单位代码	10475
学　　号	104753170654
分 类 号	O212.1

河南大學

硕 士 学 位 论 文

高维数据均值和协差阵检验的经验似然方法

Empirical Likelihood Method for Testing Mean and Covariance Matrix of High-dimensional Data

学科、专业：统计学

研 究 方 向：数理统计

申请学位类别：理学硕士

申　请　人：马莹莹

指 导 教 师：解俊山 副教授

二〇二〇 年 六 月

Empirical Likelihood Method for Testing Mean and Covariance Matrix of High-dimensional Data

A Dissertation Submitted to

the Graduate School of Henan University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

By

Ma Yingying

Supervisor: Associate Professor Xie Junshan

Date: June, 2020

关于学位论文独创声明和学术诚信承诺

本人向河南大学提出硕士学位申请。本人郑重声明：所呈交的学位论文是本人在导师的指导下独立完成的，对所研究的课题有新的见解。据我所知，除文中特别加以说明、标注和致谢的地方外，论文中不包括其他人已经发表或撰写过的研究成果，也不包括其他人为获得任何教育、科研机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

在此本人郑重承诺：所呈交的学位论文不存在舞弊作伪行为，文责自负。

学位申请人（学位论文作者）签名：

2020年 6月 19日

关于学位论文著作权使用授权书

本人经河南大学审核批准授予硕士学位。作为学位论文的作者，本人完全了解并同意河南大学有关保留、使用学位论文的要求，即河南大学有权向国家图书馆、科研信息机构、数据收集机构和本校图书馆等提供学位论文（纸质文本和电子文本）以供公众检索、查阅。本人授权河南大学出于宣扬、展览学校学术发展和进行学术交流等目的，可以采取影印、缩印、扫描和拷贝等复制手段保存、汇编学位论文（纸质文本和电子文本）。

（涉及保密内容的学位论文在解密后适用本授权书）

学位获得者（学位论文作者）签名：

2020年 6月 9日

学位论文指导教师签名：

2020年 6月 19日

摘 要

随着大数据时代的到来,需要处理的数据量及数据的维数也在不断扩大,即统计文献中所讨论的"大 p、大 n"现象.然而,经典的统计分析方法对于"大 p、大 n"类高维数据的研究已不再适用,此类高维数据的统计分析方法研究是一个十分有意义的课题.

本文主要利用非参经验似然的方法和统计渐近理论方法及工具,研究有克罗内克积相关结构的高维矩阵型数据协差阵的假设检验,以及固定维两样本均值线性组合的检验问题.

对于矩阵型数据协差阵的恒等检验问题, Touloumis et al.[4] 借鉴 Chen et al.[6] 提出的 U 统计量,在考虑列独立的条件下,对假定非参模型进行球形和恒等假设检验,得出检验统计量是渐近正态的.本文针对 Touloumis et al.[4] 提出的模型,考虑通过对随机矩阵进行拉直处理,构造估计方程,用经验似然方法来对这一特殊数据的协方差矩阵进行检验,得出似然比统计量渐近服从卡方分布.并通过数值模拟得出该方法是有效的.

对于两样本均值线性组合的检验问题, Li et al.[31] 通过构造 U 统计量的方法对高维数据均值向量的线性组合进行研究,得出统计量是渐近正态的.本文采用四种不同的经验似然方法重新考虑这个问题,在一些约束条件下,得出似然比统计量均渐近服从卡方分布.同时还运用自助法对这四种方法在进行两样本均值线性组合的检验方面进行了模拟分析,对各种经验似然方法的优劣进行了分析.

关键词: 高维数据, 经验似然, 似然比统计量, 卡方分布

ABSTRACT

With the arrival of big data era, both the size and the dimension of the data become larger and larger, it is the phenomenon of "large p, large n" discussed in the statistical literature. However, the classical statistical analysis method is no longer applicable on studying the "large p, large n" type data. Therefore, it is a meaningful work to seek some appropriate and effective statistical analysis methods to deal with high-dimensional data.

In the paper, we mainly adopt nonparametric empirical likelihood method and statistical asymptotic theory to study the hypothesis test for the covariance matrix in high-dimensional transposable matrix-valued data with kronecker product dependence structure. At the same time, we will also consider the test on the linear combination of the mean value vectors of two population with fixed dimension.

For the identity test on the covariance matrix of the matrix-valued data, under the column independence assumption, Touloumis et al.[4] adopt the U statistics method proposed by Chen et al.[6] to consider the sphericity and identity test on the assumed nonparametric model, they proved the test statistics obey normal distribution asymptotically. In the paper, we will construct the estimation equation by straightening the random matrix, and use the empirical likelihood method to test the covariance matrix of this matrix-valued data, it is proved that likelihood ratio statistic asymptotical Chi-square distributed. Some numerical simulation show the efficiency of the proposed method.

For the test problem on the linear combination of two sample mean value vectors, a remarkable work due to Li et al.[31], who study the linear combination of mean vector of high-dimensional data by constructing U statistics. In the paper, we will adopt four types of slightly different empirical likelihood methods to revisit the test problem, it is concluded that all the likelihood ratio statistics asymptotically obey Chi-square distribution. At the same time, some simulations are conducted by the bootstrap trick, the advantages and disadvantages of each empirical likelihood method are analysed.

KEY WORDS: High-dimensional data, Empirical likelihood, Likelihood ratio statistic, Chi-square distribution

文中符号说明

tr(A)	矩阵 A 的迹
A^T	矩阵 A 的转置
$i.i.d.$	独立同分布
C	仅表示一个常数，其值在上下文中可以不同
$\\|\cdot\\|$	表示 R^p 上的 $Euclidean$ 范数
$X_n \xrightarrow{d} X$	随机变量序列 $\{X_n\}$ 依分布收敛于随机变量 X
$X_n \xrightarrow{p} X$	随机变量序列 $\{X_n\}$ 依概率收敛于随机变量 X
$a_n = o(b_n)\ (n \to \infty)$	$\lim_{n\to\infty} a_n/b_n = 0$
$a_n = O(b_n)\ (n \to \infty)$	$\lim_{n\to\infty} \sup\lvert\frac{a_n}{b_n}\rvert < \infty$

目 录

第一章 绪论

§1.1 研究背景及意义

随着科技的高速发展与社会的不断进步,高维数据在经济金融、生物医学、生产生活等各个方面的应用越来越广泛.过去传统的统计分析方法多局限于有限维数据的情形,对于维数 $p \to \infty$ 的高维数据情况已不再有效.比如,Jiang 和 Yang[12] 在对一系列高维正态总体进行似然比检验时,通过对比分析得出,经典的卡方近似方法在高维情形下不再适用.因此,针对不同类型的高维数据,采取合适有效的统计分析方法变得至关重要.

高维数据的假设检验问题目前已经是统计分析中比较普遍但却十分重要的一个问题.其中,对高维数据进行统计推断最基本的就是高维数据均值、协差阵的检验.比如,Yamada 和 Srivastava[40]、Srivastava 和 Kubokawa[33] 分别在高维正态和非正态总体下对多元方差分析进行检验;Qin[32] 在高维情况下基于独立随机样本分别对两均值向量的相等性、两协差阵的相等性以及相关问题进行了研究,建立了较为完善的理论体系.此外还有大量的文献都对高维数据均值和协差阵的假设检验进行了深入细致的研究.

对于总体均值的假设检验问题,一个经典的处理方法就是 Bai 和 Saranadasa[34] 所提出的 $Hotelling's\ T^2$ 检验方法.然而,在“大 p、小 n”的情况下,经典的 $Hotelling's\ T^2$ 检验不再有效.此时,Chen 和 Qin[13] 通过构造 U 统计量来对总体均值进行推断;之后 Peng et al.[3] 提出在维数 p 固定或者发散的情况下,将样本分为两个子样本,分别对这两个子样本构造合适的估计方程,进而对总体均值进行统计推断,得出统计量渐近服从卡方分布;当维数 $p = p_n \to \infty$ 时,Lahiri 和 Mukhopadhyay[44] 基于加权的样本均值与总体均值的假设值之间的马氏距离,施加了一个惩罚因子,证得检验统计量渐近服从正态分布;在 $p/n = c_n \to c \in (0,1)$ 情况下,Chen , Pan 和 Zhou[10] 通过对原始样本集增加两个伪观测值,来研究总体均值向量对数似然比统计量的渐近正态性;Wang et al.[8] 采用 Jacknife 经验似然法对两高维总体均值是否相等进行检验,该方法比前面几种方法的条件要弱,但数值模拟结果较好;此外,关于均值的线性组合的检验,Li et al.[31] 通过构造 U 统计量的方法解决了此类检验问题.

而对于协差阵的检验问题,Ledoit 和 Wolf[35] 研究了 p 和 n 均趋于 ∞ 时,协差阵是否等于给定矩阵这一检验问题的检验统计量的极限分布性质;Strivastava[23] 在样本量

$N=n+1$ 比维数 p 小, 且总体为多元正态分布时对协差阵进行球形检验, 得出统计量是渐近正态的; Zhang et al.[2] 在 p 固定或者发散的情况下, 通过将样本分为两个子样本, 分别构造合适的估计方程对总体协差阵进行统计推断, 得出统计量渐近服从卡方分布; Chen et al.[6] 在"大 p、小 n"的情况下对高维协差阵进行球形和恒等假设检验, 得出所构造的统计量是渐近正态的; Li 和 Chen[39] 在 $p \gg n$ 且不假定总体服从正态分布时、对两样本高维协差阵是否相等进行假设检验, 证得所构造的无偏统计量渐近服从标准正态分布; 此外, Zhang et al.[41] 将 Li 和 Chen[39] 中两样本协差阵的检验推广到多样本, 更进一步拓广了相关研究结果.

经验似然方法作为一种非参似然方法最初由 Owen[46] 提出, 用来对单个总体均值进行统计推断. 统计学家们在 Owen[1],[26],[46] 工作的基础之上, 将该方法广泛应用于统计分析的各个领域, 并证实了此方法在解决很多问题上都是非常有效的. 例如, 两样本数据的假设检验问题、数据缺失、线性回归模型、半参模型的统计推断等问题. 关于两样本问题, 除了上面所叙述的对两样本协差阵的假设检验之外, Jing[47] 证得关于两总体均值差的两样本经验似然方法是巴特利特纠偏的; Wu 和 Yan[28] 采用加权经验似然方法对两样本均值是否相等的问题进行了详细的理论证明; 同时, Wu 和 Yan[5] 还分别采取不同的经验似然方法对两样本均值差进行假设检验, 并分析对比了各个方法与经典的经验似然方法的差异及有效之处, 其中也包含了对数据缺失情况的讨论. 此外, Wang 和 Rao[18]、Wang 和 Rao[19] 分别在不同模型下对缺失数据进行了经验似然推断; Dong[21] 文章中使用经验似然方法解决了典型的 Behrens-Fisher 问题; Peng和 Schick[38] 对拟合优度检验问题采用了经验似然方法; Wang 和 Veraverbeke[42] 对半参模型进行了经验似然推断.

本文主要运用经验似然方法, 研究具有克罗内克积独立结构的高维可转置数据协差阵的假设检验, 以及固定维两样本均值线性组合的假设检验.

关于高维可转置数据, Allen 和 Tibshirani[45], Allen 和 Tibshirani[17] 分别对可转置数据进行建模, 针对不同问题研究并建立了相应的优于现有方法的统计分析算法; Touloumis et al.[4] 对高维可转置数据的行列协差阵进行球形和恒等假设检验, 得出所构造的 U 统计量是渐近正态的, 但该方法需要对某些参数进行估计, 构造统计量过程复杂, 对行协差阵或者列协差阵的迹也有条件限制, 鉴于此, 本文考虑运用经验似然方法, 构造估计方程来对总体协差阵进行推断, 并得出统计量渐近卡方分布的结果.

对于均值的线性组合问题, 现有文献中, Li et al.[31] 在高维情况下通过构造 U 统计量对 k 个总体均值向量的线性组合进行假设检验, 均值线性组合的经验似然检验还没有涉及, 因此本文考虑用几种不同的经验似然法对固定维两样本均值的线性组合进行统计分

析, 并证明了统计量是渐近服从卡方分布的.

§1.2 论文创新点

本文的主要工作有以下两个:

一、针对高维矩阵型数据协差阵的假设检验, 本文考虑用经验似然方法代替 U 统计量检验的方法, 在满足高维假设条件下, 通过对相关矩阵进行向量拉直, 构造估计方程, 定义经验似然比统计量, 得出该统计量渐近服从卡方分布. 最后, 通过经验水平和经验功效这两个指标去验证所提出的方法是合理有效的.

二、针对固定维两样本均值线性组合的假设检验, 本文针对不同情况, 采用四种不同的经验似然方法, 建立了相应的似然比统计量, 并得出似然比统计量在某些条件下渐近服从卡方分布, 最后通过自助法进行数值模拟, 验证了几种经验似然方法的有效性.

§1.3 主要内容及结构

本文的主要内容及结构如下:

第一章主要介绍了该论文所要研究问题的背景和意义, 以及本论文的创新之处.

第二章主要介绍了高维矩阵型数据协差阵的假设检验问题, 通过对 Touloumis et al.[4] 中所提出的数据模型进行向量拉直, 采用非参经验似然的方法, 构造两估计方程并运用统计极限理论, 得出对数似然比统计量渐近服从卡方分布. 最后, 通过数值模拟验证了本章所提出的经验似然方法的有效性.

第三章主要介绍了固定维两样本均值线性组合的假设检验问题, 证得针对四种经验似然方法所构造的似然比检验统计量均渐近服从卡方分布. 最后通过自助法模拟验证了各种方法的有效性.

最后一章为本文的总结与展望, 对本文的结果进行了总结, 并指出了本课题之后需努力研究的方向.

本文在各章节中的符号表达上可能稍有不同, 但分别都给出了说明, 各章节的符号之间独立存在, 所有符号的具体表达含义以各章节给出的为准.

第二章 高维矩阵型数据的检验

§2.1 模型介绍

本章主要考虑高维矩阵型数据协差阵的恒等假设检验. 假设有 r 个行变量和 c 个列变量的 $r\times c$ 矩阵型总体 X 满足非参模型

$$X=\Sigma_R^{1/2}Z\Sigma_C^{1/2}+M,$$

令 $X_1,\dots,X_n$ 为 n 个 $i.i.d.$ 的 $r\times c$ 随机矩阵样本, 因此有 $X_i=\Sigma_R^{1/2}Z_i\Sigma_C^{1/2}+M$, 其中 $M=E[X_i]$ 为 $r\times c$ 常数均值矩阵, 协差阵为 $cov[vec(X_i)]=\Omega=\Sigma_C\bigotimes\Sigma_R$, 且 Σ_R 和 Σ_C 为 X_i 的行协差阵和列协差阵, $\Sigma_m=\Sigma_m^{1/2}\Sigma_m^{1/2}$ 为正定矩阵 $(m\in\{R,C\})$, 而 $\{Z_i:i=1,\dots,n\}$ 是 $i.i.d.$ 的 $r\times c$ 随机矩阵序列.

本章将考虑的假设检验是

$$H_0:\begin{cases}\Sigma_R=I_R\\ \Sigma_C=I_C\end{cases}\Longrightarrow\Omega=I_{rc}\ \ v.s.\ H_1:not\ H_0.$$

进一步地, 对 Z_i 中每个元素施加如下限制: 随机变量 $\{Z_{iab}:a=1,\dots,r;\ b=1,\dots,c\}$ 的各阶矩满足

(1) $E[Z_{iab}]=0, E[Z_{iab}^2]=1;$

(2) $1\le E[Z_{iab}^4]<\infty, E[Z_{iab}^8]<\infty;$

(3) 对任意满足 $\sum_{v=1}^q l_v\le 8$ 的正整数 $l_1,\dots,l_q$, 伪独立条件成立. 即对 $(a_1,b_1)\neq(a_2,b_2)\neq\cdots\neq(a_q,b_q)$, 有

$$E[Z_{ia_1b_1}^{l_1}Z_{ia_2b_2}^{l_2}\dots Z_{ia_qb_q}^{l_q}]=E[Z_{ia_1b_1}^{l_1}]E[Z_{ia_2b_2}^{l_2}]\dots E[Z_{ia_qb_q}^{l_q}].$$

接下来, 考虑对随机矩阵进行向量拉直, 记

$$R_i=vec(X_i),\quad W_i=vec(Z_i),\quad i=1,\dots,n.$$

则 R_i 转化为

$$R_i=(\Sigma_C^{1/2}\bigotimes\Sigma_R^{1/2})W_i+vec(M).$$

为了后面计算证明的方便, 且由于 $H_0:\ \Sigma_R=I_r, \Sigma_C=I_c$, R_i 和 W_i 可以一致写为

$$R_i=(X_{i1},\cdots,X_{ir},X_{i,r+1},\cdots,X_{i,2r},\cdots,X_{i,(c-1)r+1},\cdots,X_{i,cr})^T$$

和

$$W_i = (Z_{i1}, \cdots, Z_{ir}, Z_{i,r+1}, \cdots, Z_{i,2r}, \cdots, Z_{i,(c-1)r+1}, \cdots, Z_{i,cr})^T.$$

下面分两种情况进行讨论:

当 M 为已知均值常数矩阵时, 定义 $Y_i = (R_i - vec(M))(R_i - vec(M))^T$, 由此得

$$\begin{aligned} E[Y_i] &= E[(\Sigma_C^{1/2} \bigotimes \Sigma_R^{1/2}) W_i W_i' (\Sigma_C^{1/2} \bigotimes \Sigma_R^{1/2})] \\ &= (\Sigma_C^{1/2} \bigotimes \Sigma_R^{1/2}) I_{rc\times rc} (\Sigma_C^{1/2} \bigotimes \Sigma_R^{1/2}) \\ &= \Sigma_C \bigotimes \Sigma_R \\ &= \Omega = cov(R_i). \end{aligned}$$

则可以构造如下两估计方程并将之应用于经验似然比检验

$$\begin{cases} \mathbf{E}[tr(Y_1 - I_{rc})(Y_2 - I_{rc})] = 0 \\ \mathbf{E}[1_{rc}^T(Y_1 + Y_2 - 2I_{rc})1_{rc}] = 0. \end{cases}$$

