

分类号\_\_\_\_\_

密级\_\_\_\_\_

UDC\_\_\_\_\_

编号\_\_\_\_\_

華中師範大學

# 硕士学位论文

## 基于 bootstrap 方法的高维 数据两样本均值检验

学位申请人姓名: 曾奎杰

申请学位学生类别: 全日制硕士

申请学位学科专业: 概率论与数理统计

指导教师姓名: 左国新 教授

李正帮 副教授



硕士学位论文  
MASTER'S THESIS

---

# 硕士学位论文

基于bootstrap 方法的高维数据

两样本均值检验

论文作者：曾銮杰

指导教师：左国新 教授 李正帮 副教授

学科专业：概率论与数理统计

研究方向：统计学

华中师范大学数学与统计学学院

2020 年 4 月



硕士学位论文  
MASTER'S THESIS

---

**Two sample mean test of high dimensional data based on  
bootstrap method**

*A thesis*

*submitted in Partial Fulfillment of the Requirements*

*for the M.S. Degree in Statistics*

**By**

**Luanjie Zeng**

**Postgraduate Program**

**School of Mathematics and Statistics**

**Central China Normal University**

**Supervisor: Zhengbang Li**

**Academic Title: Associate Professor**

**Signature** *Guofin zuo*

**Approved**

**April 2020**



## 华中师范大学学位论文原创性声明和使用授权说明

### 原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

作者签名：常泰杰

日期：2020年5月31日

### 学位论文授权使用授权书

学位论文作者完全了解华中师范大学有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属华中师范大学。学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。（保密的学位论文在解密后遵守此规定）  
保密论文注释：本学位论文属于保密，在\_\_\_\_年解密后适用本授权书。非保密论文注释：本学位论文不属于保密范围，适用本授权书。

作者签名：常泰杰

导师签名：[Signature]

日期：2020年5月31日

日期：2020年6月1日

本人已经认真阅读“CALIS高校学位论文全文数据库发布章程”，同意将本人的学位论文提交“CALIS 高校学位论文全文数据库”中全文发布，并可按“章程”中的规定享受相关权益。同意论文提交后滞后：半年；一年；二年发布。

作者签名：常泰杰

导师签名：[Signature]

日期：2020年5月31日

日期：2020年6月1日



## 摘 要

随着社会和科技的快速发展, 很多领域出现了高维数据, 例如微阵列中基因表达值和单核苷酸多态性研究等. 由于高维数据的出现, 传统的多元统计方法和理论受到了极大的挑战. 在高维均值向量检验中, 由于传统的统计极限理论都是建立在维数  $p$  远小于样本量  $n$  的情况下, 所以当  $p$  大于  $n$  时, 传统的检验统计量失去了意义并且其极限理论也不再适用. 为此, 很多研究者已提出过相应的统计量解决此问题, 其研究思路基本都是通过理论推导, 从新统计量出发去寻找该统计量的均值估计以及协方差估计. 但在高维情况下, 找出其较为精确的协方差估计较为困难. 本文探究基于 stationary bootstrap 方法在高维数据下的均值向量检验问题. 通过理论推导, 我们知道 stationary bootstrap 适用于学生化列统计量平方和的统计量. 通过 R 语言模拟实验, 并跟一部分以往的方法进行比较, 我们可发现: 基于 stationary bootstrap 方法的学生化列统计量平方和的统计量, 其性能相对很好, 因为新统计量的经验功效以及经验水平控制效果相对较好, 并且该方法避免了在理论上对统计量协方差估计的繁杂计算.

**关键词:** 高维数据; 两样本均值检验; stationary bootstrap; 协方差估计



## Abstract

With the rapid development of society and science and technology, high dimensional data have emerged in many fields, such as gene expression value and single nucleotide polymorphism in microarray. Due to the emergence of high dimensional data, the traditional multivariate statistical methods and theories have been greatly challenged. In the high-dimensional mean value vector test, when  $p$  is greater than  $n$ , the traditional test statistics lose significance and its limit theory is no longer applies, because the traditional statistical limit theory is based on the case that the dimension  $P$  is far less than the sample size  $n$ . For this reason, many researchers have proposed corresponding statistics to solve this problem. Their research ideas are basically based on theoretical derivation, starting from the new statistics to find its mean value estimation and covariance estimation. However, in the case of high dimension, it is difficult to find out the accurate covariance estimation. In this paper, we study the problem of mean vector test based on stationary bootstrap method in high dimensional data. Through theoretical derivation, we know that the stationary bootstrap is suitable for square summation of studentized column statistics. Through a simulation experiment and comparison with some previous methods, we can find that: The statistics of the sum of squares of the student column statistics based on the stationary bootstrap method has relatively good performance, because of the new statistics having relatively good experience power and experience size control, and the method avoids the complicated calculation of the covariance estimation of the statistics in theory.

**Keywords:** High-dimensional data; Two sample mean test; stationary bootstrap; covariance estimation



## 目录

摘要	I
Abstract	II
第一章 绪论	1
1.1 研究背景	1
1.2 研究现状	2
1.3 本文主要工作及结构安排	3
第二章 基于stationary bootstrap 方法的检验	4
2.1 学生化和非学生化列统计量平方和的检验方法	4
2.1.1 非学生化列统计量的平方和	4
2.1.2 学生化列统计量的平方和	5
2.2 stationary bootstrap方法和最优平均长度	6
2.2.1 stationary bootstrap 抽样方法	7
2.2.2 最优平均长度	7
2.3 stationary bootstrap方法的应用	9
2.3.1 理论支持	9
2.3.2 stationary bootstrap算法的应用	11
第三章 统计模拟和实例研究	12
3.1 模拟及分析	12
3.1.1 模拟	12
3.1.2 模拟结果及分析	13
3.2 实例及分析	30
第四章 总结与展望	32
参考文献	34



