arXiv:2302.06782v1 [math.ST] 14 Feb 2023

组序贯抽样下趋近正态分布的有限样本界

Julian Aronowitz† 和 Jay Bartroff*

†Google Inc.，美国加利福尼亚州卡尔弗城

*美国得克萨斯州奥斯汀，德克萨斯大学奥斯汀分校统计与数据科学系

最后修改于2023年2月13日

摘要

在组序贯分析中，数据以批次形式收集和分析，直至满足预先定义的停止准则。在参数设定下，推断通常依赖于反复计算的最大似然估计量（MLEs）的极限渐近多元正态性；这一结果最初由 Jennison 和 Turnbull （1997）在一般正则条件下严格证明。本文利用 Stein方法，在相同条件下，为联合组序贯最大似然估计量与相应正态分布之间针对光滑检验函数的距离提供具有最优阶的非渐近界。我们的结果假设观测值相互独立，但允许数据具有异质性（即并非同分布）。我们考察了当数据来自指数族时，所得界如何得到简化。最后，我们给出一个将多元 Kolmogorov 距离与光滑函数距离联系起来的一般结果；该结果不仅将我们的结论扩展到前一种度量，而且本身也可能具有独立的研究价值。

1 引言

序贯分析是一种强大的统计框架，其样本量在开展研究之前并不确定；在许多类型的临床试验中（Bartroff 等，2013），以及其他应用领域，它都是占主导地位的统计方法。尤其是在临床试验中，数据通常按组收集并分析，直至满足预先设定的停止准则。在关于这一主题的权威教科书中，Jennison 和 Turnbull（2000）指出，与传统的固定样本量设计相比，组序贯分析可以显著缩短开展临床试验所需的时间。

组序贯分析通常基于反复计算的最大似然估计量（每组数据收集后计算一次），因此，统计检验的所有性质（例如期望样本量、第一类错误概率、检验效能）都是这些反复计算的最大似然估计量联合分布的函数。假设独立但不一定同分布的观测值 $Y_1, \ldots, Y_n \in \mathbb{R}^t$ 被划分为 K 组，其中 n_k 表示截至第 k 组（含该组）所观察到的观测值数量，$k = 1, \ldots, K$。设 $f_i(y_i; \theta)$ 为 Y_i 在 $\theta \in \Theta \subset \mathbb{R}^d$ 条件下的概率质量函数或密度函数。令 $\hat{\theta}_k = \hat{\theta}_k(Y_1, Y_2, \ldots, Y_{n_k}) \in \mathbb{R}^{d\times 1}$ 为基于前 k 组观测值的最大似然估计量，而 $\hat{\theta}^K = [\hat{\theta}_1^\intercal, \hat{\theta}_2^\intercal, \ldots, \hat{\theta}_K^\intercal]^\intercal \in \mathbb{R}^q$（其中 $q = dK$）则为组序贯最大似然估计量。此处及下文中，$(\cdot)^\intercal$ 表示转置。尽管通常难以精确确定 $\hat{\theta}^K$ 的分布，Jennison 和 Turnbull（1997）证明，在适当的正则条件下，该分布渐近服从多元正态分布（见定理 2.3）。鉴于最大似然估计量在相同正则条件下渐近服从正态分布这一经典结果，这一结果或许并不令人意外；见下文的定理 2.2。

由此 formulation 引出的自然问题是：需要多少样本，以及应将样本分配到哪些组，才能让渐近效应开始显现，使正态近似变得适用，同时又不会引入过大的误差？对此问题的一种回答形式，是给出组序贯最大似然估计量的多元正态近似误差上界。在本研究中，我们在异质独立数据的相同正则条件下，给出了组序贯最大似然估计量与相应正态分布之间距离的最优阶界（定理 3.1）。证明依赖于 Stein方法和泰勒级数论证。我们的证明最接近Anastasiou (2018年) 的做法；该文针对多元最大似然估计量推导了类似的界，但没有涉及联合且反复计算的组序贯结构。我们的结果推广了 Anastasiou (2018年)，因为当只有一个组时，我们的界就简化为 Anastasiou (2018年) 中的结果。

在本研究中，我们给出了一个上界，

$$|\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]|$$

其中 Z 是标准多元正态分布，X 是标准化的组序贯最大似然估计量，h 是某个函数类 $\mathcal{H}$中的检验函数。特别地，对于任意三次可微函数 $h : \mathbb{R}^q \to \mathbb{R}$，我们记$\|h\| := \sup |h|$，$\|h\|_1 := \sup_i \left|\frac{\partial}{\partial x_i} h\right|$，$\|h\|_2 := \sup_{i,j} \left|\frac{\partial^2}{\partial x_i \partial x_j} h\right|$，以及 $\|h\|_3 := \sup_{i,j,k} \left|\frac{\partial^3}{\partial x_i \partial x_j \partial x_k} h\right|$。然后令

$$\mathcal{H} = \{h : \mathbb{R}^q \to \mathbb{R} : h \text{ 是三次可微的，且有界 } \|h\|, \|h\|_1, \|h\|_2, \|h\|_3\}. \quad (1)$$

一元最大似然估计量的非渐近界最早由Anastasiou 和 Reinert (2017年) 提出，随后由 Anastasiou (2018年) 推广到多元情形。在 Anastasiou 和 Ley (2017年) 和Anastasiou 和 Gaunt(2018年) 的研究中，在最大似然估计量具有特殊形式这一附加假设下，分别对一元和多元情形的界进行了改进和简化。令 $\boldsymbol{Y} = (\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_n})$ 为由n 个独立同分布的 t维随机向量构成的随机样本，$\Theta \subset \mathbb{R}^d$，且 $\hat{} \ \boldsymbol{\theta}_n(\boldsymbol{Y})$ 为相应的最大似然估计量。这些作者考虑的特殊形式为

$$p\big(\hat{\boldsymbol{\theta}}_n(\boldsymbol{Y})\big) = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{Y_i}) = \frac{1}{n}\sum_{i=1}^{n} (g_1(\boldsymbol{y_i}), \ldots, g_d(\boldsymbol{y_i}))^\intercal \tag{2}$$

对于某些 $g:\ \mathbb{R}^t \to \mathbb{R}^d$ 以及 $p:\ \Theta \to \mathbb{R}^d$。符合这种形式的最大似然估计量的一个简单例子是均值和方差未知的独立正态分布数据。Anastasiou (2017年) 考虑了服从 m依赖的、针对同分布标量参数最大似然估计量的数据。组序贯最大似然估计量必然是多元的，因此这一特定结果无法推广到我们的设定。

Anastasiou 和 Reinert (2017年)、Anastasiou (2018年)、Anastasiou 和 Ley (2017年)、Anastasiou 和 Gaunt (2018年)，以及 Anastasiou(2017年) 中的每个界都具有最优阶 $\mathcal{O}(n^{-1/2})$，其定义基于略有不同的 $\mathcal{H}$；一般而言，多维界需要额外的有界导数。所有这些结果都使用 Stein 方法技术，并要求 $h \in \mathcal{H}$ 至少有界且绝对连续。Pinelis (2017年) 针对一元最大似然估计量，给出了关于 Kolmogorov 距离的正态性收敛速率最优阶界，$\mathcal{H} = \{\mathbb{I}\{x \le a\}:\ a \in \mathbb{R}\}$。与上述结果不同，这个界并未使用 Stein 方法。我们的推论5.1 可视为将这一结果推广到多元最大似然估计量和组序贯设定；据我们所知，这是首个通过任何方法得到的此类结果。

由特征函数推导出的界（乌利亚诺夫，1979，1986，1987）在我们的情形中或许有用，但存在两个潜在障碍。首先，在多变量情形下，这些方法要求独立性。由于我们的目标之一是开发一种方法，以便日后能够应用于具有某种形式依赖性的观测数据，因此特征函数方法似乎无法满足这一目标。其次，这些方法无法为所得的界提供显式常数。

本文其余部分安排如下。在第 2 节，我们介绍关于最大似然估计量和Stein方法所需的其余背景知识。在第 3 节，我们给出主要结果，并在第 4 节将其具体应用于指数族分布和指数分布。在第 5 节，我们给出多元 Kolmogorov 距离与光滑函数距离之间关系的结果，并将结果推广至前一种度量，该度量

以柯尔莫哥洛夫距离给出多元最大似然估计量正态性的界。

2 背景

2.1 Stein方法

下面详述的结果大量运用了 Stein方法的界，并结合多元泰勒级数论证。本文不试图在此完整介绍 Stein方法，而是着重讨论相关的多元结果。对更全面介绍感兴趣的读者可参阅 Chen 等人 (2010)。

多元正态逼近中的 Stein方法依赖于以下多元版 Stein方程。设 $f:\mathbb{R}^q \to \mathbb{R}$ 二阶可微，且 D^2 为 f的二阶导数，即海森矩阵。那么，对于 $h:\mathbb{R}^q \to \mathbb{R}$，多元 Stein方程为

$$\mathrm{Tr}D^2 f(\boldsymbol{w}) - \boldsymbol{w}\cdot\nabla f(\boldsymbol{w}) = h(\boldsymbol{w}) - Nh.$$

这里 Nh 表示期望 $\mathbb{E}h(Z)$，其中 Z 是一个 q维标准正态向量。以下由 Goldstein 和 Rinott (1996) 给出的界，将 Stein方程解 f 的导数与检验函数 h的导数联系起来:

$$\left\| \frac{\partial^k f(\boldsymbol{w})}{\prod_{j=1}^k \partial w_{i_j}} \right\| \le \frac{1}{k} \left\| \frac{\partial^k h(\boldsymbol{w})}{\prod_{j=1}^k \partial w_{i_j}} \right\|, \quad k \ge 1, \tag{3}$$

当 h 的 k 阶偏导数存在时。

2.1.1 可交换对

将 Stein方法应用于随机向量 W 通常需要构造一个耦合随机变量 W'，并利用它与 W 之间的关系，得到 W 与其极限分布之间距离的界。一个可交换对是指一个耦合（W，W'），使得（W，W'）与（W'，W）具有相同的分布。我们采用的主要结果由 Reinert 和 R¨ollin（2009）提出（见下文定理 2.1），它要求满足某种松弛线性条件

$$\mathbb{E}[W' - W \mid W] = -\Lambda W + R \tag{4}$$

关于可交换向量对 (W, W')，其中 Λ 是一个可逆矩阵，R 是一个“误差”项。以下对于矩阵 X，令 $|X|:=\sup_i$

，　$|X_{ij}|_j$

定理 2.1 （Reinert 和 R¨ollin （2009））。 假设 （W， W'）是一个可交换的 $\mathbb{R}^d$值随机向量对，满足

$$\mathbb{E}W = 0, \quad \mathbb{E}WW^\intercal = \Sigma$$

其中 $\Sigma \in \mathbb{R}^{d\times d}$ 为对称正定矩阵。进一步假设 (4) 对可逆矩阵 Λ 以及一个 $\sigma(W)$- 可测随机向量 R 成立。那么，如果 Z 服从 d 维标准正态分布，则对于每个三次可微函数 h,

$$\left|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{\frac{1}{2}}Z)\right| \leq \frac{\|h\|_2}{4}A + \frac{\|h\|_3}{12}B + \left(\|h\|_1 + \frac{1}{2}d\,|\Sigma|^{1/2}\,\|h\|_2\right)C,$$

其中， $\lambda^{(i)} := \sum_{m=1}^{d} \left|(\Lambda^{-1})_{m,\ i}\right|$,

$$A = \sum_{i,j=1}^{d} \lambda^{(i)} \sqrt{\mathrm{Var}\left[\mathbb{E}\{(W_i' - W_i)(W_j' - W_j) \mid W\}\right]},$$

$$B = \sum_{i,j,k=1}^{d} \lambda^{(i)} \mathbb{E}\left|(W_i' - W_i)(W_j' - W_j)(W_k' - W_k)\right|,$$

$$C = \sum_{i=1}^{d} \lambda^{(i)} \sqrt{\mathrm{Var}\left[R_i\right]}$$

对于给定的随机向量 W，定理中使用哪个 W'“最好”可能并不明显，而且不同的选择会得到不同的界。从常数 A、 B 和 C 通常可以看出，如果某种 W' 能使 R 的方差较小，且使 W' 与 W“接近”，那么这种 W' 就比较理想。当 W 是随机向量之和时，一种常见的做法是从各个加数中均匀随机地选取一个，用其独立同分布副本替换它，并令 W' 为所得的和；这也是我们在此采用的方法。

2.2 最大似然估计量的渐近正态性

对于 $Y_1, Y_2, \ldots, Y_n \in \mathbb{R}^t$，假设它们相互独立，但不一定同分布；令 $f_i(y_i; \theta)$为 Y_i的概率质量函数或密度函数，其中参数空间 Θ 是 $\mathbb{R}^d$的一个开子集。假设观测值 $Y_1, \ldots, Y_n$ 被划分为 K 组， n_k 表示截至并包括第 k 组已观测到的观测值数量，其中 $k = 1, \ldots, K$ 满足$n = n_K$。例如，
$\{Y_1, Y_2, \ldots, Y_{n_3}\}$ 表示前3组中的所有观测值。令$\hat{}$ $\theta_k = \hat{}\theta_k(Y_1, Y_2, \ldots, Y_{n_k})$ $\in \mathbb{R}^d$表示基于前 k 组观测值的最大似然估计量，
$G_k = \{n_{k-1} + 1, \ldots, n_k\}$表示第 k组中观测值的指标集合。似然函数

在分析 k 时，似然函数为 $L_k(\theta; \boldsymbol{y}) = \prod_{i=1}^{n_k} f_i(y_i|\theta)$，对数似然函数为$\ell_k(\theta, \boldsymbol{y}) = \log(L_k(\theta; \boldsymbol{y}))$。

下面给出众所周知的正则条件，它们足以保证固定样本最大似然估计量的渐近正态性。我们将这些条件记录为定理 2.2。这些条件对于组序贯设定下的渐近正态性也足够；这是 Jennison 和 Turnbull(1997) 的一项结果，我们在此将其记录为定理 2.3。

(R1) $\hat{\theta}_n(Y) \xrightarrow{p} \theta_0$，当 $n \to \infty$ 时，其中 θ_0 是真实参数向量。
(R2) 得分函数 $S_i(Y_i; \theta) = \nabla_\theta \log f_i(Y_i; \theta) \in \mathbb{R}^{d\times 1}$ 和信息矩阵 $I_i(\theta) = \mathrm{Var}_\theta[S_i(Y_i; \theta)] \in \mathbb{R}^{d\times d}$，$i = 1, \ldots, n$ 均几乎必然存在（相对于概率测度 $\mathbb{P}$）。

(R3) $I_i(\theta)$ 是关于 $\theta, \forall i = 1, 2, \ldots, n$ 的连续函数（相对于 $\mathbb{P}$ 几乎必然），并且是关于 Y_i 的可测函数。
(R4) $\mathbb{E}_\theta[S_i(Y_i; \theta)] = 0_d \in \mathbb{R}^{d\times 1}$ 其中 0_d 是全零的 d 列向量。