为了获得独立配对样本数据 (Y_1, Y_2), 考虑将样本分为两子样本, 两子样本的样本容量为 $m = [n/2]$. 即对于 $i = 1, \ldots, m$, 定义 $R_i(\Omega) = (e_i(\Omega), v_i(\Omega))^T$, 其中

$$e_i(\Omega) = tr[(Y_i - \Omega)(Y_{i+m} - \Omega)], \quad v_i(\Omega) = 1_p^T(Y_i + Y_{i+m} - 2\Omega)1_p.$$

下面基于 $R_i(\Omega)_{i=1}^m$ 给出 Ω 的经验似然比函数如下

$$L_1(\Omega) = \sup\{\prod_{i=1}^m (mp_i) : \sum_{i=1}^m p_i = 1, \sum_{i=1}^m p_i R_i(\Omega) = 0, p_1 \geq 0, \ldots, p_m \geq 0\}.$$

当 M 为未知均值常数矩阵时, 记 $\bar{R}^1 = \frac{1}{m}\sum_{i=1}^m R_i$, $\bar{R}^2 = \frac{1}{m}\sum_{i=m+1}^{2m} R_i$. 定义

$$\begin{aligned} Y_i^* &= (R_i - \bar{R}^1)(R_i - \bar{R}^1)^T, \\ Y_{m+i}^* &= (R_{m+i} - \bar{R}^2)(R_{m+i} - \bar{R}^2)^T, \ i = 1, \ldots, m. \end{aligned}$$

由此得

$$\begin{aligned} E(Y_i^*) &= E[(R_i - \bar{R}^1)(R_i - \bar{R}^1)^T] \\ &= E[(R_i - vec(M)) + (vec(M) - \bar{R}^1)][(R_i - vec(M)) + (vec(M) - \bar{R}^1)]^T \\ &= E[(R_i - vec(M))(R_i - vec(M))^T - 2(R_i - vec(M))(\bar{R}^1 - vec(M))^T \\ &+ (\bar{R}^1 - vec(M))(\bar{R}^1 - vec(M))^T] \\ &= \Omega - \frac{2}{m}\Omega + \frac{1}{m}\Omega \\ &= \frac{m-1}{m}\Omega, \end{aligned}$$

这里定义 $R_i^*(\Omega)=(e_i^*(\Omega),v_i^*(\Omega))^T$. 其中

$$e_i^*(\Omega)=tr[(Y_i-\frac{m-1}{m}\Omega)(Y_{m+i}-\frac{m-1}{m}\Omega)],\quad v_i^*(\Omega)=1_{rc}^T(Y_i+Y_{m+i}-\frac{2(m-1)}{m}\Omega)1_{rc}.$$

基于 $R_i^*(\Omega)$, 给出 Ω 的经验似然比函数

$$L_2(\Omega)=\sup\{\prod_{i=1}^m(mp_i):\sum_{i=1}^m p_i=1,\sum_{i=1}^m p_iR_i^*(\Omega)=0,p_1\geq 0,\ldots,p_m\geq 0\}.$$

接下来主要给出对数似然比统计量 $-2\log L_i(I_{rc})$ 的渐近分布, $i=1,2$.

§2.2 主要结果

本章主要给出两个定理、四个引理以及两个推论. 首先第一个定理主要通过泰勒展式及统计极限理论来证明, 当维数 p 和样本容量 n 都趋于无穷时, 在样本八阶矩独立条件下, 似然比检验统计量渐近服从卡方分布.

定理2.2.1 假设 $E[v_1(\Omega)]>0$, 且对某 $\delta>0$ 有

$$max\{E|e_1(\Omega)|^{2+\delta}/(E[e_1^2(\Omega)])^{(2+\delta)/2},E|v_1(\Omega)|^{2+\delta}/(E[v_1^2(\Omega)])^{(2+\delta)/2}\}=o(m^{(\delta+\min 2,\delta)/4}),\tag{2-1}$$

则在 H_0 的条件下, 有 $-2\log L_1(I_{rc})\xrightarrow{d}\chi_2^2$, 其中 $n\to\infty$. 此外, 如果还有

$$(tr(\Omega^2))^2=o(N^2E[e_1(\Omega)^2]),\quad (\sum_{i=1}^{rc}\sum_{j=1}^{rc}\sigma_{ij})^2=o(NE[v_1(\Omega)^2]).\tag{2-2}$$

则在 H_0 成立的条件下, $-2\log L_2(I_{rc})\xrightarrow{d}\chi_2^2$, 其中 $n\to\infty$.

引理2.2.1 在定理 2.2.1 条件 (2-1) 下, 有

$$\frac{1}{\sqrt{m}}\sum_{i=1}^m(\frac{e_i}{\sqrt{\pi_{11}}},\frac{v_i}{\sqrt{\pi_{22}}})^T\xrightarrow{d}N(0,I_2),\tag{2-3}$$

$$\frac{\sum_{i=1}^m e_i^2}{m\pi_{11}}-1\xrightarrow{p}0,\qquad \frac{\sum_{i=1}^m v_i^2}{m\pi_{22}}-1\xrightarrow{p}0,\qquad \frac{\sum_{i=1}^m e_iv_i}{m\sqrt{\pi_{11}\pi_{22}}}\xrightarrow{p}0,\tag{2-4}$$

$$\max_{1\leq i\leq m}|e_i/\sqrt{\pi_{11}}|=o_p(m^{1/2}),\qquad \max_{1\leq i\leq m}|v_i/\sqrt{\pi_{22}}|=o_p(m^{1/2}).\tag{2-5}$$

引理2.2.2 在定理 2.2.1 条件 (2-1) 和 (2-2) 下, 有

$$\frac{1}{\sqrt{m}}\sum_{i=1}^m(\frac{e_i^*}{\sqrt{\pi_{11}}},\frac{v_i^*}{\sqrt{\pi_{22}}})^T\xrightarrow{d}N(0,I_2),\tag{2-6}$$

$$\frac{\sum_{i=1}^{m} e_i^{*2}}{m\pi_{11}} - 1 \xrightarrow{p} 0, \quad \frac{\sum_{i=1}^{m} v_i^{*2}}{m\pi_{22}} - 1 \xrightarrow{p} 0, \quad \frac{\sum_{i=1}^{m} e_i^* v_i^*}{m\sqrt{\pi_{11}\pi_{22}}} \xrightarrow{p} 0, \tag{2-7}$$

$$\max_{1\le i\le m} |e_i^*/\sqrt{\pi_{11}}| = o_p(m^{1/2}), \quad \max_{1\le i\le m} |v_i^*/\sqrt{\pi_{22}}| = o_p(m^{1/2}). \tag{2-8}$$

记 $vec(M) = (\mu_1, \cdots, \mu_r, \cdots, \mu_{rc})^T$. 将 $rc \times rc$ 矩阵 Y_1 写为 $p = r^2c^2$ 维向量. $\Theta = (\theta_{ij})_{r^2c^2 \times r^2c^2}$ 是该向量的协差阵. 下面给出满足定理 2.2.1 成立的三个条件:

(A1) $\liminf_{N\to\infty} \frac{1}{p} tr\Theta^2 > 0$ 且 $\liminf_{n\to\infty} \frac{1}{p} 1_p^T \Theta 1_p > 0$;

(A2) 对某 $\delta > 0$, $\frac{1}{rc}\sum_{i=1}^{rc}\sum_{j=1}^{rc} E|(R_{1,i}-\mu_i)(R_{1,j}-\mu_j)-\sigma_{ij}|^{2+\delta} = O(1)$;

(A3) $rc = o(n^{(\delta+\min(2,\delta))/4(2+\delta)})$.

推论2.2.1 在条件 $(A1)-(A3)$ 下, 若 H_0 成立, 则 $-2\log L_1(I_{rc}) \xrightarrow{d} \chi_2^2, n\to\infty$. 此外, 对某常数 $C_0 > 0$, 若 $\max_{1\le i\le rc} \sigma_{ii} < C_0$, 则 $-2\log L_2(I_{rc}) \xrightarrow{d} \chi_2^2, n\to\infty$.

推论2.2.2 若 $\sum_{i=1}^{rc}\sum_{j=1}^{rc} \sigma_{ij} > 0$, 则在 H_0 下有 $-2\log L_j(I_{rc}) \xrightarrow{d} \chi_2^2$, $n\to\infty$, $j = 1, 2$.

引理2.2.3 在推论 2.2.2 条件下有 $Ee_1^4/(Ee_1^2)^2 = O(1)$, $Ev_1^4/(Ev_1^2)^2 = O(1)$.

引理2.2.4 在推论 2.2.1 条件下, 对任意的 $\delta > 0$, 有

$$\begin{aligned} E|e_1|^{2+\delta} &\le p^\delta \Big(\sum_{i=1}^{rc}\sum_{j=1}^{rc} E|R_{1i}R_{1j}-\sigma_{ij}|^{2+\delta}\Big)^2, \\ E|v_1|^{2+\delta} &\le 2^{4+\delta} p^{1+\delta} \sum_{i=1}^{rc}\sum_{j=1}^{rc} E|R_{1i}R_{1j}-\sigma_{ij}|^{2+\delta}. \end{aligned}$$

定理2.2.2 在推论 2.2.2 和 H_1 成立的条件下, 有

$$P\{-2\log L_j(I_{rc}) > \xi_{1-\alpha}\} = P\{\chi_{2,v}^2 > \xi_{1-\alpha}\} + o(1),\ n\to\infty,\ j = 1, 2. \tag{2-9}$$

§2.3 定理证明

不失一般性, 假定 $M = \mathbf{0}$. 记 $e_i(I_{rc}) = e_i$, $v_i(I_{rc}) = v_i$, $e_i^*(I_{rc}) = e_i^*$, $v_i^*(I_{rc}) = v_i^*$, $\pi_{11} = E[e_1^2(I_{rc})]$, $\pi_{22} = E[v_1^2(I_{rc})]$.

定理 2.2.1 的证明: 结合引理 2.2.1 和 2.2.2 证明该定理如下:

令 $\hat{e}_i=\frac{e_i}{\sqrt{\pi_{11}}}$, $\hat{v}_i=\frac{v_i}{\sqrt{\pi_{22}}}$, $\hat{R}_i=(\hat{e}_i,\hat{v}_i)^T$, $i=1,\cdots,m$, 则

$$\begin{aligned}
-2\log L_1(I_{rc}) &= 2\sum_{i=1}^m \log(1+\rho^T\hat{R}_i)\\
&= 2\sum_{i=1}^m [\rho^T\hat{R}_i-\frac{1}{2}(\rho^T\hat{R}_i)^2+o((\rho^T\hat{R}_i)^2)]\\
&= 2\sum_{i=1}^m \rho^T\hat{R}_i-\sum_{i=1}^m(\rho^T\hat{R}_i)^2(1-2o(1))\\
&= 2\sum_{i=1}^m \rho^T\hat{R}_i-(1+o_p(1))\sum_{i=1}^m(\rho^T\hat{R}_i)^2,
\end{aligned}$$

其中 $\rho=(\rho_1,\rho_2)^T$ 满足

$$\frac{1}{m}\sum_{i=1}^m\frac{\hat{R}_i}{1+\rho^T\hat{R}_i}=0. \tag{2-10}$$

下面由 (2-10) 计算 $\sum_{i=1}^m\rho^T\hat{R}_i$ 如下

$$\begin{aligned}
\frac{1}{m}\sum_{i=1}^m\frac{\rho^T\hat{R}_i}{1+\rho^T\hat{R}_i} &= \frac{1}{m}\sum_{i=1}^m\rho^T\hat{R}_i[1-\rho^T\hat{R}_i+\frac{(\rho^T\hat{R}_i)^2}{1+\rho^T\hat{R}_i}]\\
&= \frac{1}{m}\sum_{i=1}^m\rho^T\hat{R}_i-\frac{1}{m}\sum_{i=1}^m(\rho^T\hat{R}_i)^2(1-\frac{\rho^T\hat{R}_i}{1+\rho^T\hat{R}_i})\\
&= \frac{1}{m}\sum_{i=1}^m\rho^T\hat{R}_i-\frac{1}{m}\sum_{i=1}^m(\rho^T\hat{R}_i)^2(1+o_p(1))\\
&= 0,
\end{aligned}$$

则可解得

$$\sum_{i=1}^m\rho^T\hat{R}_i=\sum_{i=1}^m(\rho^T\hat{R}_i)^2(1+o_p(1)).$$

故

$$\begin{aligned}
-2\log L_1(I_{rc}) &= 2\sum_{i=1}^m \rho^T\hat{R}_i-(1+o_p(1))\sum_{i=1}^m(\rho^T\hat{R}_i)^2\\
&= (1+o_p(1))\sum_{i=1}^m(\rho^T\hat{R}_i)^2\\
&= (1+o_p(1))\rho^T(\sum_{i=1}^m\hat{R}_i\hat{R}_i^{\,T})\rho.
\end{aligned}$$

同样地, 再根据 (2-10) 来解 ρ 如下

$$\begin{aligned}
\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{R}_i}{1+\rho^T\hat{R}_i} &= \frac{1}{m}\sum_{i=1}^{m}\hat{R}_i[1-\rho^T\hat{R}_i+\frac{(\rho^T\hat{R}_i)^2}{1+\rho^T\hat{R}_i}] \\
&= \frac{1}{m}\sum_{i=1}^{m}\hat{R}_i-\frac{1}{m}\sum_{i=1}^{m}\hat{R}_i\hat{R}_i^T\rho+\sum_{i=1}^{m}\frac{\hat{R}_i(\rho^T\hat{R}_i)^2}{1+\rho^T\hat{R}_i} \\
&= \frac{1}{m}\sum_{i=1}^{m}\hat{R}_i-\frac{1}{m}\sum_{i=1}^{m}\hat{R}_i\hat{R}_i^T\rho+O_p(\max_{1\le i\le m}\|\frac{\hat{R}_i}{1+\rho^T\hat{R}_i}\|\cdot\frac{1}{m}\sum_{i=1}^{m}(\rho^T\hat{R}_i)^2) \\
&= \frac{1}{m}\sum_{i=1}^{m}\hat{R}_i-\frac{1}{m}\sum_{i=1}^{m}\hat{R}_i\hat{R}_i^T\rho+o_p(m^{1/2}\cdot\rho^T(\frac{1}{m}\sum_{i=1}^{m}\hat{R}_i\hat{R}_i^T)\rho) \\
&= \frac{1}{m}\sum_{i=1}^{m}\hat{R}_i-\frac{1}{m}\sum_{i=1}^{m}\hat{R}_i\hat{R}_i^T\rho+o_p(m^{-1/2}) \\
&= 0,
\end{aligned}$$

则可得 ρ 的表达式如下

$$\rho=(\frac{1}{m}\sum_{i=1}^{m}\hat{R}_i\hat{R}_i^T)^{-1}\cdot\frac{1}{m}\sum_{i=1}^{m}\hat{R}_i+o_p(m^{-1/2}).$$

将 ρ 带入 $-2\log L_1(I_{rc})$ 得

$$\begin{aligned}
-2\log L_1(I_{rc}) &= (1+o_p(1))(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\hat{R}_i)^T(\frac{1}{m}\sum_{i=1}^{m}\hat{R}_i\hat{R}_i^T)^{-1} \\
&\quad\cdot(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\hat{R}_i)+o_p(1)\xrightarrow{d}\chi_2^2,\ n\to\infty.
\end{aligned}$$

同理, 由引理 2.2.2 可证得

$$-2\log L_2(I_{rc})\xrightarrow{d}\chi_2^2,\ n\to\infty.$$

证毕. ■

注2.2.1 以上证明过程中, 由引理 2.2.1 中 (2-5) 有

$$\max_{1\le i\le m}|\hat{R}_i|=o_p(m^{1/2}). \tag{2-11}$$

根据引理 2.2.1 中 (2-3) 和 (2-4) 最后一式知

$$\|\frac{1}{m}\sum_{i=1}^{m}\hat{R}_i\hat{R}_i^T-I_2\|\xrightarrow{p}0.$$

此外, 由 Owen[1] 可知

$$\|\rho\| = O_p(m^{-1/2}). \tag{2-12}$$

则根据 (2-11) 和 (2-12) 得 $\rho^T\hat{R}_i$ 为无穷小, 即定理 2.2.1 证明中 $\log(1+\rho^T\hat{R}_i)$ 可泰勒展开, 且有

$$\begin{aligned}
\max_{1\le i\le m} |\frac{\rho^T\hat{R}_i}{1+\rho^T\hat{R}_i}| &= o_p(1), \\
\max_{1\le i\le m} \|\frac{\hat{R}_i}{1+\rho^T\hat{R}_i}\| &= o_p(m^{1/2}).
\end{aligned}$$

引理 2.2.1 的证明: 要证 (2-3) 成立, 即证对任意常数 a, b, 有 $a^2+b^2 \neq 0$, $\frac{1}{\sqrt{m}}\sum_{i=1}^m (a\frac{e_i}{\sqrt{\pi_{11}}} + b\frac{v_i}{\sqrt{\pi_{22}}})$. 则由 Lyapunov 中心极限定理可知, 只需证明 $\exists\,\delta>0$, 使得 $\frac{1}{B_n^{2+\delta}}\sum_{i=1}^m E|a\frac{e_i}{\sqrt{\pi_{11}}} + b\frac{v_i}{\sqrt{\pi_{22}}}|^{2+\delta} \to 0, m\to\infty$. 这里 $B_n = m^{1/2}$, $B_n^2 = \sum_{i=1}^m \sigma_k^2 = m$. 即证当 $m\to\infty$ 时, $\frac{1}{m^{2+\delta}/2}\sum_{i=1}^m E|a\frac{e_i}{\sqrt{\pi_{11}}} + b\frac{v_i}{\sqrt{\pi_{22}}}|^{2+\delta} \to 0$. 即

$$\begin{aligned}
&\frac{1}{m^{(2+\delta)/2}}\sum_{i=1}^m E\mid a\frac{e_i}{\sqrt{\pi_{11}}} + b\frac{v_i}{\sqrt{\pi_{22}}}\mid^{2+\delta} \\
=\; & \frac{1}{m^{\delta/2}} E\mid a\frac{e_1}{\sqrt{\pi_{11}}} + b\frac{v_1}{\sqrt{\pi_{22}}}\mid^{2+\delta} \\
\le\; & \frac{2^{1+\delta}}{m^{\delta/2}} E(\mid a\frac{e_1}{\sqrt{\pi_{11}}}\mid^{2+\delta} + \mid b\frac{v_1}{\sqrt{\pi_{22}}}\mid^{2+\delta}) \\
\le\; & \frac{(2\mid a\mid)^{2+\delta}}{m^{\delta/2}} E\mid \frac{e_1}{\sqrt{\pi_{11}}}\mid^{2+\delta} + \frac{(2\mid b\mid)^{2+\delta}}{m^{\delta/2}} E\mid \frac{v_1}{\sqrt{\pi_{22}}}\mid^{2+\delta} \\
\to\; & 0,\ m\to\infty.
\end{aligned}$$