附录 .....	37
致谢 .....	43





## 第一章 绪论

### 1.1 研究背景

假设检验是由 K.Pearson 于 20 世纪初提出, 由 Fisher, Neyman 和 E.Pearson 等完善的一套理论, 是统计推断的其中一个主要内容, 其中显著性检验是最基本的一种方法. 它的原理是对总体特征作出某种假设, 通过利用从该总体中随机抽出的样本信息, 然后对此假设应该被接受与否作出判断. 随着科学技术的不断发展, 在很多领域上高维数据的出现越来越频繁, 例如医学, 计算生物学, 金融分析, 风险控制等等. 高维数据又称为大  $p$  小  $n$ , 即数据的维数  $p$  大于数据个数  $n$ . 例如 DNA, 蛋白质的测试数据, 经济数据, 金融数据, 社会调查数据等等都属于高维数据. 传统的一维经典检验方法和多维经典检验方法比如  $t$  检验,  $u$  检验, Hotelling  $T^2$  [1] 检验等在高维数据情形下根本没意义或者不适用, 这是维数灾难带来的巨大影响. 下面看一个例子:

例 设  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  为来自  $p$  维正态分布  $N_p(\boldsymbol{\mu}, \mathbf{I})$  的一个容量为  $n$  的样本. 令样本均值为  $\bar{\mathbf{Y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}_i$ . 记  $\bar{y}_i$  和  $\mu_i$  分别为  $\bar{\mathbf{Y}}$  和  $\boldsymbol{\mu}$  的第  $i$  分量, 则

$$\bar{y}_i - \mu_i \sim N(0, \frac{1}{n}).$$

因此

$$\bar{y}_i - \mu_i = O_P(\frac{1}{\sqrt{n}}),$$

$$\|\bar{\mathbf{Y}} - \boldsymbol{\mu}\|^2 = \sum_{i=1}^p (\bar{y}_i - \mu_i)^2 = \sum_{i=1}^p O_P(\frac{1}{n}) = p \cdot O_P(\frac{1}{n}).$$

显然, 当样本容量  $n$  保持不变, 维数  $p$  增大时, 样本均值  $\bar{\mathbf{Y}}$  作为总体均值  $\boldsymbol{\mu}$  的估计将变得很不好. 所以为了使  $\|\bar{\mathbf{Y}} - \boldsymbol{\mu}\|^2 \xrightarrow{P} 0$  能够在  $p$  增加的情况下仍成立,  $n$  必须与  $p$  保持同样的速度或更快的速度增加.

从上面的例子可以看出高维数据给统计推断带来的影响: 在经典背景下建立的极限理论不能简单地应用在高维数据上. 更多更详细的资料, 可以参考文献 Donoho (2000)[2] 和 Fan(2006)[3].



## 1.2 研究现状

在多元统计分析中,关于总体均值和总体协方差矩阵的统计推断是非常重要的问题.下面介绍两样本均值问题:

假设  $\mathbf{X}^{(1)} = (X_{11}, \dots, X_{1p})^T, \dots, \mathbf{X}^{(m)} = (X_{m1}, \dots, X_{mp})^T$  是来自均值为  $\boldsymbol{\mu}_1 = (\mu_{11}, \dots, \mu_{1p})^T$ , 协方差阵为  $\boldsymbol{\Sigma}_1 = (\sigma_{1:q,s})_{q,s=1,\dots,p}$  的总体的容量为  $m$  的独立样本,  $\mathbf{Y}^{(1)} = (Y_{11}, \dots, Y_{1p})^T, \dots, \mathbf{Y}^{(n)} = (Y_{n1}, \dots, Y_{np})^T$  是来自均值为  $\boldsymbol{\mu}_2 = (\mu_{21}, \dots, \mu_{2p})^T$ , 协方差阵为  $\boldsymbol{\Sigma}_2 = (\sigma_{2:q,s})_{q,s=1,\dots,p}$  的总体的容量为  $n$  的独立样本. 其中  $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$  是未知的非奇异矩阵,  $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$  也是未知的. 我们感兴趣的检验问题是:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ VS } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

记  $\bar{\mathbf{X}} = \frac{1}{m} \sum_{i=1}^m \mathbf{X}^{(i)}, \mathbf{S}_1 = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{X}^{(i)} - \bar{\mathbf{X}})(\mathbf{X}^{(i)} - \bar{\mathbf{X}})^T, \bar{\mathbf{Y}} = (\hat{\sigma}_{1:q,s})_{q,s=1,\dots,p} = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}^{(i)}, \mathbf{S}_2 = (\hat{\sigma}_{2:q,s})_{q,s=1,\dots,p} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{Y}^{(i)} - \bar{\mathbf{Y}})(\mathbf{Y}^{(i)} - \bar{\mathbf{Y}})^T.$

当  $p$  小于  $m, n$  且  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$  时, *Hotelling* 检验为:

$$T = \frac{mn}{m+n} (\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T \mathbf{S}^{-1} (\bar{\mathbf{X}} - \bar{\mathbf{Y}}),$$

其中  $\mathbf{S} = \frac{1}{m+n-2} ((m-1)\mathbf{S}_1 + (n-1)\mathbf{S}_2).$

当  $p$  小于  $m, n$  时, *Hotelling* 检验拥有很多优良性质,例如:在原假设下检验统计量的精确分布能够得到;在线性变换下检验统计量是不变检验;且是一致最有效不变检验;检验统计量是可容许检验.