(R5) $I_i(\theta) = \mathbb{E}_\theta[[\nabla \log f_i(Y_i; \theta)][\nabla \log f_i(Y_i; \theta)]^\intercal] = \mathbb{E}_\theta[-\nabla_\theta S_i^\intercal(Y_i; \theta)]$.
(R6) 对于

$$\bar{I}_n(k, \theta) = \frac{1}{n}\sum_{i=1}^{n_k} I_i(\theta) \; k = 1, \ldots, K,$$

存在一个矩阵 $I(k, \theta) = \lim_{n\to\infty} I_n(k, \theta)$。此外，对于所有 θ，$I_n(k, \theta)$ 和 $I(k, \theta)$ 均为对称矩阵，且 $I(k, \theta)$ 为正定矩阵。
(R7) 对某个 $\delta > 0$, $\frac{\sum_i \mathbb{E}_{\theta_0}|\boldsymbol{\lambda}^\intercal S_i(Y_i, \theta)|^{2+\delta}}{n^{\frac{2+\delta}{2}}} \longrightarrow 0$，对所有 $\boldsymbol{\lambda} \in \mathbb{R}^d$均成立。

(R8) 令 $\|\cdot\|$ 为 $\mathbb{R}^d$ 上通常的欧几里得范数，则对于 $i, j, u \in \{1, 2, \ldots, d\}$，存在 $\epsilon > 0, C > 0, \delta > 0$ 以及随机变量$B_{i,j,u}(Y_u)$，使得

$$\text{(i)} \;\; \sup\left\{\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log(f_u(Y_u; t))\right| \mid \|t - \theta_0\| \le \epsilon\right\} \le B_{i,j,u}(Y_u),$$
$$\text{(ii)} \;\; \mathbb{E}\,|B_{i,j,u}(Y_u)|^{1+\delta} \le C.$$

最大似然估计量渐近服从正态分布的经典结果如下。证明参见 van der Vaart (2007年)，有关充分条件的进一步讨论参见 Hoadley (1971年)。

定理 2.2。设 Y_1，Y_2，……，Y_n 为独立随机向量，其概率密度（或质量）函数为 $f_i(y_i; \theta)$，其中 $\theta \in \Theta \subset \mathbb{R}^d$。假设最大似然估计量存在且唯一，并且正则条件 *(R1)*-*(R8)*成立。另设 $Z \sim \mathcal{N}_d(0_d, 1_{d\times d})$ 为标准的多元 d 维正态分布。则，

$$\sqrt{n}[\bar{I}_n(\theta_0)]^{\frac{1}{2}}\left(\hat{\theta}_n(Y) - \theta_0\right) \xrightarrow[n\to\infty]{d} Z$$

定理 2.3（Jennison 和 Turnbull（1997））。假设观测值Y_i 相互独立，服从分布 $f_i(y_i;\theta)$，其中 θ 的维数为 d，且在分析 k时可获得观测值 $Y_1,\ldots,Y_{n_k}$，$k=1,\ldots,K$。令 $n_k-n_{k-1}\to\infty$ 满足：对所有 $k=1,\ldots,K$，$(n_k-n_{k-1})/n\to\gamma_k\in(0,1)$。此外，令 θ_k 表示基于 $Y_1,\ldots,Y_{n_k}$得到的 θ 的最大似然估计量。假设分布 f_i 足够正则，使得对每个 k，条件 (R1)-(R8) 均成立。另设 $Z\sim\mathcal{N}_q(0_q,1_{q\times q})$，且 $\theta_0^K:=[\theta_0^\intercal,\theta_0^\intercal,\ldots,\theta_0^\intercal]^\intercal\in\mathbb{R}^q$。那么，$\hat{}\theta^K=[\hat{}\theta_1^\intercal,\ldots,\hat{}\theta_K^\intercal]^\intercal$ 渐近服从多元正态分布，

$$\sqrt{n}[J_n]^{-\frac{1}{2}}(\hat{\theta}^K-\theta_0^K)\xrightarrow[n\to\infty]{d}Z$$

其中，

$$J_n=J_n(\theta_0)=\begin{bmatrix}\bar{I}_n^{-1}(1,\theta_0) & \bar{I}_n^{-1}(2,\theta_0) & \bar{I}_n^{-1}(3,\theta_0) & \ldots & \bar{I}_n^{-1}(K,\theta_0)\\ \bar{I}_n^{-1}(2,\theta_0) & \bar{I}_n^{-1}(2,\theta_0) & \bar{I}_n^{-1}(3,\theta_0) & & \\ \bar{I}_n^{-1}(3,\theta_0) & \bar{I}_n^{-1}(3,\theta_0) & \bar{I}_n^{-1}(3,\theta_0) & & \\ & \vdots & & \ddots & \\ \bar{I}_n^{-1}(K,\theta_0) & \bar{I}_n^{-1}(K,\theta_0) & \bar{I}_n^{-1}(K,\theta_0) & \ldots & \bar{I}_n^{-1}(K,\theta_0)\end{bmatrix}.$$

3 组序贯最大似然估计量的正态近似界

定理 2.3 中Jennison 和 Turnbull(1997) 关于组序贯最大似然估计量的渐近理论保证，在满足适当条件时具有渐近正态性。本节给出在相同条件下正态近似的误差界结果。在第 3.1 节中，我们给出主要结果，即定理 3.1；其证明推迟至附录A ，但在定理之后，我们将概述一个论证，说明该界在渐近意义下为 $\mathcal{O}(n^{-1/2})$。第 3.2 节讨论 Gaunt (2016) 的一项结果，该结果允许放宽对检验函数所需导数数量的要求 h，并将其应用于定理 2.1 ，从而在定理 3.3中实现类似的放宽，这一结果或许也有独立的意义。随后，在定理 3.3 中，我们将这一放宽传递给定理 3.1。

3.1 主要结果

为便于表述，我们引入以下额外记号。对于$\theta_0^K:=[\theta_0^\intercal,\theta_0^\intercal,\ldots,\theta_0^\intercal]^\intercal\in\mathbb{R}^q$，令 $Q_i:=[\hat{}\theta^K-\theta_0^K]_i$和 $Q_{(m)}:=\max_{i\in\{1,\ldots,q\}}Q_i$。令 Y_i'为 Y的一个独立副本，

$$\xi_{ij}=[n^{-\frac{1}{2}}S_i(Y_i,\theta_0)]_j,\quad\text{and}\quad\xi_{ij}'=[n^{-\frac{1}{2}}S_i(Y_i',\theta_0)]_j.$$

定义

$$A = \begin{bmatrix} 1_{d\times d} & 0 & 0 & \dots & 0 \\ 1_{d\times d} & 1_{d\times d} & 0 & & 0 \\ 1_{d\times d} & 1_{d\times d} & 1_{d\times d} & & 0 \\ & \vdots & & \ddots & \\ 1_{d\times d} & 1_{d\times d} & 1_{d\times d} & \dots & 1_{d\times d} \end{bmatrix} \in \mathbb{R}^{q\times q},$$

$$\tilde{J}_n = \tilde{J}_n(\theta) = \begin{bmatrix} \bar{I}_n(1,\theta) & \bar{I}_n(1,\theta) & \bar{I}_n(1,\theta) & & \bar{I}_n(1,\theta) \\ \bar{I}_n(1,\theta) & \bar{I}_n(2,\theta) & \bar{I}_n(2,\theta) & \dots & \bar{I}_n(2,\theta) \\ \bar{I}_n(1,\theta) & \bar{I}_n(2,\theta) & \bar{I}_n(3,\theta) & & \bar{I}_n(3,\theta) \\ & \vdots & & \ddots & \\ \bar{I}_n(1,\theta) & \bar{I}_n(2,\theta) & \bar{I}_n(3,\theta) & \dots & \bar{I}_n(K,\theta) \end{bmatrix} \in \mathbb{R}^{q\times q}$$

其中， $1_{d\times d}$ 是 d维单位矩阵。我们将矩阵 W 的第 n行、第 m列记为

$$n\text{th row: } W_{n*} = W_{n,*}$$
$$m\text{th columns: } W_{*m} = W_{*,m}.$$

当 W 定义为分块矩阵时，我们将第 i 块行、第 j块列中的子矩阵记为 $W_{[i][j]}$ 。类似地，当 $W \in \mathbb{R}^q$ 定义为分块向量时，我们将第 i个子向量记为 $W_{[i]} \in \mathbb{R}^d$ 。我们将第 n块行和第 m块列记为

$$n\text{th block row: } W_{[n][*]} = W_{[n],[*]}$$
$$m\text{th block columns: } W_{[*][m]} = W_{[*],[m]}.$$

我们已具备所需的记号，可以陈述本文的主要结果。

定理 3.1. 设 $Y_1, Y_2, \dots, Y_n$ 是相互独立但分布可能不相同的 $\mathbb{R}^t$ 值随机向量，其概率密度（或概率质量）函数为 $f_i(y_i|\theta)$，参数空间 Θ 是 $\mathbb{R}^d$ 的一个开子集。假设在分析时点 $k = 1, \dots, K$ 可获得观测值 $Y_1, Y_2, \dots, Y_{n_k}$。假设最大似然估计量 θ_k 存在且唯一，并且在每个分析时点 k 均满足条件 (R1)-(R6)。此外，假设对任意 θ_0，存在 $0 < \epsilon = \epsilon(\theta_0)$ 及函数 $M^k_{iuj}(y)$，$\forall i, u, j = 1, 2, \dots, d$，使得对所有满足 $|\theta_j - \theta_{0,j}| < \epsilon \ \forall j = 1, 2, \dots, d$ 的 $\theta \in \Theta$，均有 $\left|\frac{\partial^3}{\partial\theta_i\partial\theta_u\partial\theta_j}\ell_k(\theta, y)\right| \le M^k_{iuj}(y)$。还假设对所有 $k = 1, \dots, K$，均有 $\mathbb{E}[\left(\frac{\partial^3}{\partial\theta_i\partial\theta_u\partial\theta_j}M^k_{iuj}(Y)\right)^2 \mid |Q_{(m)}| < \epsilon] < \infty$。令 $\{Y'_i, i = 1, 2, \dots, n\}$ 为 $\{Y_i, i = 1, 2, \dots, n\}$ 的一个独立副本。对于 $Z \sim N_q(0_q, 1_{q\times q})$ $h \in \mathcal{H}$，其中 $\mathcal{H}$ 如 (1) 所示，有

成立：

$$\left|\mathbb{E}[h(\sqrt{n}J_n^{-\frac{1}{2}}(\hat{\theta}^K-\theta_0^K))]-\mathbb{E}[h(Z)]\right|\le\frac{\|h\|_1}{\sqrt{n}}K_1(\theta_0)$$
$$+\frac{q^2c^2\|h\|_2}{4}K_2(\theta_0)+\frac{q^3c^3\|h\|_3}{12}K_3(\theta_0)+\frac{2\|h\|}{\epsilon^2}\mathbb{E}[\sum_{j=1}^{q}Q_j^2],\quad(5)$$

其中

$$K_1(\theta_0)=\sum_{k_1=1}^{K}\sum_{k_2=k_1}^{\min\{k_1+1,K\}}\sum_{l=1}^{d}\sum_{j=1}^{d}\left|\bar{I}_n^{-\frac{1}{2}}(G_{k_2};\theta_0)_{lj}\right|\Bigg\{\Bigg($$
$$\sum_{i=1}^{d}\sqrt{\mathbb{E}[[\hat{\theta}_{[k_1]}-\theta_0]_i]^2\mathbb{E}\Big(\frac{\partial^2}{\partial\theta_i\partial\theta_l}\ell_{k_1}(\theta_0,Y)+n\bar{I}_n(k_1,\theta_0)_{li}\Big)^2}\Bigg)$$
$$+\frac{1}{2}\sum_{i=1}^{d}\sum_{u=1}^{d}\left(\mathbb{E}\left\{([\hat{\theta}_{[k_1]}-\theta_0]_i)^2([\hat{\theta}_{[k_1]}-\theta_0]_u)^2\right\}\right)^{\frac{1}{2}}\left(\mathbb{E}[\left(M_{iul}^{k_1}(Y)\right)^2\mid\left|Q_{(m)}\right|<\epsilon]\right)^{\frac{1}{2}}\Bigg\}$$
$$K_2(\theta_0)=\sum_{k=1}^{K}\Big\{\sum_{j=1}^{d}\Big[\sum_{i\in G_k}\mathrm{Var}\left[\xi_{ij}^2\right]\Big]^{\frac{1}{2}}+2\sum_{i<j}\Big[\sum_{v\in G_k}\mathrm{Var}[\xi_{vi}\xi_{vj}]\Big]^{\frac{1}{2}}\Big\}$$
$$K_3(\theta_0)=\sum_{i=1}^{n}\mathbb{E}\Big[\sum_{j=1}^{d}\left|(\xi'_{ij}-\xi_{ij})\right|\Big]^3$$
$$c=\max_{k\in\{1,\ldots,K\}}\left|\bar{I}_n^{-1/2}(G_k;\theta_0)\right|.$$

定理的证明见附录A。

尽管定理中的界看起来很复杂，但它通常具有最优阶 $\mathcal{O}(n^{-1/2})$，这一点并不难看出。假设 $[I_n(k,\ \theta_0)]_j=\mathcal{O}(1)$。由定理 2.3，

$$\sqrt{n}\mathbb{E}[\hat{\theta}_k-\theta_0]\to 0_d.$$

ˆ因此 $\mathbb{E}[\theta_k-\theta_0]_j=o(1/\sqrt{n})$。再次由定理 2.3，

$$n\bar{I}_n^{\frac{1}{2}}(k,\theta_0)\mathrm{Cov}[\hat{\theta}_k]\bar{I}_n^{\frac{1}{2}}(k,\theta_0)\to 1_{d\times d}.$$

利用 $[\bar{I}_n(k,\theta_0)]_j=\mathcal{O}(1)$，可得 $n\mathrm{Var}\left[[\hat{\ }\theta_k]_j\right]\to 1$，因此 $\mathrm{Var}\left[[\hat{\ }\theta_k]_j\right]=\mathcal{O}(1/n)$。于是

$$\mathbb{E}([\hat{\theta}_k-\theta_0]_j^2)=\mathrm{Var}\left[[\hat{\theta}_k]_j\right]+(\mathbb{E}[\hat{\theta}_k-\theta_0]_j)^2=\mathcal{O}(n^{-1}).\quad(6)$$

由 (R2) 和 (R6) 以及 Y_1、……、Y_{n_k} 的独立性可见，

$$\mathbb{E}\Big(\frac{\partial^2}{\partial\theta_i\partial\theta_l}\ell_k(\theta_0,Y)+n\bar{I}_n(k,\theta_0)_{li}\Big)^2=\sum_{j=1}^{n_k}\mathrm{Var}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_l}\log(f_j(Y_j\mid\theta_0))\right]\quad(7)$$