要证 (2-4) 中 $\frac{\sum_{i=1}^m e_i^2}{m\pi_{11}} - 1 \xrightarrow{p} 0$ 成立, 即证对任意 $\epsilon > 0$, 有 $P(|\frac{\sum_{i=1}^m e_i^2}{m\pi_{11}} - 1| > \epsilon) \to 0$. 而

$$\begin{aligned}
&P(|\frac{\sum_{i=1}^m e_i^2}{m\pi_{11}} - 1| > \epsilon) \\
=\; & P(|\sum_{i=1}^m (e_i^2 - \pi_{11})| > \epsilon\cdot m\pi_{11}) \\
\le\; & \frac{E|\sum_{i=1}^m (e_i^2-\pi_{11})|^{(2+\delta)/2}}{(\epsilon\cdot m\pi_{11})^{(2+\delta)/2}}.
\end{aligned}$$

下面证明 $E|\sum_{i=1}^m (e_i^2-\pi_{11})|^{(2+\delta)/2}$ 为无穷小. 由于 $Ee_1^2 = \pi_{11}$ 以及当 $0<\delta\le 2$ 时,

根据 VonBahr-Essenn's 不等式 $E \mid S_n \mid^r \leq 2\sum_{v=1}^n E \mid X_v \mid^r$ 得

$$\begin{aligned} & E|\sum_{i=1}^m (e_i^2 - \pi_{11})|^{(2+\delta)/2} \\ \leq & \ 2\sum_{i=1}^m E|e_i^2 - \pi_{11}|^{(2+\delta)/2} \\ = & \ 2mE|e_i^2 - \pi_{11}|^{(2+\delta)/2} \\ = & \ O(mE|e_1|^{2+\delta}). \end{aligned}$$

当 $\delta > 2$ 时, 由 $E \mid S_n \mid^r \leq Cn^{r/2-1}\sum_{i=1}^n E \mid X_i \mid^r$ 得

$$\begin{aligned} & E|\sum_{i=1}^m (e_i^2 - \pi_{11}|^{(2+\delta)/2} \\ \leq & \ C \cdot m^{\frac{2+\delta}{4}} E \mid e_i^2 - \pi_{11} \mid^{(2+\delta)/2} \\ = & \ O(m^{\frac{2+\delta}{4}} E \mid e_1 \mid^{2+\delta}), \end{aligned}$$

则

$$\begin{aligned} & E|\sum_{i=1}^m (e_i^2 - \pi_{11})|^{(2+\delta)/2} \\ = & \ O(m^{2+\max(2,\delta)/4} \cdot E|e_1|^{2+\delta}) \\ = & \ o((m\pi_{11})^{(2+\delta)/2}). \end{aligned}$$

故

$$\begin{aligned} & P(|\frac{\sum_{i=1}^m e_i^2}{m\pi_{11}} - 1| > \epsilon) \\ \leq & \ \frac{E|\sum_{i=1}^m (e_i^2 - \pi_{11})|^{(2+\delta)/2}}{(\epsilon \cdot m\pi_{11})^{(2+\delta)/2}} \\ = & \ o(1) \to 0. \end{aligned}$$

由于 $E(e_i v_i) = 0$, 同理可证得 (2-4) 后面两式成立.

下面证明 (2-5) 中的 $\max_{1\leq i\leq m} |e_i/\sqrt{\pi_{11}}| = o_p(m^{1/2})$ 成立, 另一式同理可证.

要证 $\max_{1\leq i\leq m} |e_i/\sqrt{\pi_{11}}| = o_p(m^{1/2})$, 即证

$$P(\max_{1\leq i\leq m} |e_i/\sqrt{\pi_{11}}| \geq x) \to 0.$$

由 (2-4) 证明过程可知当 $a=1, b=0$ 时, 有

$$\begin{aligned}
&\frac{1}{m^{\delta/2}}E|\frac{e_1}{\sqrt{\pi_{11}}}|^{2+\delta}\\
\leq\ & \frac{2^{2+\delta}}{m^{\delta/2}}E|\frac{e_1}{\sqrt{\pi_{11}}}|^{2+\delta}\\
\to\ & 0,\ m\to\infty.
\end{aligned}$$

故

$$E|\frac{e_1}{\sqrt{\pi_{11}}}|^{2+\delta}=o(m^{\delta/2}).$$

而

$$\begin{aligned}
&P(\max_{1\leq i\leq m}|e_i/\sqrt{\pi_{11}}|\geq x)\\
\leq\ & m\cdot P(|\frac{e_1}{\sqrt{\pi_{11}}}|\geq x)\\
\leq\ & m\cdot\frac{1}{x^{2+\delta}}E|\frac{e_1}{\sqrt{\pi_{11}}}|^{2+\delta}.
\end{aligned}$$

故

$$P(\max_{1\leq i\leq m}|e_i/\sqrt{\pi_{11}}|\geq x)\to 0.$$

证毕. ■

引理 2.2.2 的证明: 首先证明 (2-6) 成立. 记

$$\bar{R}^1=(\bar{R}^1_{\cdot 1},\cdots,\bar{R}^1_{\cdot rc})^T,\quad \bar{R}^2=(\bar{R}^2_{\cdot 1},\cdots,\bar{R}^2_{\cdot rc})^T,$$

其中 $\bar{R}^1=\frac{1}{m}\sum_{i=1}^m R_i$, $\bar{R}^2=\frac{1}{m}\sum_{i=m+1}^{2m}R_i$. 下面将 $\{e_i^*\}$ 近似写为 $\{e_i\}$.

$$\begin{aligned}
e_i^* &= tr[(Y_i^*-\frac{(m-1)\Omega}{m})(Y_{i+m}^*-\frac{(m-1)\Omega}{m})]\\
&=\sum_{j=1}^{rc}\sum_{k=1}^{rc}[(R_{ij}-\bar{R}^1_{\cdot j})(R_{ik}-\bar{R}^1_{\cdot k})-\frac{(m-1)\sigma_{jk}}{m}]\\
&\quad\cdot[(R_{m+i,j}-\bar{R}^2_{\cdot j})(R_{m+i,k}-\bar{R}^2_{\cdot k})-\frac{(m-1)\sigma_{jk}}{m}]\\
&=\sum_{j=1}^{rc}\sum_{k=1}^{rc}\{[(R_{ij}-\mu_j)-(\bar{R}^1_{\cdot j}-\mu_j)][(R_{ik}-\mu_k)-(\bar{R}^1_{\cdot k}-\mu_k)]-\frac{(m-1)\sigma_{jk}}{m}\}\\
&\quad\cdot\{[(R_{m+i,j}-\mu_j)-(\bar{R}^2_{\cdot j}-\mu_j)][(R_{m+i,k}-\mu_k)-(\bar{R}^2_{\cdot k}-\mu_k)]-\frac{(m-1)\sigma_{jk}}{m}\}\\
&=e_i+\sum_{j=1}^{8}T_{ij},
\end{aligned}$$

这里

$$\begin{aligned}
T_{i1} &= -2\sum_{j=1}^{rc}\sum_{k=1}^{rc}[(R_{ij}-\mu_j)(R_{ik}-\mu_k)-\sigma_{jk}][(R_{m+i,j}-\mu_j)(\bar{R}^2_{\cdot k}-\mu_k)-\frac{\sigma_{jk}}{m}],\\
T_{i2} &= -2\sum_{j=1}^{rc}\sum_{k=1}^{rc}[(R_{m+i,j}-\mu_j)(R_{m+i,k}-\mu_k)-\sigma_{jk}][(R_{ij}-\mu_j)(\bar{R}^1_{\cdot k}-\mu_k)-\frac{\sigma_{jk}}{m}],\\
T_{i3} &= \sum_{j=1}^{rc}\sum_{k=1}^{rc}[(R_{ij}-\mu_j)(\bar{R}^1_{\cdot k}-\mu_k)+(R_{ik}-\mu_k)(\bar{R}^1_{\cdot j}-\mu_j)-\frac{2\sigma_{jk}}{m}]\\
&\quad\cdot[(R_{m+i,j}-\mu_j)(\bar{R}^2_{\cdot k}-\mu_k)+(R_{m+i,k}-\mu_k)(\bar{R}^2_{\cdot j}-\mu_j)-\frac{2\sigma_{jk}}{m}],\\
T_{i4} &= \sum_{j=1}^{rc}\sum_{k=1}^{rc}[(R_{ij}-\mu_j)(R_{ik}-\mu_k)-\sigma_{jk}][(\bar{R}^2_{\cdot j}-\mu_j)(\bar{R}^2_{\cdot k}-\mu_k)-\frac{\sigma_{jk}}{m}],\\
T_{i5} &= \sum_{j=1}^{rc}\sum_{k=1}^{rc}[(R_{m+i,j}-\mu_j)(R_{m+i,k}-\mu_k)-\sigma_{jk}][(\bar{R}^1_{\cdot j}-\mu_j)(\bar{R}^1_{\cdot k}-\mu_k)-\frac{\sigma_{jk}}{m}],\\
T_{i6} &= \sum_{j=1}^{rc}\sum_{k=1}^{rc}[(\bar{R}^1_{\cdot j}-\mu_j)(\bar{R}^1_{\cdot k}-\mu_k)-\frac{\sigma_{jk}}{m}][(\bar{R}^2_{\cdot j}-\mu_j)(\bar{R}^2_{\cdot k}-\mu_k)-\frac{\sigma_{jk}}{m}],\\
T_{i7} &= -2\sum_{j=1}^{rc}\sum_{k=1}^{rc}[(R_{ij}-\mu_j)(\bar{R}^1_{\cdot k}-\mu_k)-\sigma_{jk}][(\bar{R}^2_{\cdot j}-\mu_j)(\bar{R}^2_{\cdot k}-\mu_k)-\frac{\sigma_{jk}}{m}],\\
T_{i8} &= -2\sum_{j=1}^{rc}\sum_{k=1}^{rc}[(R_{m+i,j}-\mu_j)(\bar{R}^2_{\cdot k}-\mu_k)-\sigma_{jk}][(\bar{R}^1_{\cdot j}-\mu_j)(\bar{R}^1_{\cdot k}-\mu_k)-\frac{\sigma_{jk}}{m}].
\end{aligned}$$

接下来, 证明对所有的 $1\le j\le 8$, 当 $(tr(\Omega^2))^2=o(m^2Ee_1^2)$ 时, 有

$$E(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}T_{ij})^2=o(Ee_1^2),\quad \frac{1}{m}\sum_{i=1}^{m}ET_{ij}^2=o(Ee_1^2). \tag{2-13}$$

这里仅给出 (2-13) 对于 T_{i1} 成立的详细证明, 其他项与 T_{i1} 证明类似. 令

$$\delta_{ij}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$$

由于 $R_i, i=1,\cdots,m$ 是 $i.i.d.$ 随机向量, 则

$$\begin{aligned}
E(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}T_{i1})^2 &= \frac{1}{m}\sum_{i=1}^{m}ET_{i1}^2\\
&= \frac{4}{m^3}\sum_{i=1}^{m}E\{\sum_{j=1}^{rc}\sum_{k=1}^{rc}[(R_{ij}-\mu_j)(R_{ik}-\mu_k)-\sigma_{jk}]\\
&\quad\cdot[\sum_{l=1}^{m}(R_{m+i,j}-\mu_j)(R_{m+l,k}-\mu_k)-\sigma_{jk}]\}^2,
\end{aligned}$$

上式等于

$$\begin{aligned}
&\frac{4}{m^3}\sum_{i=1}^{m}E\{\sum_{j,j'=1}^{rc}\sum_{k,k'=1}^{rc}\sum_{l,l'=1}^{m}[(R_{ij}-\mu_j)(R_{ik}-\mu_k)-\sigma_{jk}]\\
&\cdot[(R_{m+i,j}-\mu_j)(R_{m+l,k}-\mu_k)-\sigma_{jk}\delta_{il}]\\
&\cdot[(R_{ij'}-\mu_{j'})(R_{ik'}-\mu_{k'})-\sigma_{j'k'}][(R_{m+i,j'}-\mu_{j'})(R_{m+l',k'}-\mu_{k'})-\sigma_{j'k'}\delta_{il'}]\}\\
=\ & I_1+I_2.
\end{aligned}$$

这里

$$\begin{aligned}
I_1 &= \frac{4}{m^3}\sum_{i=1}^{m}E\{\sum_{j=1}^{rc}\sum_{k=1}^{rc}[(R_{ij}-\mu_j)(R_{ik}-\mu_k)-\sigma_{jk}]\\
&\quad\cdot[(R_{m+i,j}-\mu_j)(R_{m+i,k}-\mu_k)-\sigma_{jk}]\}^2\\
&= \frac{4}{m^3}\sum_{i=1}^{m}Ee_1^2\\
&= \frac{4}{m^2}Ee_1^2\\
&= o(Ee_1^2)
\end{aligned}$$

和

$$\begin{aligned}
I_2 &= \frac{4}{m^3}\sum_{i=1}^{m}E\Big\{\sum_{j,j'=1}^{rc}\sum_{k,k'=1}^{rc}\sum_{l\neq i,l=1}^{m}\\
&\quad\{[(R_{ij}-\mu_j)(R_{ik}-\mu_k)-\sigma_{jk}][(R_{ij'}-\mu_{j'})(R_{ik'}-\mu_{k'})-\sigma_{j'k'}]\\
&\quad\cdot[(R_{m+i,j}-\mu_j)(R_{m+l,k}-\mu_k)][(R_{m+i,j'}-\mu_{j'})(R_{m+l,k'}-\mu_{k'})]\}\Big\}\\
&= \frac{4m(m-1)}{m^3}\sum_{j,j'=1}^{rc}\sum_{k,k'=1}^{rc}\sigma_{jj'}\sigma_{kk'}E\{[(R_{1j}-\mu_j)(R_{1k}-\mu_k)-\sigma_{jk}]\\
&\quad\cdot[(R_{1j'}-\mu_{j'})(R_{1k'}-\mu_{k'})-\sigma_{j'k'}]\}\\
&\leq \frac{4}{m}(\sum_{j,j'=1}^{rc}\sum_{k,k'=1}^{rc}\sigma_{jj'}^2\sigma_{kk'}^2)^{\frac{1}{2}}(Ee_1^2)^{\frac{1}{2}}\\
&= \frac{4}{m}(\sum_{j,j'=1}^{rc}\sigma_{jj'}^2)(Ee_1^2)^{\frac{1}{2}}\\
&= o(Ee_1^2).
\end{aligned}$$

由 $(tr(\Omega^2))^2=o(m^2Ee_1^2)$ 可知,(2-13) 对 T_{i1} 成立. 与 $\{e_i^*\}$ 类似, 有

$$\begin{aligned}
\Delta_{i1} &= -\frac{2}{m}\sum_{j,k=1}^{rc}\sum_{l\neq i}^{m}[(R_{ij}-\mu_j)(R_{lk}-\mu_k)+(R_{m+i,j}-\mu_j)(R_{m+l,k}-\mu_k)],\\
\Delta_{i2} &= \frac{1}{m^2}\sum_{j,k=1}^{rc}\sum_{s\neq t}^{m}[(R_{sj}-\mu_j)(R_{tk}-\mu_k)+(R_{m+s,j}-\mu_j)(R_{m+t,k}-\mu_k)].
\end{aligned}$$

则

$$v_i^* = \frac{(m-1)v_i}{m} + \Delta_{i1} + \Delta_{i2}.$$

当 $(\sum_{j,k=1}^{rc}\sigma_{jk})^2 = o(mEv_1^2)$ 时, 有

$$\begin{aligned}
& E(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\Delta_{i1})^2 \\
= & \frac{8}{m^3}\sum_{i,i'=1}^{m}\sum_{j,k=1}^{rc}\sum_{j',k'=1}^{rc}\sum_{l\neq i}^{m}\sum_{l'\neq i'}^{m} E[(R_{ij}-\mu_j)(R_{lk}-\mu_k)(R_{i'j'}-\mu_{j'})(R_{l'k'}-\mu_{k'})] \\
= & \frac{8}{m^3}\sum_{i=1}^{m}\sum_{l\neq i}^{m}\sum_{j,k=1}^{rc}\sum_{j',k'=1}^{rc} E[(R_{ij}-\mu_j)(R_{lk}-\mu_k)(R_{ij'}-\mu_{j'})(R_{lk'}-\mu_{k'})] \\
& +\frac{8}{m^3}\sum_{i=1}^{m}\sum_{l\neq i}^{m}\sum_{j,k=1}^{rc}\sum_{j',k'=1}^{rc} E[(R_{ij}-\mu_j)(R_{lk}-\mu_k)(R_{lj'}-\mu_{j'})(R_{ik'}-\mu_{k'})] \\
= & \frac{16(m-1)}{m^2}(\sum_{j,k=1}^{rc}\sigma_{jk})^2 \\
= & o(Ev_1^2) \qquad (2\text{-}14)
\end{aligned}$$

和

$$\begin{aligned}
& \frac{1}{m}\sum_{i=1}^{m}E\Delta_{i1}^2 \\
= & \frac{8}{m^3}\sum_{i=1}^{m}\sum_{l\neq i}^{m}\sum_{j,k=1}^{rc}\sum_{j',k'=1}^{rc} E[(R_{ij}-\mu_j)(R_{lk}-\mu_k)(R_{ij'}-\mu_{j'})(R_{lk'}-\mu_{k'})] \\
= & \frac{8(m-1)}{m^2}(\sum_{j,k=1}^{rc}\sigma_{jk})^2 \\
= & o(Ev_1^2). \qquad (2\text{-}15)
\end{aligned}$$

类似地, 当 $(\sum_{j,k=1}^{rc}\sigma_{jk})^2 = o(mEv_1^2)$ 时, 可得

$$E(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\Delta_{i2})^2 = o(Ev_1^2), \qquad (2\text{-}16)$$

$$\frac{1}{m}\sum_{i=1}^{m}E\Delta_{i2}^2 = o(Ev_1^2). \qquad (2\text{-}17)$$

根据 (2-14)-(2-17), 可知当 $(\sum_{j,k=1}^{rc}\sigma_{jk})^2 = o(mEv_1^2)$ 时, 对所有 $1 \le j \le 2$, 有

$$E(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\Delta_{ij})^2 = o(Ev_1^2), \quad \frac{1}{m}\sum_{i=1}^{m}E\Delta_{i2j}^2 = o(Ev_1^2). \qquad (2\text{-}18)$$

则由 (2-13), (2-18) 和引理 2.2.1, 可知对任意常数 c 和 d 有

$$
\begin{aligned}
&\frac{1}{\sqrt{m}}\sum_{i=1}^{m}(c\frac{e_i^*}{\sqrt{\pi_{11}}}+d\frac{v_i^*}{\sqrt{\pi_{22}}})\\
=\;&\frac{1}{\sqrt{m}}\sum_{i=1}^{m}(c\frac{e_i}{\sqrt{\pi_{11}}}+d\frac{v_i}{\sqrt{\pi_{22}}})+\frac{1}{\sqrt{m}}\sum_{i=1}^{m}(c\frac{\sum_{j=1}^{8}T_{ij}}{\sqrt{\pi_{11}}}+d\frac{\sum_{j=1}^{2}\Delta_{ij}}{\sqrt{\pi_{22}}})\\
=\;&\frac{1}{\sqrt{m}}\sum_{i=1}^{m}(c\frac{e_i}{\sqrt{\pi_{11}}}+d\frac{v_i}{\sqrt{\pi_{22}}})+o_p(1)\\
\xrightarrow{d}\;&N(0,c^2+d^2),\ n\to\infty.
\end{aligned}
$$

因此, 由 Cramer-Wold 引理知 (2-6) 成立. 根据 (2-13), (2-18) 和引理 2.2.1 的 (2-4), 可得 (2-7) 也成立.