但是,当  $p$  大于  $m, n$  时, *Hotelling* 检验有一个严重的缺陷:当维数大于样本容量时, *Hotelling* 检验没有定义. 当:  $p > m + n - 2$  时,由于样本协方差矩阵  $\mathbf{S}$  以概率 1 不可逆,  $T$  没有定义. Dempster 最先发现了这个问题,并且于1958年和1960年发了两篇文献 Dempster(1958)<sup>[6]</sup>, Dempster(1960)<sup>[7]</sup> 来解决这一问题,他提出了一种非精确检验(*non-exacttest*). Dempster 的非精确检验严格依赖于正态总体的假设,且涉及极为复杂的自由度估计. 后续又有很多学者进一步研究了该问题,提出了很多解决此问题的统计量和方法,这些检验统计量大致可以分为两类:一种类型是关于两均值向量之差  $\bar{\mathbf{X}} - \bar{\mathbf{Y}}$  的平方形式,例如当协方差各维方差一致时, Bai & Saranadasa(1996)<sup>[6]</sup>, Chen & Qin(2010)<sup>[7]</sup> 提出用非学生化  $\bar{X}_1 - \bar{Y}_1, \dots, \bar{X}_p - \bar{Y}_p$  各项平方之和的统计量;当协方差各维方差不一致时, Srivastava et al.(2013)<sup>[8]</sup>, Gregory and Carroll (2015)<sup>[9]</sup> 主张用学生化  $\frac{\bar{X}_1 - \bar{Y}_1}{\sqrt{\frac{\hat{\sigma}_{1:1,1}}{m} + \frac{\hat{\sigma}_{2:1,1}}{n}}}$ ,



...,  $\frac{\bar{X}_p - \bar{Y}_p}{\sqrt{\frac{\hat{\sigma}_{1:p,p}^2}{m} + \frac{\hat{\sigma}_{2:p,p}^2}{n}}}$  各项平方之和的统计量. 为了做统计推断, 非学生化统计量 (Bai & Saranadasa(1996)<sup>[6]</sup>, Chen & Qin(2010)<sup>[7]</sup>) 需要估计  $tr\Sigma_1$ ,  $tr\Sigma_2$ ,  $tr\Sigma_1^2$ ,  $tr\Sigma_2^2$  和  $tr\Sigma_1\Sigma_2$ , 学生化统计量 (Srivastava et al.(2013)<sup>[8]</sup>) 需要估计  $trR_1$ ,  $trR_2$ ,  $trR_1^2$ ,  $trR_2^2$  和  $trR_1R_2$ , 其中  $R_1$  和  $R_2$  分别是  $\Sigma_1$  和  $\Sigma_2$  相应的相关矩阵. Gregory and Carroll (2015)<sup>[9]</sup> 采用截断方法估计统计量的协方差矩阵. 这种类型检验统计量在检验  $\mu_1 - \mu_2$  非零成分相对密集的差异是很明显的. 而另一种类型的统计量是关于两均值向量之差  $\bar{X} - \bar{Y}$  的极大值形式, Cai et al.(2013)<sup>[10]</sup>, Cai et al.(2014)<sup>[11]</sup>, Chang et al.(2017)<sup>[12]</sup>, 该类型检验统计量能明显检验  $\mu_1 - \mu_2$  非零成分相对稀疏的差异.

### 1.3 本文主要工作及结构安排

因为 stationary bootstrap 方法能够避免繁琐的协方差估计的推演计算过程, 这里将尝试采用 Politis & Romao (1994)<sup>[13]</sup> 提出的 stationary bootstrap 方法来计算学生化列统计量平方和的统计量的  $p$  值. 本文主要集中精力将新方法与上一节提到的第一种类型的检验统计量做比较. 本文的架构安排如下:

第一章前两节介绍了高维数据给统计推断带来的巨大挑战, 以及介绍了与本文研究的问题的研究现状; 第二章详细介绍本文将用来比较经验水平和经验功效的统计量, 并在此基础上提出可用 stationary bootstrap 方法的统计量; 第三章对第二章提到的统计量进行模拟比较和实例分析; 第四章对此过程作一个总结并对未来工作做了一些展望.



## 第二章 基于stationary bootstrap 方法的检验

在这一部分, 回顾上面提及的四种主要统计量检验, 介绍 stationary bootstrap 方法和提出基于 stationary bootstrap 方法的两样本高维数据均值向量检验的统计量.

### 2.1 学生化和非学生化列统计量平方和的检验方法

#### 2.1.1 非学生化列统计量的平方和

Bai & Saranadasa (1996)<sup>[6]</sup>修正了 Hotelling  $T^2$ <sup>[1]</sup> 检验统计量, 利用非学生化列统计量  $\bar{X}_1 - \bar{Y}_1, \dots, \bar{X}_p - \bar{Y}_p$  提出了非学生化列统计量平方和的统计量, 记为  $M_{m,n,p} = \sum_{j=1}^p (\bar{X}_j - \bar{Y}_j)^2 - (\frac{1}{n} + \frac{1}{m}) \text{tr } S$ ,  $S = \frac{(m-1)S_1 + (n-1)S_2}{m+n-2}$ . 当  $\Sigma_1 = \Sigma_2$  和  $m, n, p \rightarrow \infty$  时, 在零假设下有

$$BS = \frac{M_{m,n,p}}{\sqrt{\widehat{\text{Var}} M_{m,n,p}}} \rightarrow N(0, 1)$$

其中  $\widehat{\text{Var}} M_{m,n,p} = \frac{2(m+n)(m+n-1)(m+n-2)}{m^2 n^2 (m+n-3)} (\text{tr } S^2 - \frac{1}{m+n-2} (\text{tr } S)^2)$ .

显然关于  $\widehat{\text{Var}}(M_{m,n,p})$  的估计是十分重要的部分.

