0}\left(\mathcal{S}(\pi,\boldsymbol{\theta}_0)<\inf_{\boldsymbol{\theta}\in\partial B(\boldsymbol{\theta}_0,\delta)}\mathcal{S}(\pi,\boldsymbol{\theta})\right)\to 1,$$

这正是我们想要的。 □

- $\mathbb{P}_{n,\boldsymbol{\theta}_0}(\widetilde{E}_{n,\delta}) \to 1$, where $\widetilde{E}_{n,\delta} := \Big\{\pi \in S_n : \nabla \mathcal{S}(\pi, \boldsymbol{\theta}) = \mathbf{0} \text{ has a root in } B(\boldsymbol{\theta}_0, \delta)\Big\}$.

  证明。首先，注意到函数 $\mathcal{S}(\pi, .)$ 在集合 $B(\boldsymbol{\theta}_0, \delta)$ 上存在全局极小值点，因为映射 $\mathcal{S}(\pi, .)$ 根据假设 (ii) 是连续的，且集合 $B(\boldsymbol{\theta}_0, \delta)$ 是紧致的。利用上述步骤，不失一般性，我们可以在集合 $E_{n,\delta}$ 上进行讨论。在此集合上，函数 $\mathcal{S}(\pi, .)$ 在集合 $B(\boldsymbol{\theta}_0, \delta)$ 上的任何全局极小值点 $\boldsymbol{\theta}_{n,\delta}$ 都必须在 $B(\boldsymbol{\theta}_0, \delta)$ 的内部取得，因此必须满足 $\nabla \mathcal{S}(\pi, \boldsymbol{\theta}_{n,\delta}) = \mathbf{0}$。因此我们已证明 $\widetilde{\ }$，即 $E_{n,\delta} \subseteq E_{n,\delta}$，至此证明完成。 □

- Under $\mathbb{P}_{n,\boldsymbol{\theta}_0}$ we have $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$.

  证明。利用上述步骤，不失一般性，我们可以在 $\widetilde{\ }$集合 $E_{n,1}$上进行分析。因此，存在 $\boldsymbol{\theta}_{n,1} \in B(\boldsymbol{\theta}_0, 1)$，使得$\nabla \mathcal{S}(\pi, \boldsymbol{\theta}_{n,1}) = \mathbf{0}$。由于函数 $\mathcal{S}(\pi, .)$ 是严格凸的，因此 $\boldsymbol{\theta}_{n,1}$ 是函数 $\mathcal{S}(\pi, .)$ 在 $\mathbb{R}^L$上的唯一全局极小值点。为了完成证明，只需证明对于任意 $\delta > 0$，我们有

  $$\mathbb{P}_{n,\boldsymbol{\theta}_0}(\|\boldsymbol{\theta}_{n,1} - \boldsymbol{\theta}_0\|_2 > \delta) \to 0.$$

  再次使用上述步骤，不失一般性地，我们也可以在集合 $\widetilde{\ }$ $\widetilde{\ }$ $\widetilde{\ }$ 上工作 $E_{n,\,\delta}$（注意 $E_{n,\,\delta} \subseteq E_{n,\,1}$）。在这个集合上，我们有 $\boldsymbol{\theta}_{n,\,1} = \boldsymbol{\theta}_{n,\,\delta}$，因此

  $$\|\boldsymbol{\theta}_{n,1} - \boldsymbol{\theta}_0\|_2 = \|\boldsymbol{\theta}_{n,\delta} - \boldsymbol{\theta}_0\|_2 \leq \delta.$$

  所需的一致性显而易见。 □

致谢

第一作者获得美国国家科学基金会资助的DMS-1712037和DMS-2113414号拨款。

参考文献

[1] Andersen, H. C. 和 Diaconis, P. (2007). 击中并运行：一种统一工具。*Journal de la soci´et´e fran∴caise de statistique* **1485–28.**[2] Awasthi, P., Blum, A., Sheffet, O. 和 Vijayaraghavan, A. (2014). 排名模型混合的学习。神经信息处理系统进展**27.**[3] Basu, R. 和 Bhatnagar, N. (2017). 随机马洛斯排列中最长单调子序列的极限定理。*Annales de l'InstitutHenri Poincar´e, Probabilit´es et Statistiques* **53** 1934–1951.[4] Besag, J. (1974). 空间相互作用与格点系统的统计分析。皇家统计学会学报：*B*辑（方法论）**36** 192–225.