这表明上述方程的左端为 $\mathcal{O}(n)$。由 (6)、(7) 和 $[\bar{I}_n(k,\theta_0)]_j = \mathcal{O}(1)$ 可得，$K_1(\theta_0) = \mathcal{O}(1)$，且 $K_3(\theta_0) = \mathcal{O}(n^{-\frac{1}{2}})$。$K_2(\theta_0) = \mathcal{O}(n^{-\frac{1}{2}})$ 取决于 ξ_{ij} 的四阶矩为 $\mathcal{O}\left(\frac{1}{n^2}\right)$。一般而言，我们有理由相信这确实成立，但尚未对此加以证明。下文将证明，对于指数族，K_2 确实为 $\mathcal{O}(n^{-1/2})$。当所有这些条件均成立时，(5) 的右端项确实为 $\mathcal{O}(n^{-1/2})$。关于该界阶数的详细信息，可参阅第 4.2 节中针对较简单的指数族情形所作的分析。

3.2 定理 3.1 的扩展

定理 2.1是证明定理 3.1的关键工具，它依赖于（3）来给多元 Stein 解的导数建立界。 Gaunt（2016，命题 2.1）为 Stein 解的导数找到了新的界，这些界对函数所需的导数少一个 h 见（3）。在定理 3.2中，我们利用这一结果放宽定理 2.1 的条件，要求h 只需二阶可微，而非三阶可微。代价是，关于 d 的界的阶数增加了 $d^{1/2}$倍。

定理 3.2。设 (W, W') 是一对 $\mathbb{R}^d$ 值随机向量的可交换对，使得

$$\mathbb{E}W = 0, \quad \mathbb{E}WW^\intercal = \Sigma$$

其中 $\Sigma \in \mathbb{R}^{d\times d}$ 为对称正定矩阵。进一步假设，存在可逆矩阵 Λ 和一个 $\sigma(W)$-可测随机向量R，使得 Z 具有 d维标准正态分布，则对于每个二阶可微函数 h，

$$\left|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{\frac{1}{2}}Z)\right| \le d^{1/2}\left|\Sigma^{-\frac{1}{2}}\right|\left(\frac{\|h\|_1}{\sqrt{\pi}}A + \frac{\|h\|_2\sqrt{2\pi}}{8}B + \left(\sqrt{\frac{\pi}{2}}\left\|h - \mathbb{E}h(\Sigma^{\frac{1}{2}}Z)\right\| + \frac{2d}{\sqrt{\pi}}|\Sigma|^{1/2}\|h\|\right)C\right),$$

其中，满足 $\lambda^{(i)} := \sum_{m=1}^d \left|(\Lambda^{-1})_{m,\ i}\right|$，

$$A = \sum_{i,j=1}^d \lambda^{(i)}\sqrt{\mathrm{Var}\left[\mathbb{E}\{(W_i' - W_i)(W_j' - W_j) \mid W\}\right]},$$

$$B = \sum_{i,j,k=1}^d \lambda^{(i)}\mathbb{E}\left|(W_i' - W_i)(W_j' - W_j)(W_k' - W_k)\right|,$$

$$C = \sum_{i=1}^d \lambda^{(i)}\sqrt{\mathrm{Var}\left[R_i\right]}$$

证明。 证明过程类似于定理 2.1的证明，其中用 Gaunt (2016，命题 2.1）代替（3）。因此省略细节。 □

我们现在将这一界的改进应用于提升主要结果，即定理3.1。

定理 3.3。设 $Y_1, Y_2, \ldots, Y_n$ 是相互独立且非同分布的 $\mathbb{R}^t$ 值随机向量，其概率密度（或质量）函数为 $f_i(y_i|\theta)$，其中参数空间 Θ 是 $\mathbb{R}^d$ 的一个开子集。假设在分析 $k = 1, \ldots, K$ 时分别获得观测值 $Y_1, Y_2, \ldots, Y_{n_k}$。假设最大似然估计量 θ_k 存在且唯一，并且在每次分析 k 时均满足条件 (R1)-(R6)。此外，假设对于任意 θ_0 ，存在 $0 < \epsilon = \epsilon(\theta_0)$ 以及函数 $M_{iuj}^k(y)$，$\forall i, u, j = 1, 2, \ldots, d$，使得对于所有满足 $|\theta_j - \theta_{0,j}| < \epsilon \ \forall j = 1, 2, \ldots, d$ 的 $\theta \in \Theta$，都有 $\left|\frac{\partial^3}{\partial\theta_i\partial\theta_u\partial\theta_j}\ell_k(\theta, y)\right| \leq M_{iuj}^k(y)$。此外，假设对于所有 $k = 1, \ldots, K$，均有 $\mathbb{E}[\left(\frac{\partial^3}{\partial\theta_i\partial\theta_u\partial\theta_j}M_{iuj}^k(Y)\right)^2 \mid \left|Q_{(m)}\right| < \epsilon] < \infty$。令 $\{Y_i', i = 1, 2, \ldots, n\}$ 为 $\{Y_i, i = 1, 2, \ldots, n\}$ 的一个独立副本。对于 $Z \sim N_q(0_q, 1_{q\times q})$ $h \in \mathcal{H}$，其中 $\mathcal{H}$ 是所有具有有界一阶和二阶导数的有界函数组成的类，有

$$\left|\mathbb{E}[h(\sqrt{n}J_n^{-\frac{1}{2}}(\hat{\theta}^K - \theta_0^K))] - \mathbb{E}[h(Z)]\right| \leq \frac{\|h\|_1}{\sqrt{n}}K_1(\theta_0) + \frac{q^{3/2}c^2\|h\|_1}{\sqrt{\pi}}K_2(\theta_0) + \frac{\sqrt{2\pi}q^{5/2}c^3\|h\|_2}{8}K_3(\theta_0) + \frac{2\|h\|}{\epsilon^2}\mathbb{E}[\sum_{j=1}^{q}Q_j^2],$$

其中，

$$K_1(\theta_0) = \sum_{k_1=1}^{K}\sum_{k_2=k_1}^{\min\{k_1+1,K\}}\sum_{l=1}^{d}\sum_{j=1}^{d}\left|\bar{I}_n^{-\frac{1}{2}}(G_{k_2};\theta_0)_{lj}\right|\Bigg\{\Bigg(\sum_{i=1}^{d}\sqrt{\mathbb{E}[[\hat{\theta}_{[k_1]} - \theta_0]_i]^2\mathbb{E}\Big(\frac{\partial^2}{\partial\theta_i\partial\theta_l}\ell_{k_1}(\theta_0, Y) + n\bar{I}_n(k_1, \theta_0)_{li}\Big)^2}\Bigg)$$

$$+ \frac{1}{2}\sum_{i=1}^{d}\sum_{u=1}^{d}\left(\mathbb{E}\{([\hat{\theta}_{[k_1]} - \theta_0]_i)^2([\hat{\theta}_{[k_1]} - \theta_0]_u)^2\}\right)^{\frac{1}{2}}\left(\mathbb{E}[\left(M_{iul}^{k_1}(Y)\right)^2 \mid \left|Q_{(m)}\right| < \epsilon]\right)^{\frac{1}{2}}\Bigg\}$$

$$K_2(\theta_0) = \sum_{k=1}^{K}\Big\{\sum_{j=1}^{d}\Big[\sum_{i\in G_k}\mathrm{Var}\left[\xi_{ij}^2\right]\Big]^{\frac{1}{2}} + 2\sum_{i<j}\Big[\sum_{v\in G_k}\mathrm{Var}[\xi_{vi}\xi_{vj}]\Big]^{\frac{1}{2}}\Big\}$$

$$K_3(\theta_0) = \sum_{i=1}^{n}\mathbb{E}\Big[\sum_{j=1}^{d}\left|(\xi_{ij}' - \xi_{ij})\right|\Big]^3$$

$$c = \max_{k\in\{1,\ldots,K\}}\left|\bar{I}_n^{-1/2}(G_k;\theta_0)\right|.$$

证明。 定理 3.1 （见附录A）的证明只需补充以下内容：用定理 3.2 代替定理 2.1 ，并注意到由（22）， $\left|\Sigma^{-1/2}\right| = c$。经过此调整后，证明的其余部分类似，因此省略细节。 □

Fang 等人(2018年)结合 Malliavin 微积分技术与可交换对方法，针对多元近似，在 Wasserstein 距离上获得了近乎最优的误差界。借助这一结果，可以进一步改进定理3.3 ，无需再要求 h具有有界的二阶导数，但代价是无法达到最优阶收敛。

4 指数族观测值的应用

本节将定理3.1 专门应用于观测值独立同分布且服从指数族的情形。随后我们将证明，该界的阶 $\mathcal{O}(n^{-1/2})$是最优的，并针对来自指数分布的寿命数据明确计算该界。

4.1 符号与设定

为使符号更符合指数族文献中的惯例，我们对部分符号作了少许调整。若对于 $\boldsymbol{\eta} \in \mathbb{R}^d$ ，Y 的密度为，则称 Y 的分布为典则多参数指数族分布，

$$f(y;\boldsymbol{\eta}) = \exp\left\{\sum_{j=1}^{d} \eta_j T_j(y) - A(\boldsymbol{\eta}) + S(y)\right\} \mathbb{I}^{\{y\in B\}} \tag{8}$$

其中， B 是与 $\boldsymbol{\eta}$无关的 y 的支撑集， $T(y) = [T_1(y), \ldots, T_d(y)]^\intercal$ 是自然充分统计量。在此，自然参数 $\boldsymbol{\eta}$ 扮演了上述关注参数 θ 的角色。只要从上下文中可以明确 η 的多维性，我们将交替使用 η和 $\boldsymbol{\eta}$ 。累积量函数 $A(\eta)$满足

$$\frac{\partial}{\partial \eta_i} A(\eta) = \mathbb{E}_\eta\left[T_i(Y)\right], \quad \mathrm{Var}_\eta\left[T(Y)\right] = H(A(\eta)) = \nabla_\eta^2 A(\eta). \tag{9}$$

为简化记号，我们对充分统计量的混合矩采用以下简化记法：

$$\begin{aligned}
\mu_i &= \mu_i(\eta) = \mathbb{E}_\eta\left[T_i(Y)\right] \\
\mu_{ij} &= \mu_{ij}(\eta) = \mathbb{E}_\eta\left[T_i(Y)T_j(Y)\right] \\
\mu_{ijk} &= \mu_{ijk}(\eta) = \mathbb{E}_\eta\left[T_i(Y)T_j(Y)T_k(Y)\right].
\end{aligned}$$

在第 4.2 节中，我们还将用到 $A(\eta)$ 的三阶偏导数，该偏导数出现在定理3.1中。我们有

$$\begin{aligned}\frac{\partial^3}{\partial\eta_i\partial\eta_j\partial\eta_k}A(\eta) &= \int \exp\{S(y)\}T_i(y)\Bigg[\\ &\left(T_j(y)-\frac{\partial}{\partial\eta_j}A(\eta)\right)\left(T_k(y)-\frac{\partial}{\partial\eta_k}A(\eta)\right)a(y,\eta)-a(y,\eta)\frac{\partial^2}{\partial\eta_j\partial\eta_k}A(\eta)\Bigg]dy\\ &= \mu_{ijk}-\mu_{ij}\mu_k-\mu_{ik}\mu_j+\mu_i\mu_j\mu_k-\operatorname{Cov}\left[T_j(Y),T_k(Y)\right]\mu_i\\ &= \mu_{ijk}-\mu_{ij}\mu_k-\mu_{ik}\mu_j-\mu_{jk}\mu_i+2\mu_i\mu_j\mu_k. \quad (10)\end{aligned}$$

定义

$$\mu^{\epsilon}_{iul}=\mu^{\epsilon}_{iul}(\eta_0)=\max_{\{\eta:|\eta_j-\eta_{0,j}|<\epsilon,\,\forall j\in\{1,\dots,d\}\}}|\mu_{ilk}-\mu_{il}\mu_k-\mu_{ik}\mu_l-\mu_{lk}\mu_i+2\mu_i\mu_l\mu_k| \quad (11)$$

$W=(T_1(Y)\dots,T_d(Y))^\intercal$ 的联合累积量生成函数为

$$K(s)=\log\mathbb{E}\left[e^{s^\intercal W}\right]=A(s+\eta)-A(\eta).$$

利用此表达式，对$A(\eta)$求导数，即可得到累积量，如下所示:

$$\nabla_{s^n}K(s)\big|_{s=0}=\nabla_{s^n}A(s+\eta)\big|_{s=0}=\nabla_{\eta^n}A(\eta).$$

因此，对 $A(\eta)$求偏导数可得 $[T_1(Y)\dots,T_d(Y)]$的联合累积量，而这些累积量又是 $[T_1(Y)\dots,T_d(Y)]$的矩的多项式函数。

现在我们转而讨论指数族的最大似然估计量，记为 $\hat{}\eta$。得分函数为

$$\nabla_\eta\ell(y^n;\eta)=\nabla_\eta\sum_{i=1}^n\left(\sum_{j=1}^d\eta_jT_j(y_i)-A(\eta)+S(y_i)\right)=\sum_{i=1}^nT(y_i)-n\nabla_\eta A(\eta)$$

并令其等于零，再利用（9）可得

$$\frac{1}{n}\sum_{i=1}^nT(y_i)=\nabla_\eta A(\eta)\big|_{\eta=\hat\eta}=\mathbb{E}_{\hat\eta}\left[T(Y)\right].$$

由于指数族的对数似然严格凹，因此如果最大似然估计量存在，它就是全局最大值。均值函数为

$$\tau(\eta)=[\tau_1(\eta),\dots,\tau_d(\eta)]^\intercal=[\mathbb{E}_\eta\left[T_1(Y)\right],\dots,\mathbb{E}_\eta\left[T_d(Y)\right]]^\intercal,$$

我们简洁地写作

$$\tau(\hat\eta)=\frac{1}{n}\sum_{i=1}^nT(y_i)\quad\text{or, equivalently,}\quad\hat\eta=\tau^{-1}\left(\frac{1}{n}\sum_{i=1}^nT(y_i)\right). \quad (12)$$

我们注意到（12）与形式（2）相符，因此可以将Anastasiou 和 Ley (2017年) 和 Anastasiou 和 Gaunt (2018年) 中得到的界用于指数族最大似然估计量的正态性距离。 Anastasiou 和 Ley (2017年) 在单参数指数族的情形下指出了这一事实，但Anastasiou 和 Gaunt (2018年) 并未明确讨论向多参数指数族的推广。

4.2 指数族最大似然估计量的正态性界

有了上述记号，我们可以在推论 4.1中陈述定理 3.1对指数族的特化形式。所得的界（13）完全可以用充分统计量 $T_1（Y）$，...，$T_d（Y）$、自然参数的真实值 η_0以及样本的三阶以内混合矩表示。不需要关于该分布的其他知识，尤其不需要累积量函数 $A（\eta）$ 和 $S（Y）$。尽管推论 4.1 中的界继承了定理 $n^{-1/2}$ 3.1的最优阶，但由于观测值是独立同分布的，直接分析该界的阶更为简单，我们将在第 4.3节的推论之后进行这一分析。