Chen & Qin(2010)<sup>[7]</sup>分析了 BS 之后, 提出了另一种非学生化平方和形式并记为:  $T_{m,n,p} = \frac{\sum_{(i \neq j)=1}^m \mathbf{X}_i^T \mathbf{X}_j}{m(m-1)} + \frac{\sum_{(i \neq j)=1}^n \mathbf{Y}_i^T \mathbf{Y}_j}{n(n-1)} - 2 \frac{\sum_{i=1}^m \sum_{j=1}^n \mathbf{X}_i^T \mathbf{Y}_j}{nm}$ . 当  $\Sigma_1 \neq \Sigma_2$  和  $m, n, p \rightarrow \infty$  时, 在零假设下有

$$CQ = \frac{T_{m,n,p} - \|\mu_1 - \mu_2\|^2}{\sqrt{\widehat{\text{Var}}(T_{m,n,p})}} \xrightarrow{d} N(0, 1),$$

其中  $\widehat{\text{Var}}(T_{m,n,p}) = \frac{2}{m(m-1)} \widehat{\text{tr}}(\Sigma_1^2) + \frac{2}{n(n-1)} \widehat{\text{tr}}(\Sigma_2^2) + \frac{4}{mn} \widehat{\text{tr}}(\Sigma_1 \Sigma_2)$ ;

$$\widehat{\text{tr}}(\Sigma_1^2) = \{m(m-1)\}^{-1} \widehat{\text{tr}} \left\{ \sum_{j \neq k}^m (\mathbf{X}_j - \bar{\mathbf{X}}_{(j,k)}) \mathbf{X}_j^T (\mathbf{X}_k - \bar{\mathbf{X}}_{(j,k)}) \mathbf{X}_k^T \right\};$$

$$\widehat{\text{tr}}(\Sigma_1 \Sigma_2) = (mn)^{-1} \widehat{\text{tr}} \left\{ \sum_{l=1}^m \sum_{k=1}^n (\mathbf{X}_l - \bar{\mathbf{X}}_{(l)}) \mathbf{X}_l^T (\mathbf{Y}_k - \bar{\mathbf{Y}}_{(k)}) \mathbf{Y}_k^T \right\};$$

$$\widehat{\text{tr}}(\Sigma_2^2) = \{n(n-1)\}^{-1} \widehat{\text{tr}} \left\{ \sum_{j \neq k}^n (\mathbf{Y}_j - \bar{\mathbf{Y}}_{(j,k)}) \mathbf{Y}_j^T (\mathbf{Y}_k - \bar{\mathbf{Y}}_{(j,k)}) \mathbf{Y}_k^T \right\};$$



$\bar{\mathbf{X}}_{(j,k)}$  是从  $m$  个样本中删除两个不同 (即  $j \neq k$ ) 样本  $\mathbf{X}^{(j)}$ ,  $\mathbf{X}^{(k)}$  之后的均值;  $\bar{\mathbf{X}}_{(l)}$  是从  $m$  个样本中删除样本  $\mathbf{X}^{(l)}$  之后的均值;  $\bar{\mathbf{Y}}_{(j,k)}$  是从  $n$  个样本中删除两个不同 (即  $j \neq k$ ) 样本  $\mathbf{Y}^{(j)}$ ,  $\mathbf{Y}^{(k)}$  之后的均值;  $\bar{\mathbf{Y}}_{(l)}$  是从  $n$  个样本中删除样本  $\mathbf{Y}^{(l)}$  之后的均值.

注意到, 在零假设下:

$$T_{m,n,p} - \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 = \sum_{j=1}^p (\bar{X}_j - \bar{Y}_j)^2 - \left( \frac{1}{m} \text{tr} \mathbf{S}_1 + \frac{1}{n} \text{tr} \mathbf{S}_2 \right).$$

因为在零假设下,

$$\begin{aligned} & T_{m,n,p} - \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 \\ &= T_{m,n,p} \\ &= \frac{\sum_{(i \neq j)=1}^m \mathbf{X}_i^T \mathbf{X}_j}{m(m-1)} + \frac{\sum_{(i \neq j)=1}^n \mathbf{Y}_i^T \mathbf{Y}_j}{n(n-1)} - 2 \frac{\sum_{i=1}^m \sum_{j=1}^n \mathbf{X}_i^T \mathbf{Y}_j}{mn} \\ &= \frac{\sum_{i,j=1}^m \mathbf{X}_i^T \mathbf{X}_j - \sum_{i=1}^m \mathbf{X}_i^T \mathbf{X}_i}{m(m-1)} + \frac{\sum_{i,j=1}^n \mathbf{Y}_i^T \mathbf{Y}_j - \sum_{i=1}^n \mathbf{Y}_i^T \mathbf{Y}_i}{n(n-1)} - 2 \frac{\sum_{i=1}^m \sum_{j=1}^n \mathbf{X}_i^T \mathbf{Y}_j}{nm} \\ &= \frac{m^2 \bar{\mathbf{X}}^T \bar{\mathbf{X}} - (m-1) \text{tr} \mathbf{S}_1 - m \bar{\mathbf{X}}^T \bar{\mathbf{X}}}{m(m-1)} + \frac{m^2 \bar{\mathbf{Y}}^T \bar{\mathbf{Y}} - (m-1) \text{tr} \mathbf{S}_2 - m \bar{\mathbf{Y}}^T \bar{\mathbf{Y}}}{n(n-1)} - 2 \frac{mn \bar{\mathbf{X}}^T \bar{\mathbf{Y}}}{nm} \\ &= (\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T (\bar{\mathbf{X}} - \bar{\mathbf{Y}}) - \left( \frac{1}{m} \text{tr} \mathbf{S}_1 + \frac{1}{n} \text{tr} \mathbf{S}_2 \right) \\ &= \sum_{j=1}^p (\bar{X}_j - \bar{Y}_j)^2 - \left( \frac{1}{m} \text{tr} \mathbf{S}_1 + \frac{1}{n} \text{tr} \mathbf{S}_2 \right). \end{aligned}$$

但  $\widehat{\text{Var}}(T_{m,n,p}) \neq \widehat{\text{Var}}(M_{m,n,p})$ , 因此 BS 和 CQ 的功效性能应该是不一样的.