下面证明 (2-8) 中的第一个等式成立, 另一式同理可证.

要证

$$\max_{1\le i\le m}|e_i^*/\sqrt{\pi_{11}}|=o_p(m^{1/2}),$$

根据 $e_i^*=e_i+\sum_{j=1}^{8}T_{ij}$, 可知即证

$$\max_{1\le i\le m}|\sum_{j=1}^{8}T_{ij}/\sqrt{\pi_{11}}|=o_p(m^{1/2}),$$

即证

$$P(\max_{1\le i\le m}\max_{1\le j\le 8}|T_{ij}|>\varepsilon\sqrt{m\pi_{11}})\to 0.$$

由 (2-13) 的第二个等式及Markov不等式, 可知对 $\forall\,\varepsilon>0$, 有

$$
\begin{aligned}
&P(\max_{1\le i\le m}\max_{1\le j\le 8}|T_{ij}|>\varepsilon\sqrt{m\pi_{11}})\\
\le\;&\sum_{i=1}^{m}\sum_{j=1}^{8}P(|T_{ij}|>\varepsilon\sqrt{m\pi_{11}})\\
\le\;&\frac{\sum_{i=1}^{m}\sum_{j=1}^{8}ER_{ij}^2}{\varepsilon^2 mEe_1^2}\\
\to\;&0.
\end{aligned}
$$

根据上式和 (2-5) 的第一个等式知, (2-8) 的第一个等式成立. 类似地, (2-8) 的第二个等式也成立. 证毕. ■

引理 2.2.3 的证明: 已知 $e_1 = tr[(Y_1 - I_{rc})(Y_{1+m} - I_{rc})]$, $v_1 = 1_{rc}^T(Y_1 + Y_{1+m} - 2I_{rc})1_{rc}$, $Y_i = (\Sigma_C^{1/2} \bigotimes \Sigma_R^{1/2})W_iW_i'(\Sigma_C^{1/2} \bigotimes \Sigma_R^{1/2}) = R_iR_i'$, 故

$$
\begin{aligned}
Ee_1^4 &= E\{tr[(Y_1 - I_{rc})(Y_{1+m} - I_{rc})]\}^4 \\
&= E\{tr[(R_1R_1' - I_{rc})(R_{1+m}R_{1+m}' - I_{rc})]\}^4 \\
&= E\{\sum_{i=1}^{rc}\sum_{j=1}^{rc}(R_{1i}R_{1j} - \delta_{ij})(R_{1+m,i}R_{1+m,j} - \delta_{ij})\}^4 \\
&\leq E\{4\sum_{i=1}^{rc}\sum_{j=1}^{rc}R_{1i}R_{1j}R_{1+m,i}R_{1+m,j}\}^4 \\
&= 256E\{\sum_{i=1}^{rc}\sum_{j=1}^{rc}R_{1i}R_{1j}R_{1+m,i}R_{1+m,j}\}^4 \\
&= 256E\{\sum_{i=1}^{rc}R_{1i}R_{1+m,i}\}^8 \\
&= 256\sum_{1\leq i_1,\cdots,i_8\leq rc}\{E(\prod_{v=1}^{8}R_{1,i_l})\}^2 \\
&:= 256\sum_{1\leq i_1,\cdots,i_8\leq rc}[T(i_1,\cdots,i_8)]^2.
\end{aligned}
$$

记 $\Omega^{1/2} = (\gamma_{ij})_{rc\times rc}$, $U_i = W_{1i}$, $i = 1,\cdots,rc$, 则 $Var(R_i) = \Omega = (\sigma_{ij})_{rc\times rc} = (\Omega^{1/2})^T\Omega^{1/2}$. 其中 $\sigma_{ij} = \sum_{k=1}^{rc}\gamma_{ik}\gamma_{kj}$. 由 $R_i = \Omega^{1/2}W_i + E(R_i)$, 可知 $R_{1i} = \sum_{j=1}^{rc}\gamma_{ij}U_j$, $i = 1,\cdots,rc$. 故 $T(i_1,\cdots,i_8) = E(\prod_{l=1}^{8}R_{1,i_l}) = E(\prod_{l=1}^{8}(\sum_{j=1}^{rc}\gamma_{ij}U_j))$. 易知, $E(\prod_{i=1}^{8}U_{p_i}) \neq 0$ 意味着序列 $p_1,\cdots,p_8$ 中, 每个 p_i 的值至少出现两次. 记 $B_l = \{(p_1,\cdots,p_l) : 1 \leq p_1,\cdots,p_l \leq c\}$, 令 S_k 为 k 置换集. 则

$$
\begin{aligned}
T(i_1,\cdots,i_8) &= \sum_{(a,b,d,e)\in B_4}\sum_{\substack{(k_1,\cdots,k_8)=\sigma(i_1,\cdots,i_8)\\ \sigma\in S_8}}\gamma_{k_1a}\gamma_{k_2a}\gamma_{k_3b}\gamma_{k_4b}\gamma_{k_5d}\gamma_{k_6d}\gamma_{k_7e}\gamma_{k_8e} \\
&\quad\cdot E(U_a^2U_b^2U_d^2U_e^2) + \sum_{\substack{(a,b,d)\in B_3\\ a\neq b}}\sum_{\substack{(k_1,\cdots,k_8)=\sigma(i_1,\cdots,i_8)\\ \sigma\in S_8}}\gamma_{k_1a}\gamma_{k_2a}\gamma_{k_3a} \\
&\quad\cdot\gamma_{k_4b}\gamma_{k_5b}\gamma_{k_6b}\gamma_{k_7d}\gamma_{k_8d}E(U_a^3U_b^3U_d^2) \\
&= T_1(i_1,\cdots,i_8) + T_2(i_1,\cdots,i_8).
\end{aligned}
$$

记 L 为 $E(U_1^8)$ 的一致界. 注意到 $\sum_{(k_1,\cdots,k_8)=\sigma(i_1,\cdots,i_8)}$ 至多由 8! 项组成, 对 $k_1,\cdots,k_8$ 的每个选择 (如: $k_1 = i_1,\cdots,k_8 = i_8$), 可以得到同样的值如下

$$
\sum_{1\leq i_1,\cdots,i_8\leq rc}\sum_{(a,b,d,e)\in B_4}\gamma_{k_1a}\gamma_{k_2a}\gamma_{k_3b}\gamma_{k_4b}\gamma_{k_5d}\gamma_{k_6d}\gamma_{k_7e}\gamma_{k_8e}E(U_a^2U_b^2U_d^2U_e^2).
$$

故

$$
\begin{aligned}
&\sum_{1\le i_1,\cdots,i_8\le rc} T_1^2(i_1,\cdots,i_8)\\
\le\ & (8!)^2 \sum_{1\le i_1,\cdots,i_8\le rc}\ \sum_{(a,b,d,e)\in B_4}\ \sum_{(a',b',d',e')\in B_4} \gamma_{i_1a}\gamma_{i_1a'}\gamma_{i_2a}\gamma_{i_2a'}\cdots\gamma_{i_8e}\gamma_{i_8e'}\\
&\cdot E(U_a^2U_b^2U_d^2U_e^2)E(U_{a'}^2U_{b'}^2U_{d'}^2U_{e'}^2)\\
=\ & (8!)^2 \sum_{(a,b,d,e)\in B_4}\ \sum_{(a',b',d',e')\in B_4} (\sigma_{aa'}\sigma_{bb'}\sigma_{dd'}\sigma_{ee'})^2 E(U_a^2U_b^2U_d^2U_e^2)E(U_{a'}^2U_{b'}^2U_{d'}^2U_{e'}^2)\\
\le\ & (8!)^2L^2 \sum_{(a,b,d,e)\in B_4}\ \sum_{(a',b',d',e')\in B_4} (\sigma_{aa'}\sigma_{bb'}\sigma_{dd'}\sigma_{ee'})^2\\
=\ & O(1)(\sum_{1\le a,a'\le rc} \sigma_{aa'}^2)^4\\
=\ & O((tr(\Omega^2))^4).
\end{aligned}
$$

同理可得

$$\sum_{1\le i_1,\cdots,i_8\le rc} T_2^2(i_1,\cdots,i_8) \le O((tr(\Omega^2))^4).$$

则

$$
\begin{aligned}
Ee_1^4 &\le 256 \sum_{1\le i_1,\cdots,i_8\le rc} T^2(i_1,\cdots,i_8)\\
&= O((tr(\Omega^2))^4).
\end{aligned}
$$

另一方面, 令 $V_i = W_{1+m,i}$, $i=1,\cdots,rc$, 同理可以证得

$$(Ee_1^2)^2 \ge O((tr(\Omega^2))^4).$$

因此有 $Ee_1^4/(Ee_1^2)^2 = O(1)$. 用同样的方式也可以得到 $Ev_1^4/(Ev_1^2)^2 = O(1)$. 证毕. ■

推论 2.2.1 的证明: 当 M 已知时, 根据引理 2.2.3, 可知在条件 $(A2)$ 下有 $E|e_1|^{2+\delta} = O(p^{2+\delta})$, $E|v_1|^{2+\delta} = O(p^{2+\delta})$. 另外, 在条件 $(A1)$ 下, 对某常数 $C>0$, 有 $\pi_{11} = tr\Theta^2 \ge pC$, $\pi_{22} = 1_p^T\Theta 1_p \ge pC$. 故可得

$$E|e_1|^{2+\delta}/\pi_{11}^{(2+\delta)/2} = O(p^{(2+\delta)/2}) = O((rc)^{2+\delta}),$$

$$E|v_1|^{2+\delta}/\pi_{22}^{(2+\delta)/2} = O(p^{(2+\delta)/2}) = O((rc)^{2+\delta}).$$

又由条件 $(A3)$ 有

$$rc = o(n^{(\delta+\min(2,\delta))/4(2+\delta)}).$$

综上, 定理 2.2.1 中 (2-1) 成立, 即推论 2.2.1 成立.

当 M 未知时, 利用定理 2.2.1 中 (2-1), 只需证 (2-2) 成立. 由 $\max_{1\le i\le rc}\sigma_{ii} < C_0$ 有

$$(tr(\Omega^2))^2 = (\sum_{1\le i,j\le rc} \sigma_{ij}^2)^2 \le p^2(\max_{1\le i\le rc} \sigma_{ii}^2) \le C_0^2 p^2, \tag{2-19}$$

$$(1_{rc}^T\Omega 1_{rc})^2 \le p^2(\max_{1\le i\le rc} \sigma_{ii}^2) \le C_0^2 p^2. \tag{2-20}$$

另外, 根据条件 $(A1)$, 可知存在某常数 $C > 0$, 使得

$$\pi_{11} = tr(\Theta^2) \ge pC\ , \quad \pi_{22} = 1_p^T\Theta 1_p \ge pC. \tag{2-21}$$

根据条件 $(A3)$, 有

$$rc = o(N^{1/4})\ , \quad p = 0(N^{1/2}).$$

综合 (2-19)-(2-21), 可得

$$m\cdot Ee_1^2 = m\pi_{11} \ge C\cdot mp \ge (tr(\Omega^2))^2,$$

$$\sqrt{m}Ev_1^2 = \sqrt{m}\pi_{22} \ge \sqrt{m}Cp \ge (1_{rc}^T\Omega 1_{rc})^2.$$

证毕. ■

推论 2.2.2 的证明: 根据引理 2.2.3, 可知当 $\delta = 2$ 时, 定理 2.2.1 中 (2-1) 条件成立, 故 M 已知时该推论成立. 当 M 未知时, 由于定理 2.2.1 中 (2-1) 也成立, 故仅需证明 (2-2) 成立.

根据引理 2.2.3 证明过程可得

$$E(e_1^2) \ge C^2(tr(\Omega^2))^2, \qquad E(v_1^2) \ge C(1_{rc}^T\Omega 1_{rc})^2.$$

则定理 2.2.1 中 (2-2) 成立. 证毕. ■

定理 2.2.2 的证明: 首先, 记

$$\pi_{11} = E(e_1^2(\Omega)), \quad \pi_{22} = E(v_1^2(\Omega)),$$

$$\begin{aligned}\nu &= m(\zeta_{m1}^2+\zeta_{m2}^2),\\ \zeta_{m1} &= tr[(\Omega-I_{rc})^2]/\sqrt{\pi_{11}},\\ \zeta_{m2} &= 2\cdot 1_{rc}^T(\Omega-I_{rc})1_{rc}/\sqrt{\pi_{22}}.\end{aligned}$$

下面只须证均值矩阵 M 已知时的情况, M 未知时证明类似.

由于非中心参数 $\nu=m(\zeta_{m1}^2+\zeta_{m2}^2)$, 故分以下两种情况进行讨论:

首先考虑 $\frac{\nu}{m}=o(1)$ 的情况, 根据 $EY_1=\Omega$, 可知在备择假设 H_1 成立下有

$$\begin{aligned}e_i(I_{rc}) &= tr[(Y_i-I_{rc})(Y_{i+m}-I_{rc})]\\ &= tr[(Y_i-\Omega+\Omega-I_{rc})(Y_{i+m}-\Omega+\Omega-I_{rc})]\\ &= e_i(\Omega)+tr[(\Omega-I_{rc})^2]+tr[(\Omega-I_{rc})(Y_i+Y_{i+m}-2\Omega)],\end{aligned}$$

同理有

$$v_i(I_{rc}) = v_i(\Omega)+2\cdot 1_{rc}^T(\Omega-I_{rc})1_{rc}.$$

则

$$\begin{aligned}\frac{1}{\sqrt{m}}\sum_{i=1}^m(\frac{e_i(\Omega)}{\sqrt{\pi_{11}}},\frac{v_i(\Omega)}{\sqrt{\pi_{22}}})^T &= \frac{1}{\sqrt{m}}\sum_{i=1}^m(\frac{e_i(\Omega)}{\sqrt{\pi_{11}}},\frac{v_i(\Omega)}{\sqrt{\pi_{22}}})^T\\ &\quad+\sqrt{m}(\zeta_{m1},\zeta_{m2})^T+\frac{1}{\sqrt{m}}\sum_{i=1}^m(\eta_i(\Omega),0)^T.\end{aligned}$$

这里 $\eta_i(\Omega)=tr[(\Omega-I_{rc})^2]+tr[(\Omega-I_{rc})(Y_i+Y_{i+m}-2\Omega)]/\sqrt{\pi_{11}}$, 且 $E[\eta_i(\Omega]=0$. 而

$$\begin{aligned}E[\eta_i(\Omega)]^2 &= 4E\{tr[(\Omega-I_{rc})(Y_i-\Omega)]^2\}/\pi_{11}\\ &\le 4E\{tr[(\Omega-I_{rc})^2]tr[(Y_i-\Omega)^2]\}/\pi_{11}\\ &= O[tr((\Omega-I_{rc})^2)/\sqrt{\pi_{11}}]\\ &= o(1)\\ &\to 0.\end{aligned}$$

考虑到满足 $\eta_i^2(\Omega)$ 独立同分布且其期望均为 0, 利用辛钦大数定律可得

$$\frac{1}{m}\sum_{i=1}^m\eta_i^2(\Omega)\xrightarrow{p}0,$$

$$\frac{\max_{1\le i\le m}|\eta_i(\Omega)|}{\sqrt{m}} \le \sqrt{\frac{\sum_{i=1}^m\eta_i^2(\Omega)}{m}}\xrightarrow{p}0.$$

令

$$V_m=\begin{pmatrix}V_{m1}\\ V_{m2}\end{pmatrix}=\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left\{\begin{pmatrix}\frac{e_i(I_{rc})}{\sqrt{\pi_{11}}}\\ \frac{v_i(I_{rc})}{\sqrt{\pi_{22}}}\end{pmatrix}-\begin{pmatrix}\zeta_{m1}\\ \zeta_{m1}\end{pmatrix}\right\}$$

$$=\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\hat{R}_i-\sqrt{m}\begin{pmatrix}\zeta_{m1}\\ \zeta_{m1}\end{pmatrix}.$$

根据引理 2.2.1 可知

$$V_m\xrightarrow{d}N(0,I_2).\tag{2-22}$$

利用定理 2.2.1 证明知

$$\begin{aligned}&-2\log L_1(I_{rc})\\ =&\ (1+o_p(1))\Big(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\hat{R}_i\Big)^T\Big(\frac{1}{m}\sum_{i=1}^{m}\hat{R}_i\hat{R}_i^{\,T}\Big)^{-1}\Big(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\hat{R}_i\Big)+o_p(1)\\ =&\ (1+\zeta_{m1}^2+\zeta_{m2}^2)^{-1}\\ &\cdot[(1+\zeta_{m2}^2)(V_{m1}+\sqrt{m}\zeta_{m1})^2-2\zeta_{m1}\zeta_{m2}(V_{m1}+\sqrt{m}\zeta_{m1})(V_{m2}+\sqrt{m}\zeta_{m2})\\ &+(1+\zeta_{m1}^2)(V_{m2}+\sqrt{m}\zeta_{m2})^2]+o_p(1)\\ =&\ (V_{m1}+\sqrt{m}\zeta_{m1})^2(1+o_p(1))+(V_{m2}+\sqrt{m}\zeta_{m2})^2(1+o_p(1))+o_p(1).\end{aligned}$$

接下来分非中心参数 ν 有限和无限两个方面来进行讨论:

若 $\nu=m(\zeta_{m1}^2+\zeta_{m2}^2)$ 有限, 且其极限为 ν_0, 则由 (2-22) 和 $-2\log L_1(I_{rc})$ 表达式, 可知 $-2\log L_2(I_{rc})\xrightarrow{d}\chi^2_{2,\nu_0}$.