推论 4.1. 设 Y_1、Y_2、...、Y_n 为具有概率密度（或质量）函数 $f(y|\eta)$ 的独立同分布观测值，其中参数空间是 $\mathbb{R}^d$ 的一个开子集。假设 f 是由式 (8) 定义的指数族，并且在分析 $k = 1$、...、K 时可获得观测值 Y_1、Y_2、...、Y_{n_k}。假设最大似然估计量 $\hat{}\eta_k$ 存在，且在每次分析 k 时均满足条件 (R1)-(R6)。设 $\{Y_i', i = 1, 2, \ldots, n\}$ 是 $\{Y_i, i = 1, 2, \ldots, n\}$ 的一个独立副本。对于 $Z \sim N_q($ $0_q, 1_{q\times q})$ $h \in \mathcal{H}$，其中 $\mathcal{H}$ 如式 (1) 所示，μ^{ϵ}_{iul} 由式 (11) 定义，有

$$\left|\mathbb{E}[h(\sqrt{n}J_n^{-\frac{1}{2}}(\hat{\eta}^K - \eta_0^K))] - \mathbb{E}[h(Z)]\right| \leq \frac{\|h\|_1}{\sqrt{n}}K_1(\eta_0) + \frac{q^2c^2\|h\|_2}{4}K_2(\eta_0) + \frac{q^3c^3\|h\|_3}{12}K_3(\eta_0) + \frac{2\|h\|}{\epsilon^2}\mathbb{E}\Big[\sum_{j=1}^{q} Q_j^2\Big], \quad (13)$$

其中,

$$K_1(\eta_0) = \frac{1}{2}\sum_{k_1=1}^{K}\sum_{k_2=k_1}^{\min\{k_1+1,K\}} n_{k_1}\left(\frac{|G_{k_2}|}{n}\right)^{-\frac{1}{2}}\sum_{l=1}^{d}\sum_{j=1}^{d}\mathrm{Var}^{-\frac{1}{2}}\left[T(Y)\right]_{lj}\Bigg\{$$

$$\sum_{i=1}^{d}\sum_{u=1}^{d}\mu^{\epsilon}_{iul}\left(\mathbb{E}\left[\left(\tau^{-1}\left(\frac{1}{n_{k_1}}\sum_{s=1}^{n_{k_1}}T(y_s)\right)_i - \eta_{0,i}\right)^2 \times\right.\right.$$

$$\left.\left.\left(\tau^{-1}\left(\frac{1}{n_{k_1}}\sum_{s=1}^{n_{k_1}}T(y_s)\right)_u - \eta_{0,u}\right)^2\right]\right)^{\frac{1}{2}}\Bigg\}$$

$$K_2(\eta) = \frac{1}{\sqrt{n}}\sum_{k=1}^{K}\left(\frac{|G_k|}{n}\right)^{\frac{1}{2}}\Bigg\{\sum_{j=1}^{d}\left[\mathrm{Var}\left[(T_j(Y)-\mu_j(\eta_0))^2\right]\right]^{\frac{1}{2}} +$$

$$2\sum_{i<j}\left[\mathrm{Var}\left[(T_i(Y)-\mu_i(\eta_0))(T_j(Y)-\mu_j(\eta_0))\right]\right]^{\frac{1}{2}}\Bigg\}$$

$$K_3(\eta_0) = \frac{1}{\sqrt{n}}\mathbb{E}\left[\sum_{j=1}^{d}\left|T_j(Y')-T_j(Y)\right|\right]^3$$

$$c = \left|\mathrm{Var}^{-\frac{1}{2}}\left[T(Y)\right]\right|\max_{k\in\{1,\ldots,K\}}\left(\frac{|G_k|}{n}\right)^{-\frac{1}{2}}.$$

4.3 推论4.1 中的界的阶为 $\mathcal{O}(n^{-1/2})$

在第 3 节中，我们分析了定理 3.1 中界的阶，发现它在 $n^{-1/2}$ 一些温和假设下的阶为。将其特化到指数族的情形时，它仍然是 $\mathcal{O}(n^{-1/2})$ 并不令人意外，在此可以更直接地看出这一点。令 $|G_k|/n \to \gamma_k \in (0,1)$，如定理 2.3所示，界中的项 c 满足

$$c \to \left|\mathrm{Var}^{-\frac{1}{2}}\left[T(Y)\right]\right|\max_{k\in\{1,\ldots,K\}}\gamma_k^{-\frac{1}{2}}.$$

因此 $c = \mathcal{O}(1)$。同理， $K_2(\eta) = \mathcal{O}(n^{-1/2})$。函数$K_3(\eta)$ 恰好等于一个常数乘以 $n^{-1/2}$ ，因此 $\mathcal{O}(n^{-1/2})$ 也是如此。界中的最后一项，

$$\frac{2\|h\|}{\epsilon^2}\mathbb{E}\left[\sum_{j=1}^{q}Q_j^2\right]$$

取决于.....的顺序

$$\mathbb{E}\left[Q_j^2\right] = \mathbb{E}\left[(\hat{\eta}_j-\eta_{0,j})^2\right] = \mathrm{Var}\left[\hat{\eta}_j\right] + \left(\mathbb{E}\left[Q_j\right]\right)^2 = \mathcal{O}\left(n^{-1}\right).$$

这样只需考虑 $K_1(\eta_0)$。要使整个界为$\mathcal{O}(n^{-1/2})$， $K_1(\eta)$应为 $\mathcal{O}(1)$，因为 $K_1(\eta)$ 乘以了 $n^{-1/2}$。当且仅当

$$\left(\mathbb{E}\left[\left(\tau^{-1}\left(\frac{1}{n_{k_1}}\sum_{s=1}^{n_{k_1}}T(y_s)\right)_i-\eta_{0,i}\right)^2\left(\tau^{-1}\left(\frac{1}{n_{k_1}}\sum_{s=1}^{n_{k_1}}T(y_s)\right)_u-\eta_{0,u}\right)^2\right]\right)^{\frac{1}{2}}$$

是 $\mathcal{O}(n^{-1})$。根据柯西-施瓦茨不等式，上式有如下界

$$\left(\mathbb{E}\left[(\hat{\eta}_i-\eta_{0,i})^4\right]\mathbb{E}\left[(\hat{\eta}_u-\eta_{0,u})^4\right]\right)^{\frac{1}{4}}.$$

因此，所需的阶数可由以下事实推出：最大似然估计量某个分量的四阶中心矩为 $\mathcal{O}\left(n^{-2}\right)$；一元指数情形可参见 de A. Cysneiros 等人 (2001，方程 2)，一般多元情形则可参见 Peers 和 Iqbal (1985)。

界中的 ϵ 应视具体情况而定，可以选择使 K_1 项与最后的 $\frac{2\|h\|}{\epsilon^2}$ 项之和最小。注意，较小的 ϵ 会使最后一项增大，同时使 μ^{ϵ}_{iul} 在K_2中减小。

4.4 示例：指数分布

本节将推论 4.1 应用于观测值独立同分布于指数分布的情形。我们采用自然参数化

$$T(y)=-y,\quad A(\eta)=-\log(\eta),\quad S(y)=0,\quad B=[0,\infty).$$

我们在下面的推论中记录推论 4.1 在该分布中的应用，其中我们利用了 $\hat{}$ $\eta_{n_k}\sim$ 逆伽马分布$(n_k,n_k\lambda)$ 和$\mu^{\epsilon}_{iul}=2/(\eta_0-\epsilon)^3$。其余计算与推论 4.1 中的计算类似且更简单，因此我们省略不述。

推论 4.2. 设 $Y_1,Y_2,\ldots,Y_n$ 相互独立，且 $Y_i\sim$服从指数分布(η_0)，其中 $\mathbb{E}[Y_i]=\frac{1}{\eta_0}$。假设观测值 $Y_1,Y_2,\ldots,Y_{n_k}$ 在分析 $k=1,\ldots,K$时可用。对于 $Z\sim N_q(0_q,1_{q\times q})$ $h\in\mathcal{H}$，其中 $\mathcal{H}$ 如 (1)所示，有

$$\left|\mathbb{E}[h(\sqrt{n}J_n^{-\frac{1}{2}}(\hat{\eta}^K-\eta_0^K))]-\mathbb{E}[h(Z)]\right|\le\frac{\|h\|_1}{\sqrt{n}}K_1(\eta_0)+\frac{2K^2c^2\|h\|_2}{\eta_0^2\sqrt{n}}K_2(\eta_0)$$
$$+\frac{2\eta_0^2\|h\|}{\epsilon^2}\sum_{k=1}^{K}\left(\frac{1}{n_k}\right)\frac{n_k^2+2n_k}{(n_k-1)(n_k-2)},$$

其中,

$$K_1(\eta_0) = \frac{\eta_0^3\sqrt{3}}{(\eta_0-\epsilon)^3}\sum_{k_1=1}^{K}\sum_{k_2=k_1}^{\min\{k_1+1,K\}}\left(\frac{|G_{k_2}|}{n}\right)^{-\frac{1}{2}}\sqrt{\frac{n_{k_1}^4+(\frac{46}{3})n_{k_1}^3+8n_{k_1}^2}{(n_{k_1}-1)\cdots(n_{k_1}-4)}}$$

$$K_2(\eta) = \sum_{k=1}^{K}\left(\frac{|G_k|}{n}\right)^{\frac{1}{2}}$$

$$c = \max_{k\in\{1,\ldots,K\}}\left(\frac{|G_k|}{n}\right)^{-\frac{1}{2}}.$$

5 多元柯尔莫哥洛夫距离／光滑函数界

本节给出多元柯尔莫哥洛夫距离的一般上界

$$d_{kol}(W,Z) = \sup_{\boldsymbol{x}\in\mathbb{R}^p}|\mathbb{I}\{W\le\boldsymbol{x}\} - \mathbb{I}\{Z\le\boldsymbol{x}\}| \tag{14}$$

两个 p维随机向量 W， Z可用光滑函数类距离来表示，例如第3 节和4节中的结果。在（14）中，$\mathbb{I}\{W\le\boldsymbol{x}\}$ 是 p维“下象限”
$\{W\le\boldsymbol{x}\} = \{W_1\le x_1,\ldots,W_p\le x_p\}$的指示函数。这些界本身可能也具有独立的研究价值，但我们在此给出它们还有另一个原因：它们为将组序贯最大似然估计量的光滑函数界应用于正态分布和柯尔莫哥洛夫距离打开了大门，而这在统计应用中可能很有价值。

首先，我们稍微修改一下记号，以更明确地体现其对维数和导数阶数的依赖。对于 $\boldsymbol{k}=(k_1,\ldots,k_p)\in\mathbb{N}_0^p$，令$|\boldsymbol{k}| = \sum_{i=1}^p k_i$；对于偏导数为 $\mathbb{R}^p\to\mathbb{R}$ 的函数 h

$$h^{\boldsymbol{k}}(\boldsymbol{x}) = \frac{\partial^{k_1+\cdots+k_p}h}{\partial^{k_1}x_1\ldots\partial^{k_p}x_p}\quad \text{exist for all}\quad 0\le|\boldsymbol{k}|\le m,$$

以及 $\|\cdot\|$ 上确界范数，令 $L_m^\infty(\mathbb{R}^p)$ 表示满足以下条件的所有函数$h:\mathbb{R}^p\to\mathbb{R}$ 的集合

$$\|h\|_{L_m^\infty(\mathbb{R}^p)} = \max_{0\le|\boldsymbol{k}|\le m}\left\|h^{(\boldsymbol{k})}\right\|$$

为有限值。对于 $\mathbb{R}^p$ 中的随机向量 X 和 Y，令

$$\mathcal{H}_{m,\infty,p} = \{h\in L_m^\infty(\mathbb{R}^p) : \|h\|_{L_m^\infty(\mathbb{R}^p)}\le 1\},$$

定义

$$d_{m,\infty,p}(X,Y) = \|\mathcal{L}(X)-\mathcal{L}(Y)\|_{\mathcal{H}_{m,\infty,p}} = \sup_{h\in\mathcal{H}_{m,\infty,p}}\left|\mathbb{E}\left[h(X)\right]-\mathbb{E}\left[h(Y)\right]\right|.$$

承接前几节，例如，定理 3.1给出了标准化分组序贯最大似然估计量与 q 维标准正态分布之间的界，其中 $d_{3,\ \infty,\ q}$；定理 3.3 则给出了 $d_{2,\ \infty,\ q}$ 的界。

我们首先在定理 5.1中陈述一般结果，然后在推论 $m=3$ 5.1中针对该情形进行具体说明。

定理 5.1. 假设 W 和 Z 是 p维、 $p \geq 2$的随机向量，且 Z 的密度有界于常数 C_1。那么存在一个常数 $C_2 \geq 1$，它仅依赖于 $m, m \geq 1$，使得

$$d_{kol}(W,Z) \leq d_{m,\infty,p}(W,Z)^{\frac{p-1}{m+p-1}} \left(C_2^p + p + C_1 d_{m,\infty,p}(W,Z)^{\frac{1}{m+p-1}} \right).$$

证明。假设 $d_{m,\ \infty,\ p}(W,Z) \leq 1$，否则该界显然成立。考虑 $S_m : \mathbb{R} \to \mathbb{R}$，满足 $S_m \in L_m^\infty(\mathbb{R})$，并且

$$S_m(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ f(x) & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } 1 \leq x \end{cases}$$

其中 f 是某个函数，且对所有 $0 \leq x \leq 1$都有 $0 \leq f(x) \leq 1$。令$h_{z,\ \alpha}(x) = S_m(-x/\alpha + (\alpha+z)/\alpha)$，注意到由于 $h_{z,\ \alpha}(x)$ 支配$\mathbb{I}^{\{x \leq z\}}$，

$$\begin{aligned} P(W \leq \boldsymbol{z}) - P(Z \leq \boldsymbol{z}) \leq \mathbb{E}\left[\prod_{i=1}^p h_{z_i,\alpha}(W_i)\right] - \mathbb{E}\left[\prod_{i=1}^p h_{z_i,\alpha}(Z_i)\right] \\ + \mathbb{E}\left[\prod_{i=1}^p h_{z_i,\alpha}(Z_i)\right] - P(Z \leq \boldsymbol{z}). \end{aligned} \tag{15}$$