### 2.1.2 学生化列统计量的平方和

Srivastava et al.(2013)<sup>[8]</sup> 基于学生化列统计量  $\frac{\bar{X}_1 - \bar{Y}_1}{\sqrt{\frac{\hat{\sigma}_{1:1,1}}{m} + \frac{\hat{\sigma}_{2:1,1}}{n}}}, \dots, \frac{\bar{X}_p - \bar{Y}_p}{\sqrt{\frac{\hat{\sigma}_{1:p,p}}{m} + \frac{\hat{\sigma}_{2:p,p}}{n}}}$ , 提出了学生化平方和形式的统计量, 记为  $A_{m,n,p} = \left( \frac{\bar{X}_1 - \bar{Y}_1}{\sqrt{\frac{\hat{\sigma}_{1:1,1}}{m} + \frac{\hat{\sigma}_{2:1,1}}{n}}} \right)^2 + \dots +$



$\left(\frac{\bar{X}_p - \bar{Y}_p}{\sqrt{\frac{\hat{\sigma}_{1:p,p}^2}{m} + \frac{\hat{\sigma}_{2:p,p}^2}{n}}}\right)^2 - p$ . 当  $m, n, p \rightarrow \infty$  时, 在零假设下有:

$$SR = \frac{A_{m,n,p}}{\sqrt{\widehat{\text{Var}}(A_{m,n,p})}} \xrightarrow{d} N(0, 1),$$

其中  $\widehat{\text{Var}}(A_{m,n,p}) = \frac{2 \text{tr} \mathbf{R}^2}{p} - \frac{2}{p(m-1)m^2} (\text{tr} \hat{\mathbf{D}}^{-1} \mathbf{S}_1)^2 - \frac{2}{p(n-1)n^2} (\text{tr} \hat{\mathbf{D}}^{-1} \mathbf{S}_2)^2$ ,  $\mathbf{R} = \hat{\mathbf{D}}^{-1/2} \left(\frac{\hat{\mathbf{S}}_1}{m} + \frac{\hat{\mathbf{S}}_2}{n}\right) \hat{\mathbf{D}}^{-1/2}$ ,  $\hat{\mathbf{D}} = \frac{\hat{\mathbf{D}}_1}{m} + \frac{\hat{\mathbf{D}}_2}{n}$  和  $\hat{\mathbf{D}}_i = \text{diag}(\hat{\sigma}_{i:1,1}, \dots, \hat{\sigma}_{i:p,p})$  ( $i = 1, 2$ ).

显然, 在 SR 中, 关于  $\widehat{\text{Var}} A_{m,n,p}$  的估计是十分重要的, 当然这个估计值并不好估计出来.

Gregory et al.(2015)<sup>[9]</sup>考察了 SR 之后, 提出了另一种学生化列统计量的平方之和的统计量, 记为  $B_{m,n,p} = \left(\frac{\bar{X}_1 - \bar{Y}_1}{\sqrt{\frac{\hat{\sigma}_{1:1,1}^2}{m} + \frac{\hat{\sigma}_{2:1,1}^2}{n}}}\right)^2 + \dots + \left(\frac{\bar{X}_p - \bar{Y}_p}{\sqrt{\frac{\hat{\sigma}_{1:p,p}^2}{m} + \frac{\hat{\sigma}_{2:p,p}^2}{n}}}\right)^2 - p \hat{\xi}_{m,n,p}$ , 其中, 当  $p$  ( $p = O((m+n)^2)$ ) 为稳健中阶版本时,  $\hat{\xi}_{m,n,p} = 1$ ; 当  $p$  ( $p = O((m+n)^6)$ ) 为高阶版本时,  $\hat{\xi}_{m,n,p}$  等于其它的. 当  $m, n, p \rightarrow \infty$  时, 在零假设下有:

$$GCT = \frac{B_{m,n,p}}{\sqrt{\widehat{\text{Var}}(B_{m,n,p})}} \xrightarrow{d} N(0, 1),$$

其中  $\widehat{\text{Var}}(B_{m,n,p}) = \sqrt{p} \hat{\zeta}_{m,n}^2$ ,

$\hat{\zeta}_{m,n}^2 \equiv \sum_{|k| < L} w(k/L) \hat{\gamma}(k)$ ,  $\hat{\gamma}(k) = (p-k)^{-1} \sum_{j=1}^{p-k} (t_{s,j}^2 - Q_{m,n,p}) (t_{s,(j+k)}^2 - Q_{m,n,p})$ ,  $t_{s,1} = \frac{(\bar{X}_1 - \bar{Y}_1)^2}{\frac{\hat{\sigma}_{1:1,1}^2}{m} + \frac{\hat{\sigma}_{2:1,1}^2}{n}}, \dots, t_{s,p} = \frac{(\bar{X}_p - \bar{Y}_p)^2}{\frac{\hat{\sigma}_{1:p,p}^2}{m} + \frac{\hat{\sigma}_{2:p,p}^2}{n}}$ ,  $Q_{m,n,p} = \frac{1}{p} \sum_{q=1}^p \left(\frac{\bar{X}_q - \bar{Y}_q}{\sqrt{\frac{\hat{\sigma}_{1:q,q}^2}{m} + \frac{\hat{\sigma}_{2:q,q}^2}{n}}}\right)^2$ ,  $w(x)$  是  $x$  的阶梯偶函数, 满足:  $w(0) = 1$ ; 对于所有的  $x$ ,  $|w(x)| \leq 1$ ; 还有当  $|x| > 1$  时,  $w(x) = 0$ ,  $L$  是用户选择滞后窗口大小.