若 $\nu\to\infty$, 则 (2-9) 右边极限为 1, 接下来求其左边极限如下

$$\begin{aligned}&-2\log L_1(I_{rc})\\ \geq&\ \Big(\frac{m\zeta_{m1}^2}{2}-V_{m1}^2\Big)(1+o_p(1))+\Big(\frac{m\zeta_{m2}^2}{2}-V_{m2}^2\Big)(1+o_p(1))+o_p(1)\\ =&\ \frac{\nu}{2}(1+o_p(1))-(V_{m1}^2+V_{m2}^2)(1+o_p(1))+o_p(1)\\ \xrightarrow{p}&\ \infty,\end{aligned}$$

这意味着 (2-9) 左边极限也为 1. 故当 $\frac{\nu}{m}=o(1)$, 即 $\nu=o(m)$ 时 (2-9) 成立.

下面证明 $\liminf\limits_{m\to\infty}\nu/m>0$ 时的情况:

由于 $\nu = m(\zeta_{m1}^2 + \zeta_{m2}^2)$, 首先我们用放缩的方法证明 $\liminf_{m\to\infty} \zeta_{m2}^2 > 0$ 的情况, 那么 $\liminf_{m\to\infty} \zeta_{m1} > 0$ 同理也可证得.

由于 $\sum_{i=1}^m p_i R_i(I_{rc}) = 0 \Rightarrow \sum_{i=1}^m p_i v_i(I_{rc}) = 0$, 则可对 $L_1(I_{rc})$ 进行如下放缩

$$\begin{aligned} L_1(I_{rc}) &\le \sup\{\prod_{i=1}^m (mp_i) : \sum_{i=1}^m p_i = 1, \sum_{i=1}^m p_i v_i(I_{rc}) = 0, p_1 \ge 0, \ldots, p_m \ge 0\} \\ &= \sup\{\prod_{i=1}^m (mp_i) : \sum_{i=1}^m p_i = 1, \sum_{i=1}^m p_i v_i(I_{rc})/\sqrt{\pi_{22}} = 0, p_1 \ge 0, \ldots, p_m \ge 0\}. \end{aligned}$$

接下来定义

$$L^*(\theta) = \sup\{\prod_{i=1}^m (mp_i) : \sum_{i=1}^m p_i = 1, \sum_{i=1}^m p_i(\frac{v_i(I_{rc})}{\sqrt{\pi_{22}}} - \zeta_{n2}) = \theta, p_1 \ge 0, \ldots, p_m \ge 0\}.$$

记 $\theta^* = \frac{1}{m}\sum_{i=1}^m (\frac{v_i(I_{rc})}{\sqrt{\pi_{22}}} - \zeta_{n2})$, 可以得到

$$\log L^*(\theta^*) = 0. \tag{2-23}$$

由于 $E\{\frac{v_i(I_{rc})}{\sqrt{\pi_{22}}} - \zeta_{n2}\} = E\{\frac{v_i(I_{rc})}{\sqrt{\pi_{22}}}\} = 0$, 且 $E\{\frac{v_i(I_{rc})}{\sqrt{\pi_{22}}} - \zeta_{n2}\}^2 = 1$, 则由切比雪夫不等式有

$$P(|\theta^*| > m^{-2/5}) \le m^{-4/5} \to 0. \tag{2-24}$$

与证明 $-2\log L_1(I_{rc}) \xrightarrow{p} \infty$ 一样, 可证得 $-2\log L^*(\theta_1^*) \xrightarrow{p} \infty$ 和 $-2\log L^*(\theta_2^*) \xrightarrow{p} \infty$, 其中 $\theta_1^* = m^{-1/4}$ 和 $\theta_2^* = -m^{-1/4}$ 满足 $m(\theta_1^*)^2 = o(m)$ 和 $m(\theta_2^*)^2 = o(m)$. 由于对任意的 c 来说, 集合 $\{\theta : -2\log L^*(\theta) \le c\} =: I_c$ 是凸集, 可取 $c = \min\{-2\log L^*(\theta_1^*), -2\log L^*(\theta_2^*)\}/2$. 根据 (2-23), 可知 $\theta^* \in I_c$. 故若 $-\zeta_{m2} \in I_c$, 则对任意的 $a \in [0,1]$, 有 $(1-a)\theta^* - a\zeta_{m2} \in I_c$. 这意味着 θ_1^* 和 θ_2^* 中其中一个一定属于 I_c, 因此有

$$\begin{aligned} P(|\theta^*| \le m^{-2/5}, -\zeta_{m2} \in I_c) &\le P(\theta_1^* \in I_c \text{ 或 } \theta_2^* \in I_c) \\ &= P(\min\{-2\log L^*(\theta_1^*), -2\log L^*(\theta_2^*)\} = 0) \\ &\to 0. \end{aligned}$$

另一方面, 根据 (2-24) 可得

$$\begin{aligned} P(-2\log L^*(-\zeta_{m2}) > c) &= P(-\zeta_{m2} \notin I_c) \\ &\ge 1 - P(|\theta^*| \le m^{-2/5}, -\zeta_{m2} \in I_c) - P(|\theta^*| > m^{-2/5}) \\ &\to 1, \end{aligned}$$

故

$$-2\log L^*(-\zeta_{m2}) \xrightarrow{p} \infty.$$

又由于

$$c = \min\{-2\log L^*(\theta_1^*), -2\log L^*(\theta_2^*)\}/2 \xrightarrow{p} \infty,$$

故

$$\begin{aligned} P(-2\log L_1(I_{rc}) > \xi_{1-\alpha}) &\geq P(-2\log L^*(-\zeta_{m2}) > \xi_{1-\alpha}) \\ &\to 1. \end{aligned}$$

证毕. ■

§2.4 数值研究

不失一般性, 假定均值矩阵 M 为零矩阵. 分别在正态总体和非正态总体两种情况下对所提出的非参数假设检验的有效性进行评估, 即通过数值模拟给出了不同分布下经验水平和经验功效的大小. 考虑以下分布:

正态分布: 随机矩阵 Z_i 中每一元素 $Z_{iab} \overset{i.i.d.}{\sim} N(0,1)$;

非正态分布: $Z_{iab} = (Z_{iab}^* - 9)/3$, 其中 $Z_{iab}^* \overset{i.i.d.}{\sim} Gamma(9,1)$.

对三元数组 (N,r,c), 模拟中令 $N = 20,40,60,80$; $r = 2,4,6,8,10$; $c = 10,20,40$. 此外, 在模拟求该检验的经验功效时, 对于 Σ_R 和 Σ_C 的选取, 考虑下面三种情况:

(a) Σ_R 和 Σ_C 均为三对角矩阵且对角线元素均为 0.1;

(b) $\Sigma_R = I_r$, Σ_C 为对角矩阵且其前 $c/10$ 的元素等于 2;

(c) $\Sigma_C = I_c, \Sigma_R = 10 \cdot I_r + 0.1 \cdot 1_r 1_r^T$.

在每个模拟方案中, 均设计程序使其重复运行 1000 次, 从而得出名义显著性水平为 0.05 时的经验水平和经验功效, 具体结果如下:

表 2-1 分别在两种分布下求得了该经验似然方法的经验水平, 很明显可以看出, 尽管样本量较小 (比如样本量为 20) 时, 经验水平值偏大一点, 但是, 随着维数 $r \times c$ 和样本量 n 逐渐增大, 这两种分布下的经验水平值均越来越接近于显著性水平 0.05, 这说明本文使用的非参方法从指标经验水平上来看是有效的.

表 2-1 : $H_0 : \Omega = I_{rc}$ $v.s.$ $H_1 : \Omega \neq I_{rc}$ 的经验水平

		r				
n	c	2	4	6	8	10
			正态分布			
20	10	0.076	0.068	0.084	0.069	0.078
	20	0.086	0.083	0.075	0.065	0.072
	40	0.085	0.076	0.081	0.074	0.070
40	10	0.072	0.073	0.067	0.068	0.067
	20	0.071	0.069	0.067	0.063	0.056
	40	0.074	0.072	0.062	0.068	0.065
60	10	0.061	0.058	0.052	0.053	0.047
	20	0.060	0.058	0.057	0.054	0.049
	40	0.053	0.054	0.052	0.051	0.051
80	10	0.057	0.055	0.051	0.050	0.050
	20	0.058	0.051	0.050	0.049	0.046
	40	0.052	0.047	0.046	0.051	0.050
			非正态分布			
20	10	0.081	0.075	0.074	0.079	0.064
	20	0.077	0.082	0.081	0.073	0.063
	40	0.077	0.074	0.078	0.080	0.079
40	10	0.075	0.069	0.065	0.069	0.057
	20	0.071	0.070	0.064	0.061	0.058
	40	0.068	0.067	0.062	0.065	0.057
60	10	0.065	0.065	0.060	0.048	0.052
	20	0.066	0.055	0.051	0.059	0.047
	40	0.062	0.059	0.056	0.053	0.054
80	10	0.063	0.054	0.055	0.047	0.053
	20	0.056	0.052	0.049	0.047	0.052
	40	0.052	0.051	0.052	0.050	0.050

表 2-2 – 2-4 均是对检验的功效进行估计的, 且分别对应于情况 (a)-(c). 其中, 两种分布下的模拟结果均表明, 随着维数 $r \times c$ 和样本量 n 逐渐增大, 经验功效值整体上也逐渐

增大并接近于 1, 且正态分布下的结果要更好一点, 非正态分布下的经验功效值也不错, 这些足以说明本文所提出方法的有效性.

表 2-2 : $H_0 : \Omega = I_{rc}$ $v.s.$ $H_1 : \Omega \neq I_{rc}$ 的经验功效

		r				
n	c	2	4	6	8	10
			正态分布			
20	10	0.930	0.965	0.969	0.972	0.973
	20	0.961	0.964	0.968	0.970	0.974
	40	0.960	0.967	0.969	0.975	0.979
40	10	0.964	0.968	0.974	0.974	0.980
	20	0.964	0.969	0.970	0.977	0.986
	40	0.965	0.966	0.967	0.979	0.988
60	10	0.972	0.978	0.980	0.981	0.984
	20	0.974	0.982	0.984	0.985	0.989
	40	0.977	0.972	0.985	0.986	0.991
80	10	0.980	0.981	0.988	0.989	0.994
	20	0.981	0.982	0.989	0.992	0.996
	40	0.983	0.986	0.985	0.997	0.998
			非正态分布			
20	10	0.922	0.927	0.930	0.919	0.934
	20	0.923	0.929	0.935	0.930	0.935
	40	0.932	0.930	0.933	0.930	0.938
40	10	0.930	0.935	0.934	0.933	0.944
	20	0.937	0.939	0.934	0.936	0.940
	40	0.941	0.946	0.939	0.938	0.943
60	10	0.947	0.951	0.953	0.959	0.954
	20	0.949	0.948	0.949	0.963	0.966
	40	0.947	0.952	0.954	0.950	0.957
80	10	0.945	0.947	0.958	0.962	0.971
	20	0.952	0.956	0.961	0.972	0.969
	40	0.960	0.968	0.977	0.976	0.981

表 2-3 : $H_0 : \Omega = I_{rc}$ $v.s.$ $H_1 : \Omega \neq I_{rc}$ 的经验功效

		r				
n	c	2	4	6	8	10
		正态分布				
20	10	0.918	0.920	0.929	0.932	0.935
	20	0.919	0.922	0.924	0.934	0.936
	40	0.921	0.928	0.936	0.942	0.945
40	10	0.929	0.934	0.937	0.941	0.950
	20	0.934	0.940	0.948	0.945	0.951
	40	0.936	0.943	0.942	0.946	0.950
60	10	0.942	0.946	0.951	0.958	0.949
	20	0.944	0.946	0.946	0.947	0.949
	40	0.947	0.949	0.948	0.950	0.955
80	10	0.949	0.951	0.953	0.956	0.962
	20	0.952	0.954	0.963	0.965	0.966
	40	0.951	0.956	0.971	0.980	0.989
		非正态分布				
20	10	0.927	0.934	0.926	0.933	0.932
	20	0.922	0.927	0.936	0.935	0.933
	40	0.928	0.930	0.934	0.936	0.944
40	10	0.930	0.933	0.946	0.937	0.940
	20	0.935	0.941	0.939	0.939	0.945
	40	0.938	0.943	0.940	0.938	0.941
60	10	0.944	0.952	0.946	0.944	0.949
	20	0.947	0.944	0.949	0.952	0.951
	40	0.949	0.946	0.953	0.957	0.959
80	10	0.951	0.966	0.967	0.971	0.969
	20	0.950	0.957	0.963	0.955	0.968
	40	0.950	0.960	0.969	0.965	0.974

表 2-4 : $H_0:\Omega=I_{rc}$ $v.s.$ $H_1:\Omega\neq I_{rc}$ 的经验功效

		r				
n	c	2	4	6	8	10
		正态分布				
20	10	0.922	0.930	0.943	0.944	0.945
	20	0.925	0.941	0.950	0.950	0.950
	40	0.926	0.947	0.947	0.950	0.958
40	10	0.946	0.947	0.951	0.954	0.956
	20	0.947	0.951	0.956	0.957	0.945
	40	0.947	0.954	0.951	0.957	0.954
60	10	0.951	0.952	0.954	0.958	0.961
	20	0.955	0.957	0.964	0.954	0.966
	40	0.954	0.955	0.958	0.962	0.974
80	10	0.961	0.962	0.966	0.972	0.984
	20	0.964	0.966	0.976	0.984	0.993
	40	0.967	0.967	0.968	0.994	0.996
		非正态分布				
20	10	0.925	0.935	0.941	0.947	0.949
	20	0.924	0.934	0.943	0.950	0.948
	40	0.938	0.942	0.945	0.952	0.951
40	10	0.939	0.942	0.947	0.954	0.952
	20	0.940	0.939	0.943	0.949	0.956
	40	0.940	0.950	0.949	0.953	0.955
60	10	0.944	0.952	0.959	0.949	0.950
	20	0.948	0.957	0.963	0.953	0.958
	40	0.954	0.962	0.969	0.971	0.970
80	10	0.955	0.972	0.956	0.966	0.973
	20	0.947	0.958	0.988	0.967	0.972
	40	0.954	0.961	0.965	0.970	0.975

综上所述, 不论从经验水平值还是经验功效值来说, 本文所使用的经验似然方法都是十分有效的.

本章主要对高维矩阵型数据协差阵的假设检验进行了研究, 在文中主要是通过向量拉直将矩阵型数据转换为向量型数据, 并通过构造估计方程来进行假设检验, 得出似然比统计量渐近服从卡方分布这一结论. 本章不足之处即并未找到合适的案例去分析, 这也是本论文后续的重点任务之一.

第三章　两样本均值线性组合的检验

§3.1　模型介绍

本章考虑固定维两样本均值线性组合的经验似然检验. 假设 $Y_{11},\cdots,Y_{1n_1}$ 和 $Y_{21},\cdots,Y_{2n_2}$ 分别为来自 Y_1 和 Y_2 的两组简单随机样本, 样本量分别为 n_1 和 n_2, 且 $E(Y_1)=\mu_1$, $E(Y_2)=\mu_2, Var(Y_1)=\sigma_1^2, Var(Y_2)=\sigma_2^2$. 下面将考虑如下检验问题

$$H_0: a\mu_1-b\mu_2=\theta. \quad v.s. \quad H_1: \ not\ H_0.$$

这里 θ 为感兴趣的参数, a 和 b 是任意不为零的常数, $n_1+n_2=n$ 为组合样本量.

在接下来的研究中, 考虑将 Wu 和 Yan[5] 对两样本均值相等性的检验拓展到两样本均值的线性组合, 分别用几种不同的经验似然方法对感兴趣的参数 θ 进行检验, 证得检验统计量在 $n\to\infty$ 时渐近服从卡方分布.

§3.2　标准两样本经验似然

令 $Y_{11},\cdots,Y_{1n_1}$ 和 $Y_{21},\cdots,Y_{2n_2}$ 分别为来自 Y_1 和 Y_2 的两独立同分布样本, 样本量分别为 n_1 和 n_2, 且 $E(Y_1)=\mu_1, E(Y_2)=\mu_2, Var(Y_1)=\sigma_1^2$, $Var(Y_2)=\sigma_2^2$. 令 $\theta=a\mu_1-b\mu_2$, a 和 b 是任意不为零的常数. 下面基于两样本给出 θ 的联合经验对数似然函数

$$l(\mathbf{p_1},\mathbf{p_2})=\sum_{j=1}^{n_1}\log(p_{1j})+\sum_{j=1}^{n_2}\log(p_{2j}),$$

其中 $\mathbf{p_1}=(p_{11},\cdots,p_{1n_1})^T$ 和 $\mathbf{p_2}=(p_{21},\cdots,p_{2n_2})^T$ 分别为两样本的概率测度集. θ 的经验对数似然比统计量定义为

$$r(\theta)=\sum_{j=1}^{n_1}\log\{n_1\hat{p}_{1j}(\theta)\}+\sum_{j=1}^{n_2}\log\{n_2\hat{p}_{2j}(\theta)\},$$

这里 $\hat{p}_{1j}(\theta)$ 和 $\hat{p}_{2j}(\theta)$ 为最大化 $l(\mathbf{p_1},\mathbf{p_2})$ 使其满足如下约束形式的解

$$\sum_{j=1}^{n_1}p_{1j}=1,\quad \sum_{j=1}^{n_2}p_{2j}=1, \tag{3-1}$$

$$a\sum_{j=1}^{n_1}p_{1j}Y_{1j}-b\sum_{j=1}^{n_2}p_{2j}Y_{2j}=\theta. \tag{3-2}$$

其中 $\theta = aE(Y_1) - bE(Y_2)$.