我们首先对（15）右端项的前两项进行界定。由于 $h_{z_i,\alpha} \in L_m^\infty(\mathbb{R})$，函数 $\boldsymbol{h}_{\boldsymbol{z},\ \alpha} := \prod_{i=1}^p h_{z_i,\alpha}$属于 $L_m^\infty(\mathbb{R}^p)$。因此

$$\begin{aligned} \|\boldsymbol{h}_{\boldsymbol{z},\alpha}\|_{L_m^\infty(\mathbb{R}^p)} &= \max_{0 \leq |\boldsymbol{k}| \leq m} \left\| \boldsymbol{h}_{\boldsymbol{z},\alpha}^{(\boldsymbol{k})} \right\| = \max_{0 \leq |\boldsymbol{k}| \leq m} \prod_i^p \left\| h_{z_i,\alpha}^{(k_i)} \right\| \\ &= \max_{0 \leq |\boldsymbol{k}| \leq m} \prod_i^p \left(\frac{1}{\alpha}\right)^{k_i} \left\| S_m^{(k_i)} \right\| \leq \max_{0 \leq |\boldsymbol{k}| \leq m} \prod_i^p \left(\frac{1}{\alpha}\right)^{k_i} \|S_m\|_{L_m^\infty(\mathbb{R})} \\ &= \left(\frac{1}{\alpha}\right)^m \|S_m\|_{L_m^\infty(\mathbb{R})}^p. \end{aligned}$$

Thus

$$\frac{\alpha^m \boldsymbol{h}_{\boldsymbol{z},\alpha}}{\|S_m\|_{L_m^\infty(\mathbb{R})}^p} \in \mathcal{H}_{m,\infty,p}.$$

于是（15）右端项的前两项可界定为

$$\mathbb{E}\left[\prod_{i=1}^{p} h_{z_i,\alpha}(W_i)\right] - \mathbb{E}\left[\prod_{i=1}^{p} h_{z_i,\alpha}(Z_i)\right] \leq \frac{C_2^p d_{m,\infty,p}(W,Z)}{\alpha^m} \tag{16}$$

对于 $C_2 = \|S_m\|_{L_m^\infty(\mathbb{R})}$。此外，$1 = \|S_m\| \leq \|S_m\|_{L_m^\infty(\mathbb{R})} = C_2$。

现在我们考虑 (15) 的最后两项。对于 $\boldsymbol{\alpha} = (\alpha, \ldots, \alpha) \in \mathbb{R}^p$,

$$\mathbb{E}\left[\prod_{i=1}^{p} h_{z_i,\alpha}(Z_i)\right] - P(Z \leq \boldsymbol{z}) \leq P(Z \leq \boldsymbol{z} + \boldsymbol{\alpha}) - P(Z \leq \boldsymbol{z}). \tag{17}$$

(17) 的右端项可由一个底面为 α^p、高为 C_1，且有 p 与各坐标轴平行、底面为 α^{p-1} 、长度为 $\Phi(z_i)$ 的“腿”的 p+1维长方体界定。因此

$$\begin{aligned} P(Z \leq \boldsymbol{z} + \boldsymbol{\alpha}) - P(Z \leq \boldsymbol{z}) &\leq C_1\alpha^p + \sum_{i=1}^{p} \Phi(z_i)\alpha^{p-1} \\ &\leq C_1\alpha^p + p\alpha^{p-1}. \end{aligned} \tag{18}$$

将 (15) 与 (16) 和 (18) 结合，得到

$$P(W \leq \boldsymbol{z}) - P(Z \leq \boldsymbol{z}) \leq \frac{C_2^p d_{m,\infty,p}(W,Z)}{\alpha^m} + C_1\alpha^p + p\alpha^{p-1}. \tag{19}$$

令 $\alpha = d_{m,\infty,p}(W,Z)^{1/(m+p-1)}$ 并代入上面的方程，得到:

$$P(W \leq \boldsymbol{z}) - P(Z \leq \boldsymbol{z}) \leq d_{m,\infty,p}(W,Z)^{\frac{p-1}{m+p-1}} \left(C_2^p + p + C_1 d_{m,\infty,p}(W,Z)^{\frac{1}{m+p-1}}\right).$$

同理,

$$P(W \leq \boldsymbol{z}) - P(Z \leq \boldsymbol{z}) \geq -d_{m,\infty,p}(W,Z)^{\frac{p-1}{m+p-1}} \left(C_2^p + p + C_1 d_{m,\infty,p}(W,Z)^{\frac{1}{m+p-1}}\right)$$

由此完成证明。 □

推论 5.1. 柯尔莫哥洛夫距离的界为 $d_{3,\infty,p}(\cdot,\cdot)$，如下所示,

$$d_{kol}(W,Z) \leq d_{3,\infty,p}(W,Z)^{\frac{p-1}{2+p}} \left(52.5^p + p + \frac{d_{3,\infty,p}(W,Z)^{\frac{1}{2+p}}}{(2\pi)^{p/2}}\right).$$

证明。 我们沿用定理 5.1的证明。令 S_m 为某个函数的 Hermite 插值多项式 f, 其在 $f(0) = 0$与 $f(1) = 1$之间满足 $f^{(i)}(0) = f^{(i)}(1) = 0$对所有 $i = 1, \ldots\ldots, m$ 均成立。因此 $S_m \in L_m^\infty(\mathbb{R})$，且

$$S_m(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x^{m+1} \sum_{k=0}^{m} \binom{m+k}{k}\binom{2m+1}{m-k}(-x)^k & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } 1 \leq x \end{cases}.$$

特别是，

$$S_3(x) = -20x^7 + 70x^6 - 84x^5 + 35x^4 \quad \text{for} \quad 0 \leq x \leq 1.$$

经过简单计算可得 $\|S_3\|_{L_3^\infty(\mathbb{R})} = 52.5$ ，结论随之得出。 □

6 结论与未来方向

我们推广了Anastasiou 和 Reinert （2017年）以及 Anastasio u （2018年）的结果，求得基于累积数据的一系列最大似然估计量的联合分布的最优阶界。该联合分布的近似正态性，是组序贯假设检验统计理论的基本假设，而组序贯假设检验是临床试验中的主流范式。我们所知，在非序贯情形下，推论 4.1 中将此界专用于指数族，得到的是涵盖该情形的最简单界。

推广这些结果的一个方向，是考察如下形式的对数似然函数

$$\sum_{i \in G_k} \log f_i(Y_i, \theta) + g_k(\mathcal{Y}_k, \theta),$$

其中 $\mathcal{Y}_k = \{Y_i : \ i \in G_k\}$。这种形式或许能提供一种放宽样本间独立性假设的方法，并适用于 Stein方法。此外，它也是某些具有随机阶段效应的广义线性混合模型（GLMMs）的对数似然函数形式，例如具有泊松响应变量和典则对数链接函数的广义线性混合模型。

参考文献

Anastasiou, A.（2017年）。m-依赖随机变量的最大似然估计量正态近似界。Statistics & Probability Letters，129:171–181。

Anastasiou, A.（2018年）。评估高维异质数据中最大似然估计量的多元正态近似。Electronic Journal of Statistics，12(2):3794–3828。

Anastasiou, A. 和 Gaunt, R. E.（2018+）。通过Delta方法对最大似然估计量进行多元正态近似。Brazilian Journal of Probability and Statistics，即将刊出。

Anastasiou, A. 和 Ley, C.（2017年）。使用 Delta 方法得到最大似然估计量渐近正态性的界。Latin American Journal of Probability and Statistics，14：153–171。

Anastasiou, A. 和 Reinert, G.（2017年）。最大似然估计量正态近似的界。Bernoulli，23(1)：191–218。Bartroff, J., Lai, T. L. 和 Shih, M.（2013）。临床试验中的序贯试验：设计与分析。施普林格，纽约。Chen, L. H., Goldstein, L. 和 Shao, Q.-M.（2010）。Stein方法的正态近似。施普林格，柏林海德堡。de A. Cysneiros, F. J., dos Santos, S. J. P. 和 Cordeiro, G. M.（2001）。单参数指数族模型中最大似然估计量的偏度与峰度。Brazilian Journal of Probability and Statistics，15(1)：85–105。Fang, X., Shao, Q.-M. 和 Xu, L.（2018年）。利用 Stein方法和 Bismut公式，在 Wasserstein 距离下进行多元近似。概率论及相关领域，在线优先发表：1–35。Gaunt, R. E.（2016）。通过 Stein方程解的新界，在矩条件下正态近似的收敛速率。理论概率杂志，26：231–247。Goldstein, L. 和 Reinert, G.（1997）。Stein方法与零偏变换及其在简单随机抽样中的应用。应用概率年刊，7(4)：935–952。Goldstein, L. 和 Rinott, Y.（1996）。利用 Stein方法和大小偏倚耦合进行多元正态近似。应用概率杂志，33(1)：1–17。Hoadley, B.（1971年）。独立非同分布情形下最大似然估计量的渐近性质。数学统计年刊，42(6)：1977–1991。Jennison, C. 和 Turnbull, B. W.（1997）。纳入协变量信息的组序贯分析。美国统计学会杂志，92(440)：1330–1341。Jennison, C. 和 Turnbull, B. W.（2000）。临床试验中的组序贯方法。Chapman & Hall，博卡拉顿，佛罗里达州。O’Brien, P. 和 Fleming, T. R.（1979）。临床试验中的多重检验程序。生物统计学，35(3)：549–56。Peers, H. W. 和 Iqbal, M.（1985）。存在干扰参数时置信限的渐近展开及其应用。英国皇家统计学会杂志：B辑（方法论），47(3)：547–554。

Pinelis, I.（2017年）。最大似然估计量正态性收敛速率的最优阶一致与非一致界。《电子统计学期刊》，11：1160–1179。Pocock, S. J.（1977）。临床试验设计与分析中的组序贯方法。《生物统计学》，64(2)：191–199。Reinert, G. 和 R¨ollin, A.（2009）。在线性条件下，利用可交换对的Stein方法进行多元正态近似。《概率年刊》，36(6)：2150–2173。Stein, C.（1972）。相依随机变量之和的分布正态近似误差界。载于《第六届伯克利数学统计与概率研讨会论文集》，第2卷：概率论，第583–602页，美国加利福尼亚州伯克利：加利福尼亚大学出版社。Ulyanov, V.（1979）。关于中心极限定理中更精确的收敛速度估计。《概率论及其应用》，23(3)：660–663。Ulyanov, V.（1986）。希尔伯特空间中非同分布随机变量之和的正态近似。《数学科学学报》，50：411–419。Ulyanov, V.（1987）。H中独立随机变量之和的分布渐近展开。《概率论及其应用》，31(1)：25–39。

van der Vaart，A. W.（2007年）。《渐近统计学》。剑桥大学出版社，美国纽约州纽约市。

附录

定理 3.1的证明

本附录提供定理 3.1的详细证明，随后给出两个辅助引理。

定理的证明 *3.1*。 根据三角不等式，我们将其界定为

$$\left|\mathbb{E}[h(\sqrt{n}J_n^{-\frac{1}{2}}(\hat{\theta}^K - \theta_0^K))] - \mathbb{E}[h(Z)]\right|$$

上界为

$$\left|\mathbb{E}[h(\frac{1}{\sqrt{n}}\tilde{J}_n^{-\frac{1}{2}}[S(1,\theta_0),\ldots,S(K,\theta_0)]^\intercal)]-\mathbb{E}[h(Z)]\right| \tag{20}$$

$$+\left|\mathbb{E}[h(\sqrt{n}J_n^{-\frac{1}{2}}(\hat{\theta}^K-\theta_0))]-\mathbb{E}[h(\frac{1}{\sqrt{n}}\tilde{J}_n^{-\frac{1}{2}}[S(1,\theta_0),\ldots,S(K,\theta_0)]^\intercal)]\right|. \tag{21}$$

其中 $S(k;\theta)=\sum_{i=1}^{n_k}S_i(Y_i;\theta)\in\mathbb{R}^d$。我们先求出（20）的一个上界。

（20）的上界。令 $\tilde{h}(w)=h(\tilde{J}_n^{-\frac{1}{2}}Aw)$，并定义

$$S(G_k;\theta)=\sum_{i\in G_k}S_i(Y_i;\theta)\in\mathbb{R}^d,$$

$$W=n^{-\frac{1}{2}}[S(G_1,\theta_0)^\intercal\ldots,S(G_K,\theta_0)^\intercal]^\intercal\in\mathbb{R}^q.$$

于是，$\tilde{h}(W)=h(n^{-\frac{1}{2}}\tilde{J}_n^{-\frac{1}{2}}[S(1,\theta_0)^\intercal\ldots,S(K,\theta_0)^\intercal]^\intercal)$。令

$$\begin{aligned}\Sigma=\mathrm{Var}\,[W]&=\frac{1}{n}\mathrm{diag}(\sum_{i\in G_1}I_i(\theta_0),\sum_{i\in G_2}I_i(\theta_0),\ldots,\sum_{i\in G_K}I_i(\theta_0))\\&=\mathrm{diag}(\bar{I}_n(G_1,\theta_0),\bar{I}_n(G_2,\theta_0),\ldots,\bar{I}_n(G_K,\theta_0))\end{aligned}$$

其中，我们定义 $\bar{I}_n(G_k,\theta)=\frac{1}{n}\sum_{i\in G_k}I_i(\theta)$。注意，这里的记号略有滥用，因为 $\bar{I}_n(k,\theta)=\frac{1}{n}\sum_{i=1}^{n_k}I_i(\theta)$。因此，

$$\tilde{h}(\Sigma^{\frac{1}{2}}Z)=h(\tilde{J}_n^{-\frac{1}{2}}A\Sigma^{\frac{1}{2}}Z).$$

现在我们断言 $\tilde{J}_n^{-\frac{1}{2}}A\Sigma^{\frac{1}{2}}=1_{q\times q}$。为说明这一点，注意到，

$$\begin{aligned}&&A\Sigma A^\intercal&=\tilde{J}_n\\\Longrightarrow\quad&&A\Sigma^{\frac{1}{2}}&=\tilde{J}_n^{\frac{1}{2}}\\\Longrightarrow\quad&&\tilde{J}_n^{-\frac{1}{2}}A\Sigma^{\frac{1}{2}}&=1_{q\times q}.\end{aligned}$$

其中，第一个等式很容易通过矩阵分块乘法验证。因此，

$$\tilde{h}(\Sigma^{\frac{1}{2}}Z)=h(Z).$$

这意味着（20）等于

$$\left|\mathbb{E}[\tilde{h}(W)]-\mathbb{E}[\tilde{h}(\Sigma^{\frac{1}{2}}Z)]\right|.$$

通过与上述类似的论证，我们可知

$$\tilde{J}_n^{-1/2}A=\Sigma^{-1/2}=\mathrm{diag}(\bar{I}_n^{-1/2}(G_1;\theta_0),\ldots,\bar{I}_n^{-1/2}(G_K;\theta_0)) \tag{22}$$

现在，我们希望将定理2.1中得到的 Stein方法界应用于上述方程。为此，我们需要 $\mathbb{E}[W]$ 和 Var $[W]$。由上文可知 Var $[W]=\Sigma$。为求得 $\mathbb{E}[W]$，注意到:

$$\mathbb{E}[W_{[i]}]=n^{-\frac{1}{2}}\mathbb{E}[S(G_i,\theta_0)]=n^{-\frac{1}{2}}\sum_{j\in G_i}\mathbb{E}[S_j(Y_j,\theta_0)]=0_d.$$