很显然, 关于  $\widehat{\text{Var}}(B_{m,n,p})$  的一个好的估计值是困难的. Gregory et al.(2015)<sup>[9]</sup>采用 Brockwell(1991)<sup>[14]</sup> 和 Politis & Romano (1995)<sup>[15]</sup>的想法来估计  $\widehat{\text{Var}}(B_{m,n,p})$ .

我们可以发现  $\widehat{\text{Var}}(A_{m,n,p})$  和  $\widehat{\text{Var}}(B_{m,n,p})$  的估计值是不同一样的, 因此 SR 和 GCT 的功效性能也应该是有所差别的.

## 2.2 stationary bootstrap方法和最优平均长度

在这一节将介绍 stationary bootstrap 方法的基本知识, 包括抽样方法以及最优平均长度.



### 2.2.1 stationary bootstrap 抽样方法

根据Politis & Romao (1994)<sup>[13]</sup>可知, stationary bootstrap 方法的抽样方法如下: 假设我们有  $p$  个观测值  $\chi_p = \{t_1, \dots, t_p\}$ , 这  $p$  个观测值是严格稳健弱相关的时间序列, 则  $p$  个 stationary bootstrap 抽样值是来自若干个 block 的. 不妨记每个 block 为  $B_{I_i, k_i} = \{t_{I_i}, \dots, t_{I_i+k_i-1}\}$ , 其中  $i = 1, 2, \dots$ , 独立同分布样本  $I_1, I_2, \dots$  是来自离散均匀分布  $\{1, 2, \dots, p\}$ , 独立同分布样本  $k_1, k_2, \dots$  是来自概率为  $q(1-q)^{m-1}$  ( $m = 1, 2, \dots$ ) 的几何分布. 令  $t_1^* = t_{I_1}$  是从  $\chi_p$  里面随机抽取的, 然后考虑"成功"概率为  $q$ , "失败"概率为  $1-q$  的二项分布, 接下来第二个 stationary bootstrap 样本  $t_2^* = t_{I_1+1}$ , 即  $t_{I_1+1}$  是  $t_{I_1}$  在  $\chi_p$  的原始位置中的下一个, 如果"成功"的话, 将令  $t_2^*$  继续添加至第一个 block 里面, 此时第一个 block 有两个元素; 但如果"失败"的话, 则令  $t_2^* = t_{I_2}$ , 即  $t_{I_2}$  是重新从  $\chi_p$  中随机抽取, 将重新启动第二个 block, 此时第一个 block 就只有一个元素. 第三个以及往后的 stationary bootstrap 样本将继续按照此原则进行下去, 直至取到  $k-1+k_2+\dots \geq p$  为止. 若  $k-1+k_2+\dots > p$ , 我们只需取前  $p$  个就可以. 如果  $I_i > p$ , 则  $t_{I_i} = t_{I_i-p}$ .

### 2.2.2 最优平均长度

假设  $\{X_n, n \in \mathbf{Z}\}$  是具有严平稳弱相关的时间序列样本观察值, 令  $\theta = \mu$ ,  $\hat{\theta}_n = \bar{X}_n$ , 其中  $\mu$  是  $\{X_n, n \in \mathbf{Z}\}$  联合分布的均值,  $\hat{\theta}_n$  是感兴趣参数  $\theta$  的估计量,  $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ . 考虑形如:  $W_n = n^{1/2}[\hat{\theta}_n - \theta]$  的统计量. 假设关于 stationary bootstrap 方法抽出来的样本为  $X_1^*, \dots, X_N^*$ , 记  $\hat{\theta}_n^* = \bar{X}_n^*$  ( $\bar{X}_n^* = N^{-1} \sum_{i=1}^N X_i^*$ ),  $\sigma^{2*} = \text{Var}(W_n^*)$ .

根据Politis D. & Romano(1994)<sup>[13]</sup>:  $E(W_n^*) = E(n^{1/2}[\hat{\theta}_n^* - \hat{\theta}_n]) = 0$ ;  
根据 White & Politis (2004)<sup>[16]</sup>: 假设对于  $\delta > 0$ ,  $E|X_t|^{6+\delta} < \infty$  和  $\sum_{k=1}^{\infty} k^2 (\alpha_X(k))^{\frac{\delta}{6+\delta}} < \infty$ . 如果随着  $N \rightarrow \infty$  有  $b \rightarrow \infty$  且  $b = o(N^{1/2})$ , 则有

$$\text{Bias}(\sigma^{2*}) = -\frac{1}{b}G + o(1/b)$$

$$\text{Var}(\sigma^{2*}) = \frac{b}{N}D_{SB} + o(b/N)$$

其中  $R(s) = E(X_t - \theta)(X_{t+|s|} - \theta)$ ,  $g(w) := \sum_{s=-\infty}^{\infty} R(s) \cos(ws)$ ,  $G = \sum_{k=-\infty}^{\infty} |k| R(k)$ ,  $D_{SB} = \left( 4g^2(0) + \frac{2}{\pi} \int_{-\pi}^{\pi} (1 + \cos w) g^2(w) dw \right)$ .