引理3.1.1 假设 Y_i 为有共同分布的独立随机变量, $i = 1, \cdots, n$. $E(Y_i^2) < \infty$. 令 $Z_n = \max_{1\le i\le n}|Y_i|$, 则 $Z_n = o(n^{\frac{1}{2}})$.

该引理的详细证明参见 Owen[1]中的引理 11.3.

定理3.1.1 假设 $\sigma_1^2 < \infty, \sigma_2^2 < \infty$, 且 $\frac{n_1}{n} \to \pi, \pi \in (0,1)$, $n \to \infty$. 则当 $n \to \infty$ 时, $-2r(\theta) \xrightarrow{d} \chi_1^2$.

定理 3.1.1 的证明: 首先, 根据 $\frac{n_1}{n} \to \pi \in (0,1)$, 可知不需要区分 $O_p(n_1^{-\frac{1}{2}}), O_p(n_2^{-\frac{1}{2}})$ 和 $O_p(n^{-\frac{1}{2}})$. 下面令 μ_0 为一固定的数且 $\mu_0 = \mu_2 + O(n^{-\frac{1}{2}})$. 用下式来代替约束条件 (3-2)

$$a\sum_{j=1}^{n_1} p_{1j}Y_{ij} = \mu_0 + \theta, \quad b\sum_{j=1}^{n_2} p_{2j}Y_{2j} = \mu_0. \tag{3-3}$$

则约束条件 (3-3) 为 (3-2) 的充分不必要条件. 原本的 $r(\theta)$ 转化为 $r(\hat{\mu}_0, \theta)$, 这里 $\hat{\mu}_0$ 是经验对数似然比统计量 $r(\mu_0, \theta)$ 关于 μ_0 的最大点 (对固定的 θ). 且 μ_0 为讨厌参数.

通过拉格朗日乘子法, 最大化 $l(\mathbf{p_1}, \mathbf{p_2})$ 使其满足 (3-1) 和 (3-3) 可得 $\tilde{p}_{1j}$ 和 $\tilde{p}_{2j}$ 如下

$$\begin{cases} \tilde{p}_{1j} = \dfrac{1}{n_1\{1 + \lambda_1(aY_{1j} - \mu_0 - \theta)\}} \\ \tilde{p}_{2j} = \dfrac{1}{n_2\{1 + \lambda_2(bY_{2j} - \mu_0)\}}. \end{cases}$$

这里 λ_1 和 λ_2 满足如下两个方程

$$\begin{cases} \dfrac{1}{n_1}\displaystyle\sum_{j=1}^{n_1} \dfrac{aY_{1j} - \mu_0 - \theta}{1 + \lambda_1(aY_{1j} - \mu_0 - \theta)} = 0 \\ \dfrac{1}{n_2}\displaystyle\sum_{j=1}^{n_2} \dfrac{bY_{2j} - \mu_0}{1 + \lambda_2(bY_{2j} - \mu_0)} = 0. \end{cases}$$

则对应的经验对数似然比统计量为

$$r(\mu_0, \theta) = -\sum_{j=1}^{n_1} log\{1 + \lambda_1(aY_{1j} - \mu_0 - \theta)\} - \sum_{j=1}^{n_2} log\{1 + \lambda_2(bY_{2j} - \mu_0)\}.$$

为求得 $\hat{\mu}_0$, 令 $\partial r(\mu_0, \theta)/\partial \mu_0 = 0$, 可得

$$n_1\lambda_1 + n_2\lambda_2 = 0. \tag{3-4}$$

根据 Owen[1] 中的定理 3.2 知

$$\begin{cases} \lambda_1 = \dfrac{1}{a^2(\sigma_1)^2}(a\bar{Y_1} - \mu_0 - \theta) + o_p(n^{-\frac{1}{2}}) \\ \lambda_2 = \dfrac{1}{b^2(\sigma_2)^2}(b\bar{Y_2} - \mu_0) + o_p(n^{-\frac{1}{2}}), \end{cases} \tag{3-5}$$

其中 $\bar{Y}_1=\frac{1}{n_1}\sum_{j=1}^{n_1}Y_{1j}$, $\bar{Y}_2=\frac{1}{n_2}\sum_{j=1}^{n_2}Y_{2j}$.

将 (3-5) 代入 (3-4) 可得 (3-5) 的解为

$$\hat{\mu}_0=\gamma(a\bar{Y}_1-\theta)+(1-\gamma)b\bar{Y}_2+o_p(n^{-\frac{1}{2}}),$$

其中

$$\gamma=(\frac{n_1}{a^2\sigma_1^2})/(\frac{n_1}{a^2\sigma_1^2}+\frac{n_2}{b^2\sigma_2^2}).$$

注意到 $\lambda_1(aY_{1j}-\hat{\mu}_0-\theta)=o_p(1)$, $\lambda_2(bY_{2j}-\hat{\mu}_0)=o_p(1)$. 因此, 在 $\mu_0=\hat{\mu}_0$ 处对 $-2r(\theta)$ 进行泰勒展开有

$$\begin{aligned}-2r(\theta)&=-2r(\hat{\mu}_0,\theta)\\&=2\sum_{j=1}^{n_1}\log\{1+\lambda_1(aY_{1j}-\hat{\mu}_0-\theta)\}+2\sum_{j=1}^{n_2}\log\{1+\lambda_2(bY_{2j}-\hat{\mu}_0)\},\end{aligned}$$

对第一项泰勒展开并将 λ_1 代入得

$$\begin{aligned}&2\sum_{j=1}^{n_1}\log\{1+\lambda_1(aY_{1j}-\hat{\mu}_0-\theta)\}\\=\;&2\sum_{j=1}^{n_1}[\lambda_1(aY_{1j}-\hat{\mu}_0-\theta)]-\sum_{j=1}^{n_1}[\lambda_1(aY_{1j}-\hat{\mu}_0-\theta)]^2+2\sum_{j=1}^{n_1}\eta_{1j}\\=\;&2n_1\lambda_1(a\bar{Y}_1-\hat{\mu}_0-\theta)-n_1a^2\lambda_1^2\sigma_1^2+2\sum_{j=1}^{n_1}\eta_{1j}\\=\;&\frac{n_1}{a^2\sigma_1^2}(a\bar{Y}_1-\hat{\mu}_0-\theta)^2-o_p(1)+2\sum_{j=1}^{n_1}\eta_{1j},\end{aligned}$$

其中 $\eta_{1j}=o([\lambda_1(aY_{1j}-\hat{\mu}_0-\theta)]^2)$.

对某个有限的 $B>0$, 有

$$P(|\eta_{1j}|\le B\cdot|\lambda_1(aY_{1j}-\hat{\mu}_0-\theta)|^3,1\le j\le n_1)\to 1,n_1\to\infty,$$

则

$$\begin{aligned}|\sum_{j=1}^{n_1}\eta_{1j}|&\le B\cdot\|\lambda_1\|^3\sum_{j=1}^{n_1}\|aY_{1j}-\hat{\mu}_0-\theta\|^3\\&=O_p(n_1^{-\frac{3}{2}})o_p(n_1^{\frac{3}{2}})\\&=o_p(1).\end{aligned}$$

故第一项即

$$
\begin{aligned}
& 2\sum_{j=1}^{n_1}\log\{1+\lambda_1(aY_{1j}-\hat{\mu}_0-\theta)\} \\
= & \frac{n_1}{a^2\sigma_1^2}(a\bar{Y}_1-\hat{\mu}_0-\theta)^2+o_p(1).
\end{aligned}
$$

同样地，第二项为

$$
\begin{aligned}
& 2\sum_{j=1}^{n_2}\log\{1+\lambda_2(bY_{2j}-\hat{\mu}_0)\} \\
= & \frac{n_2}{b^2\sigma_2^2}(b\bar{Y}_2-\hat{\mu}_0)^2+o_p(1),
\end{aligned}
$$

故将 $\hat{\mu}_0$ 代入 $-2r(\theta)$ 有

$$
\begin{aligned}
-2r(\theta) &= \frac{n_1}{a^2\sigma_1^2}(a\bar{Y}_1-\hat{\mu}_0-\theta)^2+\frac{n_2}{b^2\sigma_2^2}(b\bar{Y}_2-\hat{\mu}_0)^2+o_p(1) \\
&= (a\bar{Y}_1-b\bar{Y}_2-\theta)^2/(\frac{a^2\sigma_1^2}{n_1}+\frac{b^2\sigma_2^2}{n_2})+o_p(1) \\
&\xrightarrow{d} \chi_1^2.
\end{aligned}
$$

证毕. ■

§3.3 加权两样本经验似然

目前已有文献中，有许多针对不同问题对经验似然函数的加权方法，本节沿用 Wu 和 Yan[5] 中的加权方法，考虑将权加在每一个样本上，即所加权与样本量有关，该方法最早由 Fu, Wang 和 Wu[27] 在对多重样本的推断中提出.

首先，定义加权对数似然函数如下

$$
l_\omega(\mathbf{p_1},\mathbf{p_2})=\frac{\omega_1}{n_1}\sum_{j=1}^{n_1}\log(p_{1j})+\frac{\omega_2}{n_2}\sum_{j=1}^{n_2}\log(p_{2j}),
$$

这里 $\omega_1=\omega_2=\frac{1}{2}$. 定义加权经验对数似然比统计量为

$$
r_\omega(\theta)=\frac{\omega_1}{n_1}\sum_{j=1}^{n_1}\log\{n_1\hat{p}_{1j}(\theta)\}+\frac{\omega_2}{n_2}\sum_{j=1}^{n_2}\log\{n_2\hat{p}_{2j}(\theta)\}.
$$

其中 $\hat{p}_{1j}(\theta)$ 和 $\hat{p}_{2j}(\theta)$ 是最大化 $l_\omega(\mathbf{p_1},\mathbf{p_2})$ 使其满足约束条件 (3-1) 和 (3-2) 的解.

定理3.2.1 假设 $\sigma_1^2<\infty,\sigma_2^2<\infty$, 且 $\frac{n_1}{n}\to\pi,\pi\in(0,1)$, $n\to\infty$. 则当 $n\to\infty$ 时, $-2r_\omega(\theta)\xrightarrow{d}\chi_1^2$.

定理 3.2.1 的证明: 首先, 对规范化约束条件 (3-1) 和 (3-2) 重新构造为其替代形式.

由 (3-1) 知

$$\sum_{i=1}^{2}\omega_i\sum_{j=1}^{n_i}p_{ij}=1. \tag{3-6}$$

由 (3-2) 知

$$a\omega_1\sum_{j=1}^{n_1}p_{1j}(Y_{1j}/\omega_1)-b\omega_2\sum_{j=1}^{n_2}p_{2j}(Y_{2j}/\omega_2)=\theta. \tag{3-7}$$

由 (3-6) 和 (3-7) 有

$$\omega_1\sum_{j=1}^{n_1}p_{1j}(aY_{1j}/\omega_1-\theta)+\omega_2\sum_{j=1}^{n_2}p_{2j}(-bY_{2j}/\omega_2-\theta)=0. \tag{3-8}$$

根据 (3-5) 及 $\omega_1=\omega_2=\frac{1}{2}$ 得

$$\omega_1\sum_{j=1}^{n_1}p_{1j}(1-\omega_1)+\omega_2\sum_{j=1}^{n_2}p_{2j}(-\omega_1)=0. \tag{3-9}$$

则综合 (3-8) 和 (3-9) 知

$$\sum_{i=1}^{2}\omega_i\sum_{j=1}^{n_i}p_{ij}\mathbf{u}_{ij}=0. \tag{3-10}$$

其中 $\mathbf{u}_{ij}=\mathbf{z}_{ij}-\eta$, $i=1,2$, 且 $\mathbf{z}_{1j}=(1,aY_{1j}/\omega_1)^T$, $\mathbf{z}_{2j}=(0,-bY_{2j}/\omega_2)^T$, $\eta=(\omega_1,\theta)^T$.

接下来, 用拉格朗日乘子法在约束条件 (3-6) 和 (3-10) 下求解 $\max\{l_\omega(\mathbf{p_1},\mathbf{p_2})\}$ 得

$$p_{ij}=\frac{1}{n_i(1+\lambda^T\mathbf{u}_{ij})},$$

其中 λ 满足如下方程

$$g_1(\lambda)=\sum_{i=1}^{2}\frac{\omega_i}{n_i}\sum_{j=1}^{n_i}\frac{\mathbf{u}_{ij}}{1+\lambda^T\mathbf{u}_{ij}}=0. \tag{3-11}$$

将 $\frac{1}{1+\lambda^T\mathbf{u}_{ij}}=1-\frac{\lambda^T\mathbf{u}_{ij}}{1+\lambda^T\mathbf{u}_{ij}}$ 代入 (3-11) 即得

$$\left(\sum_{i=1}^{2}\frac{\omega_i}{n_i}\sum_{j=1}^{n_i}\frac{\mathbf{u}_{ij}\mathbf{u}_{ij}^T}{1+\lambda^T\mathbf{u}_{ij}}\right)\lambda=\sum_{i=1}^{2}\frac{\omega_i}{n_i}\sum_{j=1}^{n_i}\mathbf{u}_{ij}.$$

将 $\mathbf{u}_{1j},\mathbf{u}_{2j}$ 代入 $\sum_{i=1}^{2}\frac{\omega_i}{n_i}\sum_{j=1}^{n_i}\mathbf{u}_{ij}$ 有

$$\mathrm{U}\triangleq\sum_{i=1}^{2}\frac{\omega_i}{n_i}\sum_{j=1}^{n_i}\mathbf{u}_{ij}=(0,a\bar{Y}_1-b\bar{Y}_2-\theta)^T.$$

根据 Owen[1] 中类似证明可得 $\lambda = \mathbf{D}^{-1}\mathbf{U} + o_p(n^{-1/2})$. 这里 $\mathbf{D} = \sum_{i=1}^{2}\frac{\omega_i}{n_i}\sum_{j=1}^{n_i}\mathbf{u}_{ij}\mathbf{u}_{ij}^T$ 为 2×2 矩阵.

故对 θ 的加权经验对数似然比统计量泰勒展开, 并将 λ 和 $\mathbf{D}$ 代入有

$$\begin{aligned}
-2r_\omega(\theta) &= 2\sum_{i=1}^{2}\frac{\omega_i}{n_i}\sum_{j=1}^{n_i}\log(1+\lambda^T\mathbf{u}_{ij}) \\
&= 2\sum_{i=1}^{2}\frac{\omega_i}{n_i}\sum_{j=1}^{n_i}(\lambda^T\mathbf{u}_{ij} - \frac{1}{2}\lambda^T\mathbf{u}_{ij}\mathbf{u}_{ij}^T\lambda) + o_p(n^{-1}) \\
&= \mathbf{U}^T\mathbf{D}^{-1}\mathbf{U} + o_p(n^{-1}) \\
&= d^{(22)}(a\bar{Y}_1 - b\bar{Y}_2 - \theta)^2 + o_p(n^{-1}).
\end{aligned}$$

其中 $d^{(22)}$ 为 $\mathbf{D}^{-1}$ 的第二个对角元素.

令 $c_1 = d^{(22)}(a^2\cdot\frac{S_1^2}{n_1} + b^2\cdot\frac{S_2^2}{n_2})$, 则

$$-2r_\omega(\theta)/c_1 \xrightarrow{d} \chi_1^2.$$

这里 $\theta = aEY_1 - bEY_2$. 证毕. ■

注3.2.1 当 θ 用 $\hat{\theta} = a\bar{Y}_1 - b\bar{Y}_2$ 估计时, 比例常数 c_1 的一致估计可以用 $\hat{c}_1$ 表示, 此时仍有 $-2r_\omega(\theta)/\hat{c}_1 \xrightarrow{d} \chi_1^2$. 且 θ 的 $1-\alpha$ 水平的置信区间为

$$\mathcal{C}_2 = \{\theta | \frac{-2r_\omega(\theta)}{\hat{c}_1} \le \chi_1^2(\alpha)\}.$$

§3.4 复杂调查数据的伪经验似然方法

伪经验似然方法最早由 Chen 和 Sitter[48] 提出, 后来 Wu 和 Rao[15] 用来基于单个复杂调查样本进行假设检验或构造置信区间, 本节考虑用该方法对参数 θ 进行统计推断. 令 s_i 为从第 i 个有限总体中选取的样本单元集, n_i 为第 i 个总体的样本量, $i=1,2$. 记 $\pi_j^{(i)} = P(j\in s_i)$ 为包含概率, $j=1,\cdots,n_i, i=1,2$. 令 $d_{ij} = \frac{1}{\pi_j^{(i)}}$ 为基本设计权重, 且

$$\tilde{d}_{ij}(s_i) = d_{ij}/\sum_{k\in s_i} d_{ik}$$

为标准化设计权重. 同样地, 定义两样本伪经验似然函数

$$l_{pel}(\mathbf{p_1},\mathbf{p_2}) = \sum_{i=1}^{2}\omega_i\sum_{j=1}^{n_i}\tilde{d}_{ij}(s_i)\log(p_{ij}),$$

权重 $\omega_1 = \omega_2 = \frac{1}{2}$. 对给定总体 i, 当包含概率都相等时, $\tilde{d}_{ij}(s_i)$ 即转化为 $\frac{1}{n_i}$, 这样, $l_{pel}(\mathbf{p_1}, \mathbf{p_2})$ 即化为上节的加权经验似然函数 $l_\omega(\mathbf{p_1}, \mathbf{p_2})$.

定理3.3.1 在 Wu 和 Rao[24] 的渐近性工作和正则性条件 $C1 - C3$ 下, 假定 $\frac{n_1}{n} \to \pi$, $\pi \in (0, 1)$, $n \to \infty$. 则当 $n \to \infty$ 时, 调整的两样本伪经验似然比统计量 $-2r_{pel}(\theta)/c_2 \xrightarrow{d} \chi_1^2$.