因此， $\mathbb{E}[W]=0_q$。

应用定理 2.1 的下一步，是为 W 找到一个适当的可交换对。令 $\{Y_i'$: $i=1,\ .....,\ n\}$ 为 $\{Y_i:\ i=1,\ .....,\ n\}$ 的独立副本。令 $I\in\{1,\ldots,n\}$ 为一个与 $\{Y_i'\}_{i=1,\ .....,\ n}$ 以及 $\{Y_i\}_{i=1,\ .....,\ n}$ 独立的均匀随机变量。然后，我们将 $\frac{1}{\sqrt{n}}S_I(Y_I,\ \theta_0)$ 替换为 $\frac{1}{\sqrt{n}}S_I(Y_I',\ \theta_0)$，使得

$$W'_{[\hat{k}]}=W_{[\hat{k}]}-\frac{1}{\sqrt{n}}S_I(Y_I,\theta_0)+\frac{1}{\sqrt{n}}S_I(Y_I',\theta_0)$$

^^其中， k 是包含观测值 Y_I 和 $W'_k=W_k$ （ $k\neq k$）的组。于是，

$$\begin{aligned}
\mathbb{E}[W'_{[k]}-W_{[k]}|W]&=\sum_{j=1}^{n}\mathbb{E}[W'_{[k]}-W_{[k]}|W,I=j]\mathbb{P}(I=j)\\
&=\frac{1}{n}\sum_{j\in G_k}\mathbb{E}[W'_{[k]}-W_{[k]}|W,I=j]\\
&=\frac{1}{n}\sum_{j\in G_k}\mathbb{E}[\frac{1}{\sqrt{n}}S_I(Y_I',\theta_0)-\frac{1}{\sqrt{n}}S_I(Y_I,\theta_0)|W,I=j]\\
&=\frac{1}{n}\sum_{j\in G_k}\mathbb{E}[\frac{1}{\sqrt{n}}S_j(Y_j',\theta_0)-\frac{1}{\sqrt{n}}S_j(Y_j,\theta_0)|W]\\
&=-\frac{1}{n}\sum_{j\in G_k}\mathbb{E}[\frac{1}{\sqrt{n}}S_j(Y_j,\theta_0)|W_{[k]}]\\
&=-\frac{1}{n}\mathbb{E}[\sum_{j\in G_k}\frac{1}{\sqrt{n}}S_j(Y_j,\theta_0)|W_{[k]}]\\
&=-\frac{1}{n}\mathbb{E}[W_{[k]}|W_{[k]}]\\
&=-\frac{W_{[k]}}{n}
\end{aligned}$$

因此，如定理 2.1，$\Lambda = \frac{1}{n}1_{q\times q}$，$R = 0_{q\times q}$，并由此可得

$$\left|\mathbb{E}[\tilde{h}(W)] - \mathbb{E}[\tilde{h}(\Sigma^{\frac{1}{2}}Z)]\right| \leq$$

$$\frac{n\left\|\tilde{h}\right\|_2}{4}\sum_{k_1,k_2=1}^{K}\sum_{i,j=1}^{d}\sqrt{\mathrm{Var}\left[\mathbb{E}[([W'_{[k_1]}]_i - [W_{[k_1]}]_i)([W'_{[k_2]}]_j - [W_{[k_2]}]_j)|W]\right]} \tag{23}$$

$$+\frac{n\left\|\tilde{h}\right\|_3}{12}\sum_{k_1,k_2,k_3=1}^{K}\sum_{i,j,u=1}^{d}\mathbb{E}\left|([W'_{[k_1]}]_i - [W_{[k_1]}]_i)\cdot\right.$$

$$\left.([W'_{[k_2]}]_j - [W_{[k_2]}]_j)([W'_{[k_3]}]_u - [W_{[k_3]}]_u)\right|. \tag{24}$$

现在我们希望对两者进行界定并简化，即 (23) 和 (24)。注意，根据构造，对于 $k_1 \neq k_2$,

$$([W'_{[k_1]}]_i - [W_{[k_1]}]_i)([W'_{[k_2]}]_j - [W_{[k_2]}]_j) = 0_d.$$

类似地,

$$([W'_{[k_1]}]_i - [W_{[k_1]}]_i)([W'_{[k_2]}]_j - [W_{[k_2]}]_j)([W'_{[k_3]}]_u - [W_{[k_3]}]_u) = 0_d$$

除非 $k_1 = k_2 = k_3$。因此，(23) 等于

$$\frac{n\left\|\tilde{h}\right\|_2}{4}\sum_{k=1}^{K}\sum_{i=1}^{d}\sum_{j=1}^{d}\sqrt{\mathrm{Var}\left[\mathbb{E}[([W'_{[k]}]_i - [W_{[k]}]_i)([W'_{[k]}]_j - [W_{[k]}]_j)|W]\right]} \tag{25}$$

而 (24) 等于

$$\frac{n\left\|\tilde{h}\right\|_3}{12}\sum_{k=1}^{K}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{u=1}^{d}\mathbb{E}\left|([W'_{[k]}]_i - [W_{[k]}]_i)([W'_{[k]}]_j - [W_{[k]}]_j)([W'_{[k]}]_u - [W_{[k]}]_u)\right|. \tag{26}$$

为了用 $\|h\|_2$ 表示 $\left\|\tilde{h}\right\|_2$，并用 $\|h\|_3$ 表示 $\left\|\tilde{h}\right\|_3$，我们使用 ~_1 链式法则。令 $x = J_n{}^2 Aw$，并定义 $|X| :=$ 对任意矩阵 $_{i,j}X$ 取上确界 $|X_{ij}|$。于是,

$$\frac{\partial\tilde{h}}{\partial w_j} = \sum_{i=1}^{q}\frac{\partial h}{\partial x_i}\frac{\partial x_i}{\partial w_j} = \sum_{i=1}^{q}\frac{\partial h}{\partial x_i}[\tilde{J_n}^{-\frac{1}{2}}A]_{ij}.$$

So,

$$\begin{aligned}\frac{\partial \tilde{h}^2}{\partial w_s \partial w_j} &= \frac{\partial}{\partial w_s}\sum_{i_1=1}^{q}\frac{\partial h}{\partial x_{i_1}}[\tilde{J}_n^{-\frac{1}{2}}A]_{i_1 j} = \sum_{i_1=1}^{q}[\tilde{J}_n^{-\frac{1}{2}}A]_{i_1 j}\sum_{i_2=1}^{q}\frac{\partial h^2}{\partial x_{i_2}\partial x_{i_1}}\frac{\partial x_{i_2}}{\partial w_s} \\ &= \sum_{i_1=1}^{q}[\tilde{J}_n^{-\frac{1}{2}}A]_{i_1 j}\sum_{i_2=1}^{q}\frac{\partial h^2}{\partial x_{i_2}\partial x_{i_1}}[\tilde{J}_n^{-\frac{1}{2}}A]_{i_2 s} \\ &\le \sum_{i_1=1}^{q}\sum_{i_2=1}^{q}[\tilde{J}_n^{-\frac{1}{2}}A]_{i_1 j}[\tilde{J}_n^{-\frac{1}{2}}A]_{i_2 s}\left|\frac{\partial h^2}{\partial x_{i_2}\partial x_{i_1}}\right| \\ &\le q^2\left|\tilde{J}_n^{-\frac{1}{2}}A\right|^2\|h\|_2 \\ &= q^2c^2\|h\|_2 .\end{aligned}$$

其中最后一个等式由（22）得出，并提醒读者

$$c = \max_{k\in\{1,\dots,K\}}\left|\bar{I}_n^{-1/2}(G_k;\theta_0)\right| .$$

同理，

$$\frac{\partial \tilde{h}^3}{\partial w_u \partial w_s \partial w_j} \le q^3c^3\|h\|_3 .$$

对上述每个方程的左端取上确界，可得：

$$\begin{aligned}\left\|\tilde{h}\right\|_2 &\le q^2c^2\|h\|_2 , \\ \left\|\tilde{h}\right\|_3 &\le q^3c^3\|h\|_3 .\end{aligned}$$

因此，结合上面的结果与式（25）和（26），可知式（23）小于或等于

$$\frac{nq^2c^2\|h\|_2}{4}\sum_{k=1}^{K}\sum_{i=1}^{d}\sum_{j=1}^{d}\left\{\mathrm{Var}\left[\mathbb{E}[([W'_{[k]}]_i-[W_{[k]}]_i)([W'_{[k]}]_j-[W_{[k]}]_j)|W]\right]\right\}^{\frac{1}{2}} \tag{27}$$

且式（24）小于或等于

$$\frac{nq^3c^3\|h\|_3}{12}\sum_{k=1}^{K}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{u=1}^{d}\mathbb{E}\left|([W'_{[k]}]_i-[W_{[k]}]_i)\cdot([W'_{[k]}]_j-[W_{[k]}]_j)([W'_{[k]}]_u-[W_{[k]}]_u)\right| . \tag{28}$$

为了给（27）中的条件期望的方差设界，我们提醒读者注意以下记号

$$\begin{aligned}\xi_{ij} &= [n^{-\frac{1}{2}}S_i(Y_i,\theta_0)]_j \\ \xi'_{ij} &= [n^{-\frac{1}{2}}S_i(Y'_i,\theta_0)]_j .\end{aligned}$$

我们还记

$$C_2 = \frac{q^2 c^2 \|h\|_2}{4}$$
$$C_3 = \frac{q^3 c^3 \|h\|_3}{12}.$$

现在令 $\mathcal{A} = \sigma(Y_1, Y_2, ..., Y_n)$。那么，由于 $\sigma(W) \subset \mathcal{A}$，对于任意随机变量 X，$\text{Var}\,[\mathbb{E}[X|W]] \leq \text{Var}\,[\mathbb{E}[X|\mathcal{A}]]$。因此，(27) 小于或等于

$$nC_2 \sum_{k=1}^{K} \Big\{ \sum_{j=1}^{d} \sqrt{\text{Var}\left[\mathbb{E}[(\xi'_{Ij} - \xi_{Ij})^2 \mathbb{I}\{I \in G_k\}|\mathcal{A}]\right]} + 2 \sum_{i<j} \sqrt{\text{Var}\left[\mathbb{E}[(\xi'_{Ii} - \xi_{Ii})(\xi'_{Ij} - \xi_{Ij}) \mathbb{I}\{I \in G_k\}|\mathcal{A}]\right]} \Big\}. \quad (29)$$

由于$\{Y'_i\}_{i=1,...,n}$是$\{Y_i\}_{i=1,...,n}$的独立副本，ξ'_{ik}与$\mathcal{A}$独立，且 I 与 $\{\xi'_{Ik}, \xi_{Ik}\}$独立，(29)变为

$$nC_2 \sum_{k=1}^{K} \sum_{j=1}^{d} \left\{ \text{Var} \left[\mathbb{E}[(\xi'_{Ij})^2 \mathbb{I}\{I \in G_k\}] - 2n^{-1} \sum_{i \in G_k} \mathbb{E}[\xi'_{ij}]\mathbb{E}[\xi_{ij}|\mathcal{A}] + n^{-1} \sum_{i \in G_k} \mathbb{E}[\xi_{ij}^2|\mathcal{A}] \right] \right\}^{\frac{1}{2}}$$
$$+ 2 \sum_{i<j} \left[\text{Var} \left[\mathbb{E}[\xi'_{Ii}\xi'_{Ij} \mathbb{I}\{I \in G_k\}] - n^{-1} \sum_{v \in G_k} \mathbb{E}[\xi'_{vi}]\mathbb{E}[\xi_{vj}|\mathcal{A}] - n^{-1} \sum_{v \in G_k} \mathbb{E}[\xi'_{vj}]\mathbb{E}[\xi_{vi}|\mathcal{A}] + n^{-1} \sum_{v \in G_k} \mathbb{E}[\xi_{vi}\xi_{vj}|\mathcal{A}] \right] \right]^{\frac{1}{2}}. \quad (30)$$

利用 $\mathbb{E}[\xi'_{ij}]=0$ 及独立性，我们得到(30)等于：

$$
\begin{aligned}
nC_2\sum_{k=1}^{K}\Big\{\sum_{j=1}^{d}\Big[\frac{1}{n^2}\mathrm{Var}\left[\sum_{i\in G_k}\mathbb{E}[\xi_{ij}^2|\mathcal{A}]\right]\Big]^{\frac{1}{2}} \\
+2\sum_{i<j}\Big[\frac{1}{n^2}\mathrm{Var}[\sum_{v\in G_k}\mathbb{E}[\xi_{vi}\xi_{vj}|\mathcal{A}]]\Big]^{\frac{1}{2}}\Big\} \\
=C_2\sum_{k=1}^{K}\Big\{\sum_{j=1}^{d}\Big[\mathrm{Var}\left[\sum_{i\in G_k}\xi_{ij}^2\right]\Big]^{\frac{1}{2}}+2\sum_{i<j}\Big[\mathrm{Var}[\sum_{v\in G_k}\xi_{vi}\xi_{vj}]\Big]^{\frac{1}{2}}\Big\} \\
=C_2\sum_{k=1}^{K}\Big\{\sum_{j=1}^{d}\Big[\sum_{i\in G_k}\mathrm{Var}\left[\xi_{ij}^2\right]\Big]^{\frac{1}{2}}+2\sum_{i<j}\Big[\sum_{v\in G_k}\mathrm{Var}[\xi_{vi}\xi_{vj}]\Big]^{\frac{1}{2}}\Big\} \\
=\frac{q^2c^2\left\|h\right\|_2}{4}K_2(\theta_0) \qquad (31)
\end{aligned}
$$

其中 $K_2(\theta_0)$ 在定理陈述中定义。接下来，注意到 28 式等于

$$
\begin{aligned}
&nC_3\sum_{k=1}^{K}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{u=1}^{d}\mathbb{E}\left|(\xi'_{Ii}-\xi_{Ii})(\xi'_{Ij}-\xi_{Ij})(\xi'_{Iu}-\xi_{Iu})\right| \\
&=nC_3\sum_{k=1}^{K}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{u=1}^{d}\frac{1}{n}\sum_{v\in G_k}\mathbb{E}\left|(\xi'_{vi}-\xi_{vi})(\xi'_{vj}-\xi_{vj})(\xi'_{vu}-\xi_{vu})\right| \\
&=C_3\sum_{k=1}^{K}\sum_{v\in G_k}\mathbb{E}\Big[\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{u=1}^{d}\left|(\xi'_{vi}-\xi_{vi})\right|\left|(\xi'_{vj}-\xi_{vj})\right|\left|(\xi'_{vu}-\xi_{vu})\right|\Big] \\
&=C_3\sum_{k=1}^{K}\sum_{v\in G_k}\mathbb{E}\Big[\sum_{j=1}^{d}\left|(\xi'_{vj}-\xi_{vj})\right|\Big]^3 \\
&=C_3\sum_{i=1}^{n}\mathbb{E}\Big[\sum_{j=1}^{d}\left|(\xi'_{ij}-\xi_{ij})\right|\Big]^3 \\
&=\frac{q^3c^3\left\|h\right\|_3}{12}K_3(\theta_0)
\end{aligned}
$$