因此我们有

$$\text{MSE}(\sigma_2^*) = \frac{G^2}{b^2} + D_{SB} \frac{b}{N} + o(b^{-2}) + o(b/N)$$

解得

$$b_{opt,SB} = \left( \frac{2G^2}{D_{SB}} \right)^{1/3} N^{1/3}.$$

要得出  $b_{opt,SB}$  的估计值  $\hat{b}_{opt,SB}$ , 需要正确估计参数  $\sum_{k=-\infty}^{\infty} |k|R(k)$ ,  $g(0) = \sum_{k=-\infty}^{\infty} R(0)$  以及  $\frac{1}{\pi} \int_{-\pi}^{\pi} (1 + \cos w)g^2(w)dw$ . 在这里采用 Politis & Romano(1995)<sup>[15]</sup> 的平顶式滞后窗口, 因此可以用  $\sum_{k=-M}^M \lambda(k/M)|k|\hat{R}(k)$  估计  $\sum_{k=-\infty}^{\infty} |k|R(k)$ , 其中  $\hat{R}(k) = \frac{1}{N} \sum_{i=1}^{N-|k|} (X_i - \bar{X}_N)(X_{i+|k|} - \bar{X}_N)$ ,

$$\lambda(t) = \begin{cases} 1, & |t| \in [0, 1/2] \\ 2(1 - |t|), & |t| \in [1/2, 1] \\ 0, & |t| \in (1, \infty). \end{cases}$$

同样地, 用  $\hat{g}(w) = \sum_{k=-M}^M \lambda(k/M)\hat{R}(k)\cos(wk)$  估计  $g(w) = \sum_{k=-\infty}^{\infty} R(w)\cos(wk)$ . 将  $\hat{R}(k)$  和  $\hat{g}(w)$  代入  $G$  和  $D_{SB}$  可以得到:

$$\hat{G} = \sum_{k=-M}^M \lambda(k/M)\hat{R}(k)$$

和

$$\hat{D}_{SB} = 4\hat{g}^2(0) + \frac{2}{\pi} \int_{-\pi}^{\pi} (1 + \cos w)\hat{g}^2(w)dw$$

因此有:

$$\hat{b}_{opt,SB} = \left( \frac{2\hat{G}^2}{\hat{D}_{SB}} \right)^{1/3} N^{1/3}.$$

使用平顶式滞后窗口的原因除了快速收敛和对底层相关结构的自适应之外, 另一重要原因是其选择带宽  $M$  通过对相关图的简单检查, 在实践中是直观且可行的. 特别地, Politis & Romano(1995)<sup>[15]</sup> 建议寻找最小整数  $\hat{m}$ , 之后的相关图可以忽略不计, 即  $\hat{R}(k) \simeq 0$   $k > \hat{m}$ . 当然,  $\hat{R}(k) \simeq 0$  并不意味着  $\hat{R}(k)$  显著不同于 0, 即隐含假设检验. 引用 Politis(2001)<sup>[17]</sup> 提出的隐含假设检验的一个确切式子, 令  $\rho(k) = R(k)/R(0)$ ,  $\hat{\rho}(k) = \hat{R}(k)/\hat{R}(0)$  以及  $\hat{m}$  为最小正整数且满足  $|\hat{\rho}(\hat{m} + k)| = c\sqrt{\log N/N}$  ( $k = 1, \dots, K_N$ ), 其中  $c > 0$  是一个固定常数,  $K_N$  是  $N$  的





非降整数值函数, 满足  $K_N = o(\log N)$ . 在本文中, 采用  $c = 2$ ,  $K_N = \max(5, \sqrt{\ln N})$ . 识别相关图上的  $\hat{m}$  后, White & Politis (2004)<sup>[16]</sup> 建议取  $M = 2\hat{m}$ .

## 2.3 stationary bootstrap方法的应用

### 2.3.1 理论支持

当维数  $p$  小于  $m, n$  时,  $(m+n)^{1/2}(\bar{X} - \bar{Y})$  (Bilodeau et. al (1999)<sup>[18]</sup>) 渐近多元正态分布. 但当维数  $p$  大于  $m, n$  时, 下面将在特定的条件下, 推导  $(m+n)^{1/2}(\bar{X} - \bar{Y})$  的渐近分布.

**定理1** 设  $p$  维实数向量  $\mathbf{l} = (l_1, l_2, \dots, l_p)'_{p \times 1}$  满足  $\|\mathbf{l}\|_2^2 = \sum_{s=1}^p l_s^2 < M_p$  ( $M_p$  是与  $p$  相关的有限正实数), 定义  $\mathbf{V} = \mathbf{l}'(\bar{X} - \bar{Y})$ . 如果正整数  $p$ , 存在一个正实数  $\alpha$ , 满足条件  $\frac{\mathbf{l}'\Sigma_1\mathbf{l}}{p^\alpha} < C$  和  $\frac{\mathbf{l}'\Sigma_2\mathbf{l}}{p^\alpha} < C$ , 其中  $C$  是有限正常数,  $\Sigma_1, \Sigma_2$  分别是  $\mathbf{X}, \mathbf{Y}$  的总体协方差矩阵. 记  $h_1(l_1, l_2, \dots, l_p)$  为  $\left(\frac{(m+n)^{1/2}(\bar{X}_1 - \bar{Y}_1)}{p^{\alpha/2}}, \dots, \frac{(m+n)^{1/2}(\bar{X}_p - \bar{Y}_p)}{p^{\alpha/2}}\right)$  的特征函数. 当  $m, n, p \rightarrow \infty$  时, 在零假设下有:

$$\log[h_1(l_1, l_2, \dots, l_p)] = -\frac{\mathbf{l}'\left(\frac{\Sigma_1}{\eta} + \frac{\Sigma_2}{1-\eta}\right)\mathbf{l}}{2p^\alpha} + o(1).$$

**证明:** 对任意  $p$  维实数向量  $\mathbf{l} = (l_1, l_2, \dots, l_p)'_{p \times 1}$  满足  $\|\mathbf{l}\|_2^2 = \sum_{s=1}^p l_s^2 < M_p$ , 定义  $f_X(t)$ ,  $f_Y(t)$  分别为一维随机变量  $\mathbf{l}'\mathbf{X} = \sum_{q=1}^p l_q X_q$ ,  $\mathbf{l}'\mathbf{Y} = \sum_{q=1}^p l_q Y_q$  的特征函数. 如果对任意正整数  $p$ , 存在一个正实数  $\alpha$ , 满足条件  $\frac{\mathbf{l}'\Sigma_1\mathbf{l}}{p^\alpha} < C$  和  $\frac{\mathbf{l}'\Sigma_2\mathbf{l}}{p^\alpha} < C$ , 其中  $C$  是一个有限正值常数. 则我们可以得到

$$f_X(t) = 1 + \frac{\mathbf{l}'\boldsymbol{\mu}_1}{p^{\alpha/2}}(it) - \frac{\mathbf{l}'\Sigma_1\mathbf{l}}{2p^\alpha}t^2 + o(t^2),$$