定理 3.3.1 的证明: 最大化 $l_{pel}(\mathbf{p_1}, \mathbf{p_2})$ 使其满足约束条件 (3-1) 和 (3-2) 得

$$\hat{p}_{ij}(\theta) = \frac{\tilde{d}_{ij}(s_i)}{1 + \lambda^T \mathbf{u}_{ij}},$$

λ 为下式的解

$$g_2(\lambda) = \sum_{i=1}^{2} \omega_i \sum_{j=1}^{n_i} \frac{\tilde{d}_{ij}(s_i)\mathbf{u}_{ij}}{1 + \lambda^T \mathbf{u}_{ij}} = 0. \tag{3-12}$$

注意到

$$\hat{p}_{ij}(\hat{\theta}_{pel}) = \tilde{d}_{ij}(s_i),$$

其中

$$\hat{\theta}_{pel} = a \sum_{j=1}^{n_i} \hat{p}_{1j} Y_{1j} - b \sum_{j=1}^{n_2} \hat{p}_{2j} Y_{2j} = a\hat{\mu}_1 - b\hat{\mu}_2$$

为 θ 的最大伪经验似然估计, $\hat{\mu}_i = \sum_{j=1}^{n_i} \tilde{d}_{ij}(s_i) Y_{ij} = (\sum_{j=1}^{n_i}) d_{ij} Y_{ij} / \sum_{j=1}^{n_i}) d_{ij}$, 则 θ 的伪经验似然比统计量为

$$\begin{aligned}
r_{pel}(\theta) &= l_{pel}\{\hat{p}_1(\theta), \hat{p}_2(\theta)\} - l_{pel}\{\hat{p}_1(\hat{\theta}_{pel}), \hat{p}_2(\hat{\theta}_{pel})\} \\
&= \sum_{i=1}^{2} \omega_i \sum_{j=1}^{n_i} \tilde{d}_{ij}(s_i) \log \frac{\tilde{d}_{ij}(s_i)}{1 + \lambda^T \mathbf{u}_{ij}} - \sum_{i=1}^{2} \omega_i \sum_{j=1}^{n_i} \tilde{d}_{ij}(s_i) \log[\tilde{d}_{ij}(s_i)] \\
&= \sum_{i=1}^{2} \omega_i \sum_{j=1}^{n_i} \tilde{d}_{ij}(s_i) \log \frac{1}{1 + \lambda^T \mathbf{u}_{ij}} \\
&= -\sum_{i=1}^{2} \omega_i \sum_{j=1}^{n_i} \tilde{d}_{ij}(s_i) \log(1 + \lambda^T \mathbf{u}_{ij}).
\end{aligned}$$

由 (3-12) 有

$$\sum_{i=1}^{2} \omega_i \sum_{j=1}^{n_i} \tilde{d}_{ij}(s_i) \mathbf{u}_{ij} (1 - \frac{\lambda^T \mathbf{u}_{ij}}{1 + \lambda^T \mathbf{u}_{ij}}) = 0,$$

则

$$(\sum_{i=1}^{2}\omega_i\sum_{j=1}^{n_i}\frac{\tilde{d}_{ij}(s_i)\mathbf{u}_{ij}\mathbf{u}_{ij}^T}{1+\lambda^T\mathbf{u}_{ij}})\lambda=\sum_{i=1}^{2}\omega_i\sum_{j=1}^{n_i}\tilde{d}_{ij}(s_i)\mathbf{u}_{ij}.$$

解得

$$\lambda=\mathbf{K}^{-1}\mathbf{U}+o_p(n^{-1/2}),$$

其中

$$\mathbf{K}=\sum_{i=1}^{2}\omega_i\sum_{j=1}^{n_i}\tilde{d}_{ij}(s_i)\mathbf{u}_{ij}\mathbf{u}_{ij}^T$$

为 2×2 矩阵, $\mathbf{U}$ 形式如下

$$\mathbf{U}=\sum_{i=1}^{2}\omega_i\sum_{j=1}^{n_i}\tilde{d}_{ij}(s_i)\mathbf{u}_{ij}=\begin{pmatrix}0\\ a(\hat{\mu}_1-\mu_1)-b(\hat{\mu}_2-\mu_2)\end{pmatrix}.$$

同样地, 对 $-2r_{pel}(\theta)$ 泰勒展开并将 λ 代入得

$$-2r_{pel}(\theta)=k^{(22)}\{a\sum_{j=1}^{n_1}\tilde{d}_{1j}(s_1)Y_{1j}-b\sum_{j=1}^{n_2}\tilde{d}_{2j}(s_2-\theta)^2\}+o_p(1).$$

令

$$c_2=k^{(22)}\{a^2\cdot V(\sum_{j=1}^{n_1}\tilde{d}_{1j}(s_1)Y_{1j})+b^2\cdot V(\sum_{j=1}^{n_2}\tilde{d}_{2j}(s_2)Y_{2j})\},$$

其中 $k^{(22)}$ 为 $\mathbf{K}^{-1}$ 的第二个对角元素, $V(\cdot)$ 为基于设计的方差. 故

$$-2r_{pel}(\theta)/c_2\xrightarrow{d}\chi_1^2.$$

令 $\hat{c}_2$ 为 c_2 的一致估计量, 则 θ 的 $1-\alpha$ 水平的置信区间为

$$C_3=\{\theta|\frac{-2r_{pel}(\theta)}{\hat{c}_2}\le\chi_1^2(\alpha)\}.$$

证毕. ■

§3.5 带缺失数据的两样本经验似然

本节考虑在响应变量缺失的条件下去讨论两样本均值线性组合的假设检验. 沿用 Wu 和 Yan[5] 中对模型的假设, 假定总体 Y_i 未知, $\{(Y_{ij}, \mathbf{x}_{ij}), j=1,\dots,n_i\}$ 是来自第 i 个总体的随机样本, $i=1,2$, 响应变量 Y_{ij} 服从如下线性回归模型

$$Y_{ij} = \mathbf{x}_{ij}^T\beta_i + \epsilon_{ij}, \tag{3-13}$$

其中 $i=1,2$, β_1 和 β_2 为两总体的回归参数, 且 ϵ_{ij} 是均值为 0, 方差未知记为 τ_i^2 的独立误差项, 响应变量 Y_{ij} 随机缺失. 令

$$\delta_{ij} = \begin{cases} 1, & Y_{ij} \text{ 能被观测到} \\ 0, & Y_{ij} \text{ 缺失} \end{cases}$$

感兴趣的参数仍为 $\theta = a\mu_1 - b\mu_2$, $\mu_i = E(Y_{ij})$ 为第 i 个总体的均值.

令

$$\tilde{\beta}_i = (\sum_{j=1}^{n_i}\delta_{ij}\mathbf{x}_{ij}\mathbf{x}_{ij}^T)^{-1}\sum_{j=1}^{n_i}\delta_{ij}\mathbf{x}_{ij}Y_{ij}$$

为 β_i 的普通最小二乘估计量, 再令

$$\tilde{Y}_{ij} = \delta_{ij}Y_{ij} + (1-\delta_{ij})\mathbf{x}_{ij}^T\tilde{\beta}_i,$$

其中 $j=1,\dots,n_i$, $i=1,2$, 且用标准两样本经验似然中定义统计量 $r(\theta)$ 的方法去定义本节的检验统计量 $\tilde{r}(\theta)$, 将约束条件 (3-2) 替换为

$$a\sum_{j=1}^{n_1}p_{1j}\tilde{Y}_{1j} - b\sum_{j=1}^{n_2}p_{2j}\tilde{Y}_{2j} = \theta, \tag{3-14}$$

则当 $\delta_{ij}=1$ 时, $\tilde{Y}_{1j} = Y_{ij}$,当 $\delta_{ij}=0$ 时, $\tilde{Y}_{1j} = \mathbf{x}_{ij}^T\tilde{\beta}_i$.

定理3.4.1 假设回归模型 (3-13) 成立, 独立误差项的方差 τ_i^2 有限. 假设 $E(\|\mathbf{X}_i\|^2) < \infty$, 当 $n\to\infty$ 时, $\frac{n_1}{n}\to\pi\in(0,1)$, 则当 $n\to\infty$ 时, 统计量 $-2\tilde{r}(\theta)/c_3 \xrightarrow{d} \chi_1^2$, c_3 为比例常数.

定理 3.4.1 的证明: 令 μ_{i0} 为真实参数, μ_i 在 μ_{i0} 附近, 且满足 $\mu_i = \mu_{i0} + O(n_i^{-1/2})$.

定义

$$r(\mu_2,\theta) = \sum_{j=1}^{n_1}\log(n_1p_{1j}) + \sum_{j=1}^{n_2}\log(n_2p_{2j}),$$

θ 的联合经验对数似然函数仍为 $l(\mathbf{p_1},\mathbf{p_2})=\sum_{j=1}^{n_1}\log(p_{1j})+\sum_{j=1}^{n_2}\log(p_{2j})$.

易知

$$\sum_{j=1}^{n_1}p_{1j}(a\tilde{Y}_{1j}-b\mu_2-\theta)=0 \tag{3-15}$$

和

$$\sum_{j=1}^{n_2}p_{2j}\cdot b(\tilde{Y}_{2j}-\mu_2)=0 \tag{3-16}$$

是 (3-14) 的充分不必要条件, 则根据拉格朗日乘子法, 最大化 $l(\mathbf{p_1},\mathbf{p_2})$ 使其满足 (3-1), (3-15) 和 (3-16) 得

$$\begin{cases} p_{1j}=\dfrac{1}{n_1[1+\beta_1^T(a\tilde{Y}_{1j}-b\mu_2-\theta)]} \\ p_{2j}=\dfrac{1}{n_2[1+\beta_2^T\cdot b(\tilde{Y}_{2j}-\mu_2)]}. \end{cases}$$

其中 β_1 和 β_2 分别为下列两式的解

$$\begin{cases} \dfrac{1}{n_1}\sum_{j=1}^{n_1}\dfrac{a\tilde{Y}_{1j}-b\mu_2-\theta}{1+\beta_1^T(a\tilde{Y}_{1j}-b\mu_2-\theta)}=0 \\ \dfrac{1}{n_2}\sum_{j=1}^{n_2}\dfrac{b(\tilde{Y}_{2j}-\mu_2)}{1+\beta_2^T\cdot b(\tilde{Y}_{2j}-\mu_2)}=0. \end{cases}$$

则

$$r(\mu_2,\theta)=-\sum_{j=1}^{n_1}\log[1+\beta_1^T(a\tilde{Y}_{1j}-b\mu_2-\theta)]-\sum_{j=1}^{n_2}\log[1+\beta_2^T\cdot b(\tilde{Y}_{2j}-\mu_2)]. \tag{3-17}$$

令 $\frac{\partial r(\mu_2,\theta)}{\partial\mu_2}=0$ 得

$$n_1\beta_1+n_2\beta_2=0. \tag{3-18}$$

此外, 由 Wang 和 Rao[19] 可知

$$\begin{aligned} \frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1}(a\tilde{Y}_{1j}-b\mu_2-\theta) &\xrightarrow{d} N(0,V_1), \\ \frac{1}{n_1}\sum_{j=1}^{n_1}(a\tilde{Y}_{1j}-b\mu_2-\theta)^2 &\xrightarrow{p} U_1, \\ \max_{1\le j\le n_i}|\tilde{Y}_{ij}| &= o_p(n_i^{1/2}), \\ \beta_i &= O_p(n_i^{-1/2}). \end{aligned}$$

其中

$$
\begin{aligned}
U_1 &= a^2(S_{11}+\beta_1^T S_{14}\beta_1) - 2aS_{15}^T\beta_1(b\mu_2+\theta) + (b\mu_2+\theta)^2, \\
U_2 &= b^2(S_{21}+\beta_2^T S_{24}\beta_2) + b^2({\mu_2}^2 - 2S_{25}^T\beta_2\mu_2), \\
V_1 &= a^2(S_{11}+S_{12}^T S_{13}^{-1}S_{12}\tau_1^2 + \beta_1^T S_{14}\beta_1 + 2S_{12}^T S_{13}^{-1}S_{16}\tau_1^2) \\
&\quad -2aS_{15}^T\beta_1(b\mu_2+\theta) + (b\mu_2+\theta)^2, \\
V_2 &= b^2(S_{21}+\beta_2^T S_{24}\beta_2) + b^2({\mu_2}^2 - 2S_{25}^T\beta_2\mu_2), \\
S_{i1} &= E[\delta_{ij}(Y_{ij}-x_{ij}^T\beta_i)^2],\ \ S_{i2} = E[(1-\delta_{ij})x_{ij}],\ \ S_{i3} = E(\delta_{ij}x_{ij}x_{ij}^T), \\
S_{i4} &= E(x_{ij}x_{ij}^T),\ \ S_{i5} = E(x_{ij}),\ \ S_{i6} = E(\delta_{ij}x_{ij}),\ \ i=1,2\,;\ \ j=1,\ldots,n_i.
\end{aligned}
$$

故

$$
\beta_1 = [\sum_{j=1}^{n_1}(a\tilde{Y}_{1j} - b\mu_2 - \theta)^2]^{-1}\sum_{j=1}^{n_1}(a\tilde{Y}_{1j} - b\mu_2 - \theta) + o_p(n_1^{-/2}), \tag{3-19}
$$

$$
\beta_2 = [\sum_{j=1}^{n_2}(b(\tilde{Y}_{2j} - \mu_2))^2]^{-1}\sum_{j=1}^{n_2} b(\tilde{Y}_{2j} - \mu_2) + o_p(n_2^{-/2}). \tag{3-20}
$$

将 (3-19) 和 (3-20) 代入 (3-18) 可得 (3-18) 的解 $\tilde{\mu}_2$ 满足

$$
b\tilde{\mu}_2 = \alpha(a\tilde{Y}_{1\cdot} - \theta) + (1-\alpha)\cdot b\tilde{Y}_{2\cdot} + o_p(n^{-/2}),
$$

其中 $\alpha = (\frac{n_1}{U_1})/(\frac{n_1}{U_1} + \frac{n_2}{U_2})$. 则

$$
\begin{aligned}
b(\tilde{\mu}_2 - \mu_{20}) &= \alpha(a\tilde{Y}_{1\cdot} - b\mu_{20} - \theta) + (1-\alpha)\cdot b(\tilde{Y}_{2\cdot} - \mu_{20}) + o_p(n^{-/2}) \\
&= \alpha\cdot a(\tilde{Y}_{1\cdot} - \mu_{10}) + (1-\alpha)\cdot b(\tilde{Y}_{2\cdot} - \mu_{20}) + o_p(n^{-/2}) \\
&= O_p(n^{-/2}).
\end{aligned}
$$

注意到 $\beta_1(a\tilde{Y}_{1\cdot} - b\mu_{20} - \theta) = o_p(n^{-/2})$, $\beta_2\cdot b(\tilde{Y}_{2\cdot} - \mu_{20}) = o_p(n^{-/2})$, 故对 (3-17) 在 $\mu_2 = \tilde{\mu}_2$ 处渐近展开得

$$
\begin{aligned}
-2\tilde{r}(\theta) &= -2r(\tilde{\mu}_2, \theta) \\
&= 2\sum_{j=1}^{n_1}\log[1+\beta_1^T(a\tilde{Y}_{1j} - b\mu_2 - \theta)] + 2\sum_{j=1}^{n_2}\log[1+\beta_2^T\cdot b(\tilde{Y}_{2j} - \mu_2)] \\
&= \frac{n_1}{U_1}(a\tilde{Y}_{1\cdot} - b\tilde{\mu}_{20} - \theta)^2 + \frac{n_2}{U_2}[b(\tilde{Y}_{2\cdot} - \tilde{\mu}_{20})]^2 + o_p(1) \\
&= (a\tilde{Y}_{1\cdot} - b\tilde{Y}_{2\cdot} - \theta)^2/(\frac{U_1}{n_1} + \frac{U_2}{n_2}) + o_p(1).
\end{aligned}
$$

令 $c_3=(\frac{V_1}{n_1}+\frac{V_2}{n_2})/(\frac{U_1}{n_1}+\frac{U_2}{n_2})$, 则

$$-2\tilde{r}(\theta)\xrightarrow{d}\ \chi_1^2.$$

证毕. ■

§3.6 数值研究

在本节中, 为了检验并比较标准经验似然 (EL)、加权经验似然 (WEL)、自助校准经验似然 (BWEL) 这三种经验似然方法的优劣, 本文采用自助法进行数值模拟, 求出了置信水平为 0.95 的 $\theta=a\mu_1-b\mu_2$ 的置信区间 ($a=1, b=3$), 具体自助程序如下:

(1) 从原始样本 s_i 中用简单随机抽样方法抽取大小为 n_i 的自助样本 s_i^*, 用 $\{(d_{ij}^*,Y_{ij}^*), j\in s_i^*; i=1,2\}$ 表示自助样本数据;

(2) 定义自助版本的两样本伪经验似然函数

$$l_{pel}^*(p_1,p_2)=\omega_1\sum_{j\in s_1^*}\tilde{d}_{1j}^*(s_1^*)\log(p_{1j})+\omega_2\sum_{j\in s_2^*}\tilde{d}_{2j}^*(s_2^*)\log(p_{2j}),$$

其中 $\omega_1=\omega_2=1/2, \tilde{d}_{1j}^*(s_1^*)=d_{ij}^*/\sum_{j\in s_i^*}d_{ij}^*$.

(3) 计算相应的统计量

$$r_{pel}^*(\hat{\theta}_{pel})=\omega_1\sum_{j\in s_1^*}\tilde{d}_{1j}^*(s_1^*)\log\frac{p_{1j}}{\tilde{d}_{1j}^*(s_1^*)}+\omega_2\sum_{j\in s_2^*}\tilde{d}_{2j}^*(s_2^*)\log\frac{p_{2j}}{\tilde{d}_{2j}^*(s_2^*)}.$$

其中 p_{ij} ($i=1,2$)为最大化 $l_{pel}^*(p_1,p_2)$ 使其满足如下两个条件的解

$$\begin{cases}\sum_{j=1}^{n_1}p_{1j}=1,\ \sum_{j=1}^{n_2}p_{2j}=1\\ \sum_{i=1}^{2}\omega_i\sum_{j\in s_i^*}p_{ij}\mathbf{u}_{ij}=0.\end{cases}$$

其中 $\mathbf{u}_{ij}=\mathbf{z}_{ij}-\eta, i=1,2,$ 且 $\mathbf{z}_{1j}=(1,aY_{1j}^*/\omega_1)^T, \mathbf{z}_{2j}=(0,-bY_{2j}^*/\omega_2)^T, \eta=(\omega_1,\theta)^T$.

(4) 重复步骤 (1)-(3), 一共重复 B 次, 得到序列 $r_1^*(\hat{\theta}_{pel}),\cdots,r_B^*(\hat{\theta}_{pel})$, 记 b_α^* 为来自该序列的第 100α 样本分位数, 其中 $\alpha=0.05$.

主要考虑以下三种情况:

(1) 总体 Y_1 和 Y_2 均为正态分布;

(2) Y_1 和 Y_2 均服从对数正态分布;

(3) 两样本均选自两个有限总体, 响应变量包含许多零值.