其中， $K_3(\theta_0)$ 的定义同样见上文定理陈述。因此，我们已将主要三角不等式（20）中的第一项上界估计为

$$
\frac{q^2c^2\left\|h\right\|_2}{4}K_2(\theta_0)+\frac{q^3c^3\left\|h\right\|_3}{12}K_3(\theta_0). \qquad (32)
$$

接下来，我们将对主三角不等式（21）中的剩余项取上界。

（21）的上界。定义

$$S := [S(1,\theta_0)^\intercal, S(2,\theta_0)^\intercal, \ldots, S(K,\theta_0)^\intercal]^\intercal.$$

对.....进行二阶泰勒展开
$[S(1,\theta)^\intercal, S(2,\theta)^\intercal, \ldots, S(K,\theta)^\intercal]_j^\intercal$，在 θ_0^K处取值为$\hat{}\theta^K$，得到：

$$[S(1,\hat{\theta}_1), S(2,\hat{\theta}_2), \ldots, S(K,\hat{\theta}_K)]_j^\intercal = S_j + \sum_{i=1}^{q}[\hat{\theta}^K - \theta_0^K]_i\Big(\frac{\partial}{\partial\theta_i}S_j\Big) +$$
$$\frac{1}{2}\sum_{i=1}^{q}\sum_{u=1}^{q}[\hat{\theta}^K-\theta_0^K]_i[\hat{\theta}^K-\theta_0^K]_u\Big(\frac{\partial^2}{\partial\theta_i\partial\theta_u}[S(1,\theta),S(2,\theta),\ldots,S(K,\theta)]_j^\intercal\Big|_{\theta=\theta_0^*}\Big). \tag{33}$$

我们向读者说明，此处对符号稍作滥用：为简化本节中的表达式，我们用$\frac{\partial}{\partial\theta_i}$ 表示$\partial\ ^{\partial}_{\theta^K()_i}$，其中 θ^K $=[\theta_1,\ldots,\theta_q]^\intercal$；对于固定的 k，$S(k,\theta)$是 $\theta_{d(k-1)+1},\ldots,\theta_{dk}$的函数。例如，
$\frac{\partial}{\partial\theta_i}S_j = 0 \quad j \in \{d\ k_1 - 1\ +1\ \ldots\ dk_1\} \qquad i \in \{d\ k_2 - 1\ +$若$(),,$且$()$
$1\ \ldots$
$,,dk_2\}$ 对于 $k_1 \neq k_2$。

$^-\qquad\qquad^-\qquad\qquad\qquad

简单的分块矩阵乘法。然后，

$$\begin{aligned}
&\tilde{J}_n = J_n^* J_n J_n^* \\
\Longrightarrow\ &\tilde{J}_n^{-1} = J_n^{*-1} J_n^{-1} J_n^{*-1} \\
\Longrightarrow\ &J_n^{-1} = J_n^* \tilde{J}_n^{-1} J_n^* \\
\Longrightarrow\ &J_n^{-\frac{1}{2}} = \tilde{J}_n^{-\frac{1}{2}} J_n^* \\
\Longrightarrow\ &\tilde{J}_n^{-\frac{1}{2}} = J_n^{-\frac{1}{2}} J_n^{*-1}.
\end{aligned}$$

因此，将（34）两边乘以 $J_n^{\frac{1}{2}}/\sqrt{n}$ 可得

$$\sqrt{n} J_n^{-\frac{1}{2}} (\hat{\theta}^K - \theta_0^K) = \frac{\tilde{J}_n^{-\frac{1}{2}}}{\sqrt{n}} \Big\{ S + \sum_{i=1}^{q} [\hat{\theta}^K - \theta_0^K]_i \Big(\frac{\partial}{\partial \theta_i} S + n [J_n^*]_{*i} \Big) + \frac{1}{2} \sum_{i=1}^{q} \sum_{u=1}^{q} [\hat{\theta}^K - \theta_0^K]_i [\hat{\theta}^K - \theta_0^K]_u \Big(\frac{\partial^2}{\partial \theta_i \partial \theta_u} [S(1,\theta), S(2,\theta), \ldots, S(K,\theta)]^\intercal \Big|_{\theta=\theta_0^*} \Big) \Big\}. \tag{35}$$

回忆 $Q_i := [\hat{\theta}^K - \theta_0^K]_i$，并定义

$$R := \frac{\tilde{J}_n^{-\frac{1}{2}}}{2\sqrt{n}} \sum_{i=1}^{q} \sum_{u=1}^{q} Q_i Q_u \Big(\frac{\partial^2}{\partial \theta_i \partial \theta_u} [S(1,\theta), S(2,\theta), \ldots, S(K,\theta)]^\intercal \Big|_{\theta=\theta_0^*} \Big).$$

注意， R 是（35）右端项的最后一项。此外， R与定理 2.1 中出现的余项矩阵无关，因此在当前证明中不会用到。令

$$\begin{aligned}
T_1 &:= h\big(\sqrt{n} J_n^{-\frac{1}{2}} (\hat{\theta}^K - \theta_0^K)\big) - h\big(\frac{1}{\sqrt{n}} \tilde{J}_n^{-\frac{1}{2}} S + R\big) \\
T_2 &:= h\big(\frac{1}{\sqrt{n}} \tilde{J}_n^{-\frac{1}{2}} S + R\big) - h\big(\frac{1}{\sqrt{n}} \tilde{J}_n^{-\frac{1}{2}} S\big).
\end{aligned}$$

那么（21）等于

$$|\mathbb{E} T_1 + \mathbb{E} T_2| \le \mathbb{E} |T_1| + \mathbb{E} |T_2|.$$

我们首先讨论 T_1的一个上界。使用在 $\sqrt{n} J_n^{-\frac{1}{2}} (\hat{\theta}^K - \theta_0^K)$处求值并以 $n^{-\frac{1}{2}} \tilde{J}_n^{-\frac{1}{2}} S + R$为中心的 h一阶多元泰勒级数近似，可得，

$$h\big(\sqrt{n} J_n^{-\frac{1}{2}} (\hat{\theta}^K - \theta_0^K)\big) = h(n^{-\frac{1}{2}} \tilde{J}_n^{-\frac{1}{2}} S + R) + \sum_{j=1}^{q} \frac{\partial}{\partial \theta_j} h(x^*) \Big(\sqrt{n} J_n^{-\frac{1}{2}} (\hat{\theta}^K - \theta_0^K) - n^{-\frac{1}{2}} \tilde{J}_n^{-\frac{1}{2}} S - R \Big)_{j*},$$

其中，在本例中， $j*$ 表示第 j^{th} 行。化简得

$$|T_1| \leq \|h\|_1 \sum_{j=1}^{q} \left| \sqrt{n}[J_n^{-\frac{1}{2}}]_{j*}(\hat{\theta}^K - \theta_0^K) - n^{-\frac{1}{2}}[\tilde{J}_n^{-\frac{1}{2}}]_{j*}S - R_{j*} \right|.$$

接下来我们将用 $\sqrt{n}[J_n^{-\frac{1}{2}}]_{j*}(\hat{\theta}^K - \theta_0^K)$ 表示为 $n^{-\frac{1}{2}}[\tilde{J}_n^{-\frac{1}{2}}]_{j*}S$和两个余项。逐分量使用（35），并将其代入上述关于 $\sqrt{n}[J_n^{-\frac{1}{2}}]_{j*}(\hat{\theta}^K - \theta_0^K)$ 的表达式，得到

$$\begin{aligned}
|T_1| &\leq \|h\|_1 \sum_{j=1}^{q} \left| n^{-\frac{1}{2}}[\tilde{J}_n^{-\frac{1}{2}}]_{j*} \sum_{i=1}^{q} [\hat{\theta}^K - \theta_0^K]_i \Big(\frac{\partial}{\partial \theta_i} S + n[J_n^*]_{*i} \Big) \right| \\
&\leq \frac{\|h\|_1}{\sqrt{n}} \sum_{j=1}^{q} \sum_{l=1}^{q} \left| [\tilde{J}_n^{-\frac{1}{2}}]_{jl} \right| \sum_{i=1}^{q} \left| [\hat{\theta}^K - \theta_0^K]_i \Big([\frac{\partial}{\partial \theta_i} S]_l + n[J_n^*]_{li} \Big) \right|.
\end{aligned}$$

取期望并应用柯西-施瓦茨不等式，我们得到

$$\mathbb{E}\,|T_1| \leq \frac{\|h\|_1}{\sqrt{n}} \sum_{j=1}^{q} \sum_{l=1}^{q} \left| [\tilde{J}_n^{-\frac{1}{2}}]_{jl} \right| \sum_{i=1}^{q} \sqrt{\mathbb{E}[[\hat{\theta}^K - \theta_0^K]_i]^2 \mathbb{E}\Big([\frac{\partial}{\partial \theta_i} S]_l + n[J_n^*]_{li} \Big)^2}.$$

将矩阵写成分块形式并去除零项后，上式变为

$$\begin{aligned}
\mathbb{E}\,|T_1| \leq \frac{\|h\|_1}{\sqrt{n}} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \sum_{j=1}^{d} \sum_{l=1}^{d} \left| [[\tilde{J}_n^{-\frac{1}{2}}]_{[k_2][k_1]}]_{jl} \right| \Big\{ \\
\sum_{i=1}^{d} \sqrt{\mathbb{E}[[\hat{\theta}_{k_1} - \theta_0]_i]^2 \mathbb{E}\Big(\frac{\partial^2}{\partial \theta_i \partial \theta_l} \ell_{k_1}(\theta_0, Y) + n\bar{I}_n(k_1, \theta_0)_{li} \Big)^2} \Big\}. \quad (36)
\end{aligned}$$

现在我们讨论 T_2 的上界。根据全概率公式并利用 $Q_{(m)} = \max_{i \in \{1,\ldots,q\}} Q_i$，我们看到

$$\begin{aligned}
\mathbb{E}\,|T_2| &= \mathbb{E}[|T_2| \mid \left|Q_{(m)}\right| \geq \epsilon] \mathbb{P}(\left|Q_{(m)}\right| \geq \epsilon) + \mathbb{E}[|T_2| \mid \left|Q_{(m)}\right| < \epsilon] \mathbb{P}(\left|Q_{(m)}\right| < \epsilon) \\
&\leq 2\,\|h\|\, \mathbb{P}(\left|Q_{(m)}\right| \geq \epsilon) + \mathbb{E}[|T_2| \mid \left|Q_{(m)}\right| < \epsilon] \\
&\leq \frac{2\,\|h\|}{\epsilon^2} \mathbb{E} Q_{(m)}^2 + \mathbb{E}[|T_2| \mid \left|Q_{(m)}\right| < \epsilon] \\
&\leq \frac{2\,\|h\|}{\epsilon^2} \mathbb{E}\big[\sum_{j=1}^{q} Q_j^2 \big] + \mathbb{E}[|T_2| \mid \left|Q_{(m)}\right| < \epsilon], \quad (37)
\end{aligned}$$

其中第三行由马尔可夫不等式得出。为了对 $\mathbb{E}[|T_2| \mid \left|Q_{(m)}\right| < \epsilon]$ 求界，我们再次使用在 h 处求值的一阶泰勒展开

$(n\tilde{J}_n)^{-\frac{1}{2}}S+R$，并以$(n\tilde{J}_n)^{-\frac{1}{2}}S$为中心，得到：

$$
\begin{aligned}
&|T_2| \leq \|h\|_1 \sum_{j=1}^{q} |R_{j*}| \\
&= \frac{\|h\|_1}{2\sqrt{n}} \sum_{l=1}^{q}\sum_{j=1}^{q} \Big|[\tilde{J}_n^{-\frac{1}{2}}]_{lj}\Big| \sum_{i=1}^{q}\sum_{u=1}^{q} Q_i Q_u \Big(\frac{\partial^2}{\partial\theta_i \partial\theta_u} [S(1,\theta),\ldots,S(K,\theta)]_j^{\intercal}\Big|_{\theta=\theta_0^*} \Big).
\end{aligned} \tag{38}
$$

结合（37）和（38），可得

$$
\begin{aligned}
\mathbb{E}\,|T_2| = \frac{2\,\|h\|}{\epsilon^2}\mathbb{E}\big[\sum_{j=1}^{q} Q_j^2\big] + \frac{\|h\|_1}{2\sqrt{n}} \sum_{l=1}^{q}\sum_{j=1}^{q} \Big|[\tilde{J}_n^{-\frac{1}{2}}]_{lj}\Big| \sum_{i=1}^{q}\sum_{u=1}^{q} \Big\{ \\
\mathbb{E}\Big[Q_i Q_u \Big(\frac{\partial^2}{\partial\theta_i \partial\theta_u} [S(1,\theta),\ldots,S(K,\theta)]_j^{\intercal}\Big|_{\theta=\theta_0^*} \Big) \mid |Q_{(m)}| < \epsilon \Big] \Big\}.
\end{aligned}
$$

利用柯西-施瓦茨不等式以及 Anastasiou （2018年，引理 4.1），

$$
\begin{aligned}
\mathbb{E}\,|T_2| \leq \frac{2\,\|h\|}{\epsilon^2}\mathbb{E}\big[\sum_{j=1}^{q} Q_j^2\big] + \frac{\|h\|_1}{2\sqrt{n}} \sum_{l=1}^{q}\sum_{j=1}^{q} \Big|[\tilde{J}_n^{-\frac{1}{2}}]_{lj}\Big| \sum_{i=1}^{q}\sum_{u=1}^{q} \Big\{ \\
(\mathbb{E}[Q_i^2 Q_u^2])^{\frac{1}{2}} (\mathbb{E}[\Big(\frac{\partial^2}{\partial\theta_i \partial\theta_u} [S(1,\theta),\ldots,S(K,\theta)]_j^{\intercal}\Big|_{\theta=\theta^\dagger} \Big)^2 \mid |Q_{(m)}| < \epsilon])^{\frac{1}{2}} \Big\}.
\end{aligned}
$$

现在将矩阵写成分块形式，上式变为

$$
\begin{aligned}
\mathbb{E}\,|T_2| \leq \frac{2\,\|h\|}{\epsilon^2}\mathbb{E}\big[\sum_{j=1}^{q} Q_j^2\big] + \frac{\|h\|_1}{2\sqrt{n}} \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{l=1}^{d}\sum_{j=1}^{d} \Big|[[\tilde{J}_n^{-\frac{1}{2}}]_{[k_2][k_1]}]_{lj}\Big| \Big\{ \\
\sum_{i=1}^{q}\sum_{u=1}^{q} (\mathbb{E}[Q_i^2 Q_u^2])^{\frac{1}{2}} (\mathbb{E}[\Big(\frac{\partial^2}{\partial\theta_i \partial\theta_u} S(k_1;\theta)_j\Big|_{\theta=\theta^\dagger} \Big)^2 \mid |Q_{(m)}| < \epsilon])^{\frac{1}{2}} \Big\}.
\end{aligned}
$$