[5] Besag, J. (1975). 非格点数据的统计分析。皇家统计学会学报：D辑（统计学家）24 179–195.[6] Borga, J., Das, S., Mukherjee, S. 和 Winkler, P. (2022). 随机排列的大偏差原理。arXiv预印本 arXiv:2206.04660.[7] Chatterjee, S. (2005). 具有可交换对的集中不等式。斯坦福大学。[8] Chen, H., Branavan, S., Barzilay, R. 和 Karger, D. R. (2009). 使用潜在排列进行内容建模。《人工智能研究杂志》36 129–163.[9] Critchlow, D. E.(2012). 部分排序数据分析的度量方法 34。施普林格科学与商业媒体。[10] Critchlow, D. E., Fligner, M. A. 和 Verducci, J. S. (1991). 排名的概率模型。《数学心理学杂志》35 294–318.
[11] Diaconis, P. (1988). 概率论与统计学中的群表示。讲义专著丛书 11 i–192.[12] Diaconis, P. 和 Ram, A.(2000). 使用 Iwahori-Hecke 代数技术分析系统扫描 Metropolis 算法。《密歇根数学杂志》48 157–190.[13] Feigin, P. D. 和 Cohen, A.(1978). 关于评委间一致性模型。皇家统计学会学报：B辑（方法论）40 203–213.[14] Fligner, M. A. 和 Verducci, J. S. (1986). 基于距离的排名模型。皇家统计学会学报：B辑（方法论）48 359–369.
[15] Fligner, M. A. 和 Verducci, J. S. (1988). 多阶段排名模型。《美国统计学会杂志》83 892–901.[16] Ghosal, P., Nutz, M. 和 Bernton, E. (2022). 熵正则最优传输与 Schr¨odinger 桥的稳定性。《泛函分析杂志》283 109622.[17] Gladkich, A.和 Peled, R. (2018). 马洛斯排列的循环结构。《概率论年刊》46 1114–1169.[18] Harchaoui, Z., Liu, L. 和 Pal, S. (2020). 通过混沌分解研究熵正则最优传输的渐近性。arXiv预印本 arXiv:2011.08963.[19] He, J. (2021). 马洛斯排列循环的中心极限定理。arXiv预印本 arXiv:2112.09789.[20] Hoeffding, W. (1951). 组合中心极限定理。《数理统计年刊》558–566.[21] Kondor, R., Howard, A. 和 Jebara, T. (2007). 使用对称群表示进行多目标跟踪。载于《人工智能与统计学》211–218。PMLR。[22] Lebanon, G. 和 Lafferty, J. (2002). Cranking：使用排列上的条件概率模型合并排名。载于 ICML 2 363–370.[23] Lebanon, G. 和 Mao, Y. (2007). 部分排序数据的非参数建模。神经信息处理系统进展 20。

[24] Mallows, C. L. (1957). 非零排名模型。I.*Biometrika* **44** 114– 130.
[25] Marden, J. I. (1996). 排名数据的分析与建模.CRCPress.[26] Meila, M. 和 Bao, L. (2010). 无限排名的指数模型。*J. Mach. Learn. Res.***11** 3481–3518.[27] Meila, M., Phadnis, K., Patterson, A. 和 Bilmes, J. A.(2012). 指数模型下的共识排名。*arXiv*预印本 *arXiv:1206.5265*.
[28] Mukherjee, S. (2016). 置换上的指数族估计。统计学年刊**44** 853–875.
[29] Mukherjee, S.(2016). 随机置换的不动点与循环结构。概率电子期刊**21** 1 – 18.[30] Starr, S. (2009). S n 上 Mallows 模型的热力学极限。数学物理杂志**50** 095208.