通过覆盖率 (CP), 上 (U) 和下 (L) 尾误差率, 平均长度 (AL) 这几个指标来判断检验方法的优劣, 具体计算公式如下

$$
\begin{aligned}
L &= 100 \times \frac{1}{B}\sum_{b=1}^{B} I(\hat{\theta}_L^{(b)} \geq \theta), \\
U &= 100 \times \frac{1}{B}\sum_{b=1}^{B} I(\hat{\theta}_U^{(b)} \leq \theta), \\
CP &= 100 \times \frac{1}{B}\sum_{b=1}^{B} I(\hat{\theta}_L^{(b)} < \theta < \hat{\theta}_U^{(b)}), \\
AL &= \frac{1}{B}\sum_{b=1}^{B} (\hat{\theta}_U^{(b)} - \hat{\theta}_L^{(b)}),
\end{aligned}
$$

这里 $(\hat{\theta}_L^{(b)}, \hat{\theta}_U^{(b)})$ 是从第 b 个模拟样本中计算得的 θ 置信区间, $B = 2000$ 为模拟运行的总次数, 且 $L + U + CP = 100$.

在表 3-1 – 3-3 中, 通过设置样本量 n_1 和 n_2 的不同的值, 本文分别给出不同分布下的 CP, L, U 和 AL 值. 接下来, 对以上三种情况下的模拟结果进行具体分析如下:

对于情况 (1), 令 $Y_1 \sim N(\mu_1, \sigma_1^2)$, $Y_2 \sim N(\mu_2, \sigma_2^2)$, 其中 $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_1 = 1.5$, $\sigma_2 = 1$. 从表 3-1 可知, 除了在 $n_1 = 60$, $n_2 = 30$ 时 BWEL 方法不太敏感之外, 整体看来, WEL 法和 BWEL 法要一致优于标准的 EL 方法, 具体表现为: CP 值接近于名义值, L 和 U 值相对来说也比较平衡, AL 值没有大幅度波动.

对于情况 (2), 令 $Y_1 \sim Lognormal(\nu_1, \sigma_1^2)$, $Y_2 \sim Lognormal(\nu_2, \sigma_2^2)$, 其中 $\nu_1 = 1.1$, $\nu_2 = 1.2$, $\sigma_1^2 = 0.4$, $\sigma_2^2 = 0.2$. 分析表 3-2 可知, 样本量 $n_1 = n_2$ 时, 这三种方法的 CP 值相比样本量不等时更接近名义值; $n_1 = 30$, $n_2 = 60$ 时 BWEL 表现略差但也在 90 以上. 总体看来, EL、WEL 和 BWEL 效果均不错且依次递增.

对于情况 (3), 考虑样本量为 $N_1 = N_2 = 5000$ 的两有限总体. 对于第一个总体, 假定 3000 个响应变量 Y_{1j} 为 0, 剩余 2000 个响应变量服从均匀分布 $Uniform(0.8, 1.2)$; 对于第二个总体, 假定 4000 个响应变量 Y_{2j} 为 0, 剩余 1000 个响应变量从 $Uniform(1, 5)$中产生. 两样本均为不重复简单随机抽样. 从表 3-3 模拟结果可知: $n_1 = 30$, $n_2 = 30$ 和 $n_1 = 30$, $n_2 = 60$ 的情况下, 从指标 CP 来看, WEL 法要优于 BEL 方法; 其他情况下, 从 CP 值来看, EL、WEL 以及 BWEL 效果均不错, L 和 U 值依然呈现均匀发展态势.

表 3-1：两正态总体下 θ 的置信区间

(n_1, n_2)	CI	L	CP	U	AL
(30, 30)	EL	2.90	94.25	2.85	2.30
	WEL	2.50	94.90	2.60	2.52
	BWEL	2.45	95.25	2.30	2.99
(30, 60)	EL	2.80	93.60	3.60	1.77
	WEL	2.80	94.55	2.65	1.82
	BWEL	2.25	94.85	2.90	2.36
(60, 30)	EL	2.90	94.60	2.50	2.25
	WEL	2.45	94.65	2.90	2.28
	BWEL	3.75	92.80	3.45	2.88
(60, 60)	EL	2.65	94.45	2.90	1.80
	WEL	2.05	95.10	2.85	1.79
	BWEL	1.90	95.45	2.65	2.32
(30, 90)	EL	3.40	92.85	3.75	1.51
	WEL	2.45	95.60	1.95	1.66
	BWEL	2.15	95.35	2.50	2.33
(90, 30)	EL	2.75	95.10	2.15	2.20
	WEL	2.95	94.10	2.95	2.19
	BWEL	2.75	94.85	2.40	2.86

表 3-2 : 两对数正态总体下 θ 的置信区间

(n_1, n_2)	CI	L	CP	U	AL
$(30, 30)$	EL	1.95	95.20	2.85	1.54
	WEL	4.35	94.80	0.85	1.52
	BWEL	2.70	95.00	2.30	2.97
$(30, 60)$	EL	2.85	94.00	3.15	1.16
	WEL	3.35	94.90	1.75	1.16
	BWEL	2.80	93.70	3.50	2.34
$(60, 30)$	EL	1.80	94.80	3.40	1.49
	WEL	3.00	94.90	2.10	1.46
	BWEL	2.45	95.35	2.20	1.32
$(60, 60)$	EL	2.15	94.85	3.00	1.10
	WEL	2.90	95.65	1.45	1.09
	BWEL	1.90	95.70	2.40	2.25
$(30, 90)$	EL	2.75	94.50	2.75	1.00
	WEL	2.70	95.50	1.80	0.99
	BWEL	2.00	96.30	1.70	2.22
$(90, 30)$	EL	1.85	94.15	4.00	1.44
	WEL	3.05	94.60	2.35	1.60
	BWEL	2.70	95.20	2.10	2.36

表 3-3 : 两有限总体下 θ 的置信区间

(n_1, n_2)	CI	L	CP	U	AL
(30, 30)	EL	3.20	94.50	2.30	2.41
	WEL	1.80	96.30	1.90	2.88
	BWEL	1.95	95.85	2.20	2.97
(30, 60)	EL	3.30	94.25	2.45	1.71
	WEL	1.85	96.25	1.90	2.02
	BWEL	1.80	95.70	2.50	2.34
(60, 30)	EL	3.25	94.30	2.45	2.37
	WEL	2.40	95.25	2.35	2.86
	BWEL	2.55	95.35	2.10	1.32
(60, 60)	EL	2.95	94.60	2.45	1.70
	WEL	2.05	96.25	1.70	2.03
	BWEL	1.90	96.70	1.40	2.25
(30, 90)	EL	3.35	94.10	2.55	1.40
	WEL	2.00	96.40	1.60	1.65
	BWEL	2.00	96.30	1.70	2.22
(90, 30)	EL	3.10	93.90	3.00	2.38
	WEL	2.00	95.25	2.75	2.85
	BWEL	1.70	96.20	2.10	2.42

综上所述, 对于本章提出的检验问题, 所采用的三种经验似然方法都是行之有效的, 且从覆盖率和两个尾部误差率来看, BWEL 方法效果最显著, 其次为 WEL 法, 最后是标准的 EL 方法.

本章主要研究了固定维两样本均值线性组合的假设检验, 通过提出三种不同的非参数经验似然方法,构造相应的似然比检验统计量, 得出三种方法的检验统计量均服从于卡方分布. 然而, 对该问题的研究仅局限于有限维总体, 高维两样本问题的这几种经验似然方法并未进一步研究, 此外, 将本文相关理论结果运用到合适的实际统计问题中来解决实际问题也是本论文下一步的主要工作.

第四章　结论及展望

本文第二章通过对 Touloumis et al.[4] 中矩阵型数据进行向量拉直, 利用非参数经验似然检验方法, 在样本分量八阶矩存在且独立的前提下, 对其行列协差阵进行了检验, 得出经验似然比统计量渐近服从卡方分布. 数值模拟显示本章所使用的经验似然方法不仅在理论证明上比 U 统计量法更简便, 而且无论从经验水平还是经验功效分析, 检验效果都十分突出. 对于第三章的假设检验问题, 基于 Wu 和 Yan[5] 对固定维两样本均值相等性的推断, 在原有条件上将问题延伸至对两样本均值线性组合的推断, 证明了经验似然统计量均渐近服从卡方分布, 数值模拟显示其中的自助法检验效果较好.

尽管本文利用经验似然方法成功解决了矩阵型数据协差阵的检验问题和两样本均值的线性组合问题, 而且统计方法十分有效. 但针对此类问题, 仍存在许多问题尚未进行深入地研究, 下面主要列举三个今后需进一步去探索研究的问题.

1. 本文第二章中原假设是行列协差阵均为单位矩阵, 那么能否在行协差阵和列协差阵只有一个为单位矩阵时采用该方法进行统计检验, 去研究相应条件是否适用以及相关结论是否依然成立的问题呢?

2. 本文第三章是对固定维两样本均值的线性组合进行假设检验, 接下来考虑能否采用本文的几种经验似然方法将对两样本均值的线性组合的统计推断推广到高维, 并得到类似的结论.

3. 本文中针对这两类统计问题所进行的统计推断, 均证明了相应统计量渐近服从卡方分布这一结果. 然而这两个问题却都没有找出实际的数据和案例进行分析例证, 接下来主要任务就是, 寻找合适的与模型对应的数据并将之运用于实际, 从而更好地从实际问题方面去说明检验方法和结论的有效性.

参考文献

[1] Owen, A.B. (2001). Empirical Likelihood. Chapman and Hall/CRC.

[2] Zhang, R., Peng, L., Wang, R. (2013). Test for covariance matrix with fixed or divergent dimension. *Annals of Statistics*, **41**, 2075-2096.

[3] Peng, L., Qi, Y., Wang, F. (2014). Test for a mean vector with fixed or divergent dimension. *Statistica Science*, **29**, 113-127.

[4] Touloumis, A., Marioni, J., Tavare, S. (2014). Hypothesis testing for the covariance matrix in high-dimensional transposable data with kronecker product dependence structure. *Statistica Sinica*, **6**, 311-329.

[5] Wu, C., Yan, Y. (2012). Empirical likelihood inference for two-sample probles. *Statistics and its Interface*, **5**, 345-354.

[6] Chen, S.X., Zhang, L.X., Zhong, P.S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, **3105**, 810-819.

[7] Touloumis, A., Tavare, S., Marioni, J. (2015). Testing the mean matrix in high-dimensional transposable data. *Biometrics*, **71**, 157-166.

[8] Wang, R., Peng, L., Qi, Y. (2013). Jackknife empirical likelihood test for equality of two high dimensional means. *Statistica Sinica*, **23**, 667-690.

[9] Hjort, N.L., Mckeague, I.W., Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Annals of Statistics*, **37**, 1079-1111.

[10] Chen, B.B., Pan, G.M., Zhou, W. (2015). Large dimensional empirical likelihood, *Statistica Sinica*. **25**, 1659-1677.

[11] Wang, L., Yang, D. (2018). F-distribution calibrated empirical likelihood ratio tests for multiple hypothesis testing. *Journal of Nonparametric Statistics*, **30(3)**, 662-679.

[12] Jiang T.F., Yang F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Annals of Statistics*, **41(4)**, 2029-2074.

[13] Chen, S.X., Qin, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Annals of Statistics*, **37**, 808-835.

[14] Chen, S.X., Peng, L., Qin, Y. (2009). Empirical Likelihood methods for high dimension, *Biometrika*. **96**, 711-722.

[15] Wu, C.B., Rao, J.N.K. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *The Canadian Journal of Statistics*, **34**, 359-375.

[16] Chang, J., Chen, S.X., Chen, X. (2015). High dimensional generalized empirical likelihood for moment restrictions with dependent data. *Journal of Econometrics*, **185**, 283-304.

[17] Allen, G.I., Tibshirani, R. (2012). Inference with transposable data:Modelling the effects of row and column correlations. *Journal of the Royal Statiatical Society B*, **74**, 721-743.

[18] Wang, Q.H., Rao, J.N.K. (2002a). Empirical likelihood-based inference under imputation with missing response. *The Annals of Statistics*, **30**, 896-924.

[19] Wang, Q.H., Rao, J.N.K. (2002b). Empirical likelihood-based inference in linear models with missing data. *Scandinavian Journal of Statistics*, **29**, 563-576.

[20] Strivastava, M.S., Katayama, S., Kano, Y. (2013). A Two sample test in high dimensional data. *Journal of Multivariate Analysis*, **114**, 349-358.

[21] Dong, L.B. (2004). The Behrens-Fisher Problem: An Empirical Likelihood Approach. Technical Report, Department of Economics, University of Victoria.

[22] Liu, Y., Zou, C., Wang, Z. (2013). Calibration of the empirical likelihood for high-dimensional data. *Academic Journal*, **65**, 529-550.

[23] Strivastava, M.S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society*, **35**, 251-272.

[24] Wu, C.B., Rao, J.N.K. (2010). Bootstrap procedures for the pseudo empirical likelihood method in sample surveys. *Statistics and Probability Letters*, **80**, 1472-1478.

[25] Sang, Y.L., Xin, D., Zhao, Y.C. (2019). Jackknife empirical likelihood methods for Gini correlations and their equality testing. *Journal of Statistical Planning and Inference*, **199**, 45-59.

[26] Owen, A.B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics*, **18(1)**, 90-120.

[27] Fu, Y., Wang, X., Wu, C. (2008). Weighted Empirical Likelihood Inference for Multiple Samples. *Journal of Statistical Planning and Inference*, **139**, 1462-1473.

[28] Wu, C., Yan, Y. (2012). Weighted Empirical Likelihood Inference for Two-sample Problems. *Statistics and Its Interface*, **5**, 345-354.

[29] Dharmadhikari, S.W., Jogdeo, K. (1969). Bounds on moments of certain random variables. *The Annals of Mathematical Statistics*, **40**, 1506-1508.

[30] Wu, C.B. (2009). Empirical likelihood inference for two populations. Working Paper, University of Waterloo.

[31] Li, H.Q., Hu, J., Bai, Z.D., Yin, Y.Q., Zou, K.X. (2017). Test on the linear combinations of mean vectors in high-dimensional data. *Journal of Multivariate Analysis*, **26**, 188-208.

[32] Qin, Y.L. (2009). Statistical inference for high-dimensional data. Iowa State University.

[33] Srivastava, M.S., Kubokawa, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis*, **115**, 204-216.

[34] Bai, Z.D., Saranadasa, H. (1996). Effect of high-dimension: by an example of a two-sample problem. *Statistica Sinica*, **6**, 311-329.

[35] Ledoit, O., Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Annals of Statistics*, **30(4)**, 1081-1102.

[36] Wang, L., Peng, B., Li, R. (2015). A high-dimesnional nonparameteric multivariate test for mean vector. *Journal of the American Statistical Association*, **110**, 1658-1669.

[37] Qin, J., Zhang, B. (2007). Empirical-likelihood-based inference in missing response problems and its application in observational studies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69(1)**, 101-122.

[38] Peng, H., Schick, A. (2013). Empirical likelihood approach to goodness of fit testing. *Institute for Scientific Information*, **19(3)**, 954-981.

[39] Li, J., Chen, S. (2012). Two sample tests for high-dimensional covariance matrices. *Annals of Statistics*, **40**, 908-940.

[40] Yamada, T., Srivastava, M.S. (2012). A test for multivariate analysis of variance in high dimension. *Communications in Statistics-Theory and Methods*, **41**, 2602-2615.

[41] Zhang, C., Bai, Z., Hu, J., Wang, C. (2018). Multi-sample test for high-dimensional covariance matrices. *Communication in Statistics-Theory and Methods*, **47(13)**, 3161-3177.

[42] Wang, L., Veraverbeke, N. (2002). Empirical likelihood in a semi-parametric model for missing response data. *Communications in Statistics - Theory and Methods*, **35**, 625-639.

[43] Wu, C. (2004). Some algorithmic aspects of the empirical likelihood method in survey sampling. *Statistica Sinica*, **14**, 1057-1067.

[44] Lahiri, S.N., Mukhopadhyay, S. (2012). On the Mahalanobis-distance based penalized empirical likelihood method in high dimensions. *Statistica and Its Interface*, **5**, 331-338.

[45] Allen, G.I., Tibshirani, R. (2010). Transposable regularized covariance models with an application to missing data imputation. *The Annals of Applied Statistics*, **4**, 764-790.

[46] Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237-249.

[47] Jing, B.Y. (1995). Two-sample empirical likelihood method. *Statistics and Probability Letters*, **24**, 315-319.

[48] Chen, J., Sitter, R.R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*, **9**, 385-406.

致 谢

时光荏苒, 白驹过隙, 三年的研究生生活即将结束, 在该毕业论文完成之际, 我怀着感恩与激动之情, 真诚地向所有给与我关心和帮助的老师和同学们表示深深的感谢.

首先, 我想感谢我的导师解俊山老师. 感谢他在整个毕业论文设计的过程中, 研究生三年的学习中, 耐心细致地指导我, 督促我, 给我提了许多宝贵的意见和建议, 给了我许多引导与启发. 解老师不仅为人谦和, 待人真诚, 而且他那对工作认真负责的态度, 渊博的学识以及严谨的治学作风都使我受益匪浅. 虽然论文从选题到最终定稿遇到许多困难和障碍, 但在解老师的支持和点拨下, 最终也都拨开云雾. 再次向解老师致以最真诚的感谢和最美好的祝福.

其次, 我想感谢河南大学数学与统计学院领导们对我的培养与支持, 感谢你们为我提供了良好的学习环境和机会, 感谢院里外聘专家薛留根教授, 感谢薛老师在百忙之中多次对我论文进行指导与启发, 感谢院里同专业的各位老师对我的精心指导和耐心答疑, 这些都使我受益匪浅.

感谢一直以来朝夕相处共同成长共同进步的同学们, 感谢师兄师姐师弟师妹们对我学习和生活上的莫大鼓励与帮助. 我将心怀感恩, 砥砺前行, 不忘初心, 不负韶华.

此外, 我想感谢我的家人, 感谢你们在背后默默地支持我, 鼓励我, 在我迷茫时指引我正确的方向, 你们的支持与鼓励是我前进路上最大的动力, 我相信每一个不曾起舞的日子, 都是对生命的辜负, 生命不息, 奋斗不止, 未来可期.

最后, 诚挚地向在百忙之中评审本文的专家、老师们表示感谢!

马莹莹

2020 年 6 月