和

$$f_Y(t) = 1 + \frac{\mathbf{l}'\boldsymbol{\mu}_2}{p^{\alpha/2}}(it) - \frac{\mathbf{l}'\Sigma_2\mathbf{l}}{2p^\alpha}t^2 + o(t^2).$$

$\left(\frac{(m+n)^{1/2}(\bar{X}_1 - \bar{Y}_1)}{p^{\alpha/2}} + \dots + \frac{(m+n)^{1/2}(\bar{X}_p - \bar{Y}_p)}{p^{\alpha/2}}\right)$  的特征函数为

$$g_1(t; l_1, l_2, \dots, l_p) = \left[ f_X\left(\frac{(m+n)^{1/2}}{m}t\right) \right]^m \left[ f_Y\left(-\frac{(m+n)^{1/2}}{n}t\right) \right]^n$$



$$= \left[ 1 + \frac{\Gamma' \mu_1 (m+n)^{1/2}}{p^{\alpha/2} m} (it) - \frac{\Gamma' \Sigma_1 1 (m+n)}{2p^\alpha m^2} t^2 + o\left(\frac{(m+n)}{m^2} t^2\right) \right]^m$$

$$\left[ 1 - \frac{\Gamma' \mu_2 (m+n)^{1/2}}{p^{\alpha/2} n} (it) - \frac{\Gamma' \Sigma_2 1 (m+n)}{2p^\alpha n^2} t^2 + o\left(\frac{(m+n)}{n^2} t^2\right) \right]^n.$$

在零假设下, 当  $m, n, p \rightarrow \infty$  时, 有

$$\begin{aligned} \log[g_1(l_1, l_2, \dots, l_p)] &= m \log \left[ 1 + \frac{\Gamma' \mu_1 (m+n)^{1/2}}{p^{\alpha/2} m} (it) - \frac{\Gamma' \Sigma_1 1 (m+n)}{2p^\alpha m^2} t^2 + o\left(\frac{(m+n)}{m^2} t^2\right) \right] \\ &\quad + n \log \left[ 1 - \frac{\Gamma' \mu_2 (m+n)^{1/2}}{p^{\alpha/2} n} (it) - \frac{\Gamma' \Sigma_2 1 (m+n)}{2p^\alpha n^2} t^2 + o\left(\frac{(m+n)}{n^2} t^2\right) \right] \\ &= -\frac{\Gamma' \Sigma_1 1 (m+n)}{2p^\alpha m} t^2 - \frac{\Gamma' \Sigma_2 1 (m+n)}{2p^\alpha n} t^2 + o\left(\frac{(m+n)}{m} t^2\right) + o\left(\frac{(m+n)}{n} t^2\right) \\ &= -\frac{\Gamma'(\frac{\Sigma_1}{\theta} + \frac{\Sigma_2}{1-\theta}) 1}{2p^\alpha} t^2 + o(t^2). \end{aligned}$$

即  $\log[g_1(t; l_1, l_2, \dots, l_p)] = -\frac{\Gamma'(\frac{\Sigma_1}{\theta} + \frac{\Sigma_2}{1-\theta}) 1}{2p^\alpha} t^2 + o(t^2)$ , 令  $t = 1$ , 则有

$$\log[h_1(l_1, l_2, \dots, l_p)] = \log[g_1(1; l_1, l_2, \dots, l_p)] = -\frac{\Gamma'(\frac{\Sigma_1}{\theta} + \frac{\Sigma_2}{1-\theta}) 1}{2p^\alpha} + o(1),$$

这就完成了定理1的证明.

在定理1的条件下, 当样本量  $m, n$  很大且维数  $p$  独立于  $m, n$  时,  $(m+n)^{1/2}(\bar{X} - \bar{Y})$  渐近高斯过程. 在定理1中, 存在一个正实数  $\alpha$ , 有  $\frac{\Gamma' \Sigma_1 1}{p^\alpha} < C$  和  $\frac{\Gamma' \Sigma_2 1}{p^\alpha} < C$ , 这样的条件是能被满足的, 因为  $\frac{\Gamma' \Sigma_1 1}{p^\alpha} \leq \frac{\|\mathbb{I}\|_2^2 \max\{\lambda_1, \dots, \lambda_p\}}{p^\alpha} < \frac{M_p \max\{\lambda_1, \dots, \lambda_p\}}{p^\alpha}$ . 根据定理1, 当维数  $p$  大于样本量  $m, n$  或者  $m+n$  时,  $m+n$  趋于无穷的条件下, 统计量  $\frac{(m+n)^{1/2}(\bar{X} - \bar{Y})}{p^{\alpha/2}}$  是渐近多元正态分布的. 当  $p$  很大时,  $(m+n)^{1/2}(\bar{X}_1 - \bar{Y}_1), \dots, (m+n)^{1/2}(\bar{X}_p - \bar{Y}_p)$  渐近一个高斯序列, 因此可以采用 stationary bootstrap 算法运用于学生化列统计量平方和的统计量上.

令  $F_p(q_1, \dots, q_p)$  为变量  $((m+n)^{1/2}(\bar{X}_1 - \bar{Y}_1), \dots, (m+n)^{1/2}(\bar{X}_p - \bar{Y}_p))$  的联合分布函数, 并令  $\Psi_p(q_1, \dots, q_p)$  为  $p$  维的多元标准正态分布函数. 下面考查  $F_p(q_1, \dots