我们注意到，除非 θ_i 和 θ_u 均为

对应于 $S(k_1,\theta)$。因此，化简可得，

$$\begin{aligned}\mathbb{E}\,|T_2| \le \frac{2\,\|h\|}{\epsilon^2}\mathbb{E}\big[\sum_{j=1}^{q} Q_j^2\big] + \frac{\|h\|_1}{2\sqrt{n}}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{l=1}^{d}\sum_{j=1}^{d}\Big|[[\tilde{J}_n^{-\frac{1}{2}}]_{[k_2][k_1]}]_{lj}\Big|\Big\{ \\ \sum_{i=1}^{d}\sum_{u=1}^{d}\big(\mathbb{E}\big\{([\hat{\theta}_{k_1}-\theta_0]_i)^2([\hat{\theta}_{k_1}-\theta_0]_u)^2\big\}\big)^{\frac{1}{2}}\cdot \\ (\mathbb{E}[\big(\frac{\partial^3}{\partial\theta_i\partial\theta_u\partial\theta_j}\ell_{k_1}(\theta,Y)\Big|_{\theta=\theta_0^*}\big)^2 \mid |Q_{(m)}|<\epsilon])^{\frac{1}{2}}\Big\} \\ \le \frac{2\,\|h\|}{\epsilon^2}\mathbb{E}\big[\sum_{j=1}^{q} Q_j^2\big] + \frac{\|h\|_1}{2\sqrt{n}}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{l=1}^{d}\sum_{j=1}^{d}\Big|[[\tilde{J}_n^{-\frac{1}{2}}]_{[k_2][k_1]}]_{lj}\Big|\Big\{ \\ \sum_{i=1}^{d}\sum_{u=1}^{d}\big(\mathbb{E}\big\{([\hat{\theta}_{k_1}-\theta_0]_i)^2([\hat{\theta}_{k_1}-\theta_0]_u)^2\big\}\big)^{\frac{1}{2}}(\mathbb{E}[\big(M_{iuj}^{k_1}(Y)\big)^2 \mid |Q_{(m)}|<\epsilon])^{\frac{1}{2}}\Big\}.\end{aligned} \tag{39}$$

如上所示 $\tilde{J}_n^{-\frac{1}{2}}A\Sigma^{\frac{1}{2}} = 1_{q\times q}$，因此 $\tilde{J}_n^{-\frac{1}{2}} = \Sigma^{-\frac{1}{2}}A^{-1}$。对 A 求逆并进行简单的分块矩阵乘法，可得

$$[\tilde{J}_n^{-\frac{1}{2}}]_{[k_2][k_1]} = \begin{cases} \bar{I}_n^{-\frac{1}{2}}(G_{k_2};\theta_0) & k_2 = k_1 \\ -\bar{I}_n^{-\frac{1}{2}}(G_{k_2};\theta_0) & k_2 = k_1+1 \\ 0_{d\times d} & \text{otherwise}\ . \end{cases} \tag{40}$$

因此，(39) 可化简为

$$\begin{aligned}\mathbb{E}\,|T_2| \le \frac{2\,\|h\|}{\epsilon^2}\mathbb{E}\big[\sum_{j=1}^{q} Q_j^2\big] + \frac{\|h\|_1}{2\sqrt{n}}\sum_{k_1=1}^{K}\sum_{k_2=k_1}^{\min\{k_1+1,K\}}\sum_{l=1}^{d}\sum_{j=1}^{d}\Big|\bar{I}_n^{-\frac{1}{2}}(G_{k_2};\theta_0)_{lj}\Big|\Big\{ \\ \sum_{i=1}^{d}\sum_{u=1}^{d}\big(\mathbb{E}\big\{([\hat{\theta}_{k_1}-\theta_0]_i)^2([\hat{\theta}_{k_1}-\theta_0]_u)^2\big\}\big)^{\frac{1}{2}}(\mathbb{E}[\big(M_{iuj}^{k_1}(Y)\big)^2 \mid |Q_{(m)}|<\epsilon])^{\frac{1}{2}}\Big\}.\end{aligned} \tag{41}$$

因此，对 J_n 在 (36) 中作类似化简，并结合 (41)，可得 (21) 的上界为

$$\frac{2\,\|h\|}{\epsilon^2}\mathbb{E}\big[\sum_{j=1}^{q} Q_j^2\big] + \frac{\|h\|_1}{\sqrt{n}}K_1(\theta_0). \tag{42}$$

结合（42）和（32）即可得到（5）中的所需界。 □

在下面的引理中，我们使用多重指标记号。 n 元非负整数构成的多重指标

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \qquad (\alpha_j \in \{0, 1, 2, \ldots\}).$$

若 α 是多重指标，则，

$$|\alpha| = \alpha_1 + \alpha_2 + \ldots \alpha_n, \qquad \alpha! = \alpha_1!\alpha_2! \ldots \alpha_n!$$
$$\boldsymbol{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} \quad (\boldsymbol{x} \in \mathbb{R}^n),$$
$$\partial^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_n^{\alpha_n}}.$$

B 推论的证明 4.1

为证明该推论，我们根据定理 3.1求解 $K_1(\eta)$、 $K_2(\eta)$ 和 $K_3(\eta)$。由于观测值独立同分布，信息矩阵可简化为

$$\bar{I}_n(k;\eta) = \frac{n_k}{n} I(\eta), \quad \bar{I}(k;\eta) = I(\eta) \sum_{i=1}^{k} \gamma_k. \tag{43}$$

此外，利用（9），

$$I(\eta) = \mathbb{E}\left[-\nabla S^\intercal(Y;\eta)\right] = \mathbb{E}\left[\nabla_\eta^2 A(\eta)\right] = \nabla_\eta^2 A(\eta) = \mathrm{Var}_\eta\left[T(Y)\right] \tag{44}$$

将这些结果代入（43）可得

$$\bar{I}_n(k;\eta) = \frac{n_k}{n} \mathrm{Var}_\eta\left[T(Y)\right], \quad \bar{I}(k;\eta) = \mathrm{Var}_\eta\left[T(Y)\right] \sum_{i=1}^{k} \gamma_k \tag{45}$$

$$\bar{I}_n(G_k;\eta) = \frac{|G_k|}{n} \mathrm{Var}_\eta\left[T(Y)\right], \quad \bar{I}(k;\eta) = \gamma_k \mathrm{Var}_\eta\left[T(Y)\right]. \tag{46}$$

参见（5），我们先对以下表达式进行化简

$$K_1(\eta_0) = \sum_{k_1=1}^{K} \sum_{k_2=k_1}^{\min\{k_1+1,K\}} \sum_{l=1}^{d}\sum_{j=1}^{d} \left|\bar{I}_n^{-\frac{1}{2}}(G_{k_2};\eta_0)_{lj}\right| \Bigg\{ \Bigg($$
$$\sum_{i=1}^{d} \sqrt{\mathbb{E}[[\hat{\eta}_{[k_1]} - \eta_0]_i]^2 \mathbb{E}\Big(\frac{\partial^2}{\partial\eta_i\partial\eta_l}\ell_{k_1}(\eta_0, Y) + n\bar{I}_n(k_1,\eta_0)_{li}\Big)^2}\Bigg)$$
$$+ \frac{1}{2}\sum_{i=1}^{d}\sum_{u=1}^{d} \left(\mathbb{E}\left\{([\hat{\eta}_{[k_1]} - \eta_0]_i)^2([\hat{\eta}_{[k_1]} - \eta_0]_u)^2\right\}\right)^{\frac{1}{2}} \left(\mathbb{E}[\left(M_{iul}^{k_1}(Y)\right)^2 \mid \left|Q_{(m)}\right| < \epsilon]\right)^{\frac{1}{2}} \Bigg\}.$$

通过使用（45）并对数似然求导，

$$\frac{\partial^2}{\partial\eta_i\partial\eta_l}\ell_{k_1}(\eta_0, Y) = -n_k \frac{\partial^2}{\partial\eta_i\partial\eta_l} A(\eta) = -n\bar{I}_n(k_1,\eta_0)_{li}.$$

因此，

$$\mathbb{E}\Big(\frac{\partial^2}{\partial\eta_i\partial\eta_l}\ell_{k_1}(\eta_0,Y)+n\bar{I}_n(k_1,\eta_0)_{li}\Big)^2=0 \tag{47}$$

因此，以 i 为索引的 K_1 中的第一项求和等于 0。接下来只需化简 $\mathbb{E}\{([\hat{\eta}_{[k_1]}-\eta_0]_i)^2([\hat{\eta}_{[k_1]}-\eta_0]_u)^2\}$ 和 $\mathbb{E}[\big(M_{iul}^{k_1}(Y)\big)^2 \mid |Q_{(m)}|<\epsilon]$。根据指数族最大似然估计量的一般形式（12），

$$[\hat{\eta}_{[k_1]}-\eta_0]_i=\tau^{-1}\left(\frac{1}{n_{k_1}}\sum_{s=1}^{n_{k_1}}T(y_s)\right)_i-\eta_{0,i}.$$

因此，

$$\mathbb{E}\big\{([\hat{\eta}_{[k_1]}-\eta_0]_i)^2([\hat{\eta}_{[k_1]}-\eta_0]_u)^2\big\}=\mathbb{E}\Bigg\{\left(\tau^{-1}\left(\frac{1}{n_{k_1}}\sum_{s=1}^{n_{k_1}}T(y_s)\right)_i-\eta_{0,i}\right)^2\times \left(\tau^{-1}\left(\frac{1}{n_{k_1}}\sum_{s=1}^{n_{k_1}}T(y_s)\right)_u-\eta_{0,u}\right)^2\Bigg\} \tag{48}$$

由于不知道充分统计量 $T(y)$，我们无法进一步化简这一项。接下来，我们转而处理涉及有界函数$M_{iul}^{k_1}(Y)$的项。

对数似然取三阶偏导数，并回想可知， (10)

$$\begin{aligned}\left|\frac{\partial^3}{\partial\eta_i\partial\eta_u\partial\eta_l}\ell(\eta,y^{n_{k_1}})\right|&=n_{k_1}\left|\frac{\partial^3}{\partial\eta_i\partial\eta_u\partial\eta_l}A(\eta)\right|\\&=n_{k_1}\left|(\mu_{ilk}-\mu_{il}\mu_k-\mu_{ik}\mu_l-\mu_{lk}\mu_i+2\mu_i\mu_l\mu_k)\right|.\end{aligned}$$

注意，上述方程的右端项仅是 η 的函数，而不是观测值 Y_1、……、Y_{n_k} 的函数。于是，

$$(\mathbb{E}[\big(M_{iul}^{k_1}(Y)\big)^2 \mid |Q_{(m)}|<\epsilon])^{\frac{1}{2}}=n_{k_1}\mu_{iul}^{\epsilon}, \tag{49}$$

后者由（11）给出。结合（47）、（48）、（49）和（46），我们得到，

$$\begin{aligned}K_1(\eta_0)=&\frac{1}{2}\sum_{k_1=1}^{K}\sum_{k_2=k_1}^{\min\{k_1+1,K\}}n_{k_1}\left(\frac{|G_{k_2}|}{n}\right)^{-\frac{1}{2}}\sum_{l=1}^{d}\sum_{j=1}^{d}\mathrm{Var}^{-\frac{1}{2}}[T(Y)]_{lj}\Bigg\{\\&\sum_{i=1}^{d}\sum_{u=1}^{d}\mu_{iul}^{\epsilon}\Bigg(\mathbb{E}\Bigg[\left(\tau^{-1}\left(\frac{1}{n_{k_1}}\sum_{s=1}^{n_{k_1}}T(y_s)\right)_i-\eta_{0,i}\right)^2\times\\&\left(\tau^{-1}\left(\frac{1}{n_{k_1}}\sum_{s=1}^{n_{k_1}}T(y_s)\right)_u-\eta_{0,u}\right)^2\Bigg]\Bigg)^{\frac{1}{2}}\Bigg\}.\end{aligned} \tag{50}$$

现在我们考虑 K_2（η）在（5）中的情形。回顾可知，

$$K_2(\eta) = \sum_{k=1}^{K} \Big\{ \sum_{j=1}^{d} \Big[\sum_{i \in G_k} \mathrm{Var}\left[\xi_{ij}^2\right] \Big]^{\frac{1}{2}} + 2 \sum_{i<j} \Big[\sum_{v \in G_k} \mathrm{Var}[\xi_{vi}\xi_{vj}] \Big]^{\frac{1}{2}} \Big\}.$$

经过熟悉的计算可得

$$\begin{aligned} \mathrm{Var}\left[\xi_{ij}^2\right] &= n^{-2}\mathrm{Var}\left[(T_j(Y) - \mathbb{E}_{\eta_0}\left[T_j(Y)\right])^2\right] \\ \mathrm{Var}\left[\xi_{vi}\xi_{vj}\right] &= n^{-2}\mathrm{Var}\left[(T_i(Y) - \mathbb{E}_{\eta_0}\left[T_i(Y)\right])(T_j(Y) - \mathbb{E}_{\eta_0}\left[T_j(Y)\right])\right]. \end{aligned}$$

因此，将上述形式代入 $K_2(\eta)$ 可得

$$\begin{aligned} K_2(\eta) = \frac{1}{\sqrt{n}} \sum_{k=1}^{K} \left(\frac{|G_k|}{n} \right)^{\frac{1}{2}} \Big\{ \sum_{j=1}^{d} \Big[\mathrm{Var}\left[(T_j(Y) - \mu_j(\eta_0))^2\right] \Big]^{\frac{1}{2}} + \\ 2 \sum_{i<j} \Big[\mathrm{Var}\left[(T_i(Y) - \mu_i(\eta_0))(T_j(Y) - \mu_j(\eta_0))\right] \Big]^{\frac{1}{2}} \Big\}. \end{aligned} \tag{51}$$

最后，我们来处理 $K_3(\eta)$，

$$\sum_{i=1}^{n} \mathbb{E}\Big[\sum_{j=1}^{d} \left| (\xi'_{ij} - \xi_{ij}) \right| \Big]^3.$$

与之前一样，我们可以看到

$$\begin{aligned} \left|\xi'_{ij} - \xi_{ij}\right| &= n^{-\frac{1}{2}} \left| (T_j(Y') - \mu_j(\eta_0)) - (T_j(Y) - \mu_j(\eta_0)) \right| \\ &= \left| T_j(Y') - T_j(Y) \right|. \end{aligned}$$

因此， $K_3(\eta_0)$ 等于

$$\frac{1}{\sqrt{n}} \mathbb{E}\Big[\sum_{j=1}^{d} \left| T_j(Y') - T_j(Y) \right| \Big]^3. \tag{52}$$

结合式（50）、（51）和（52）即可得到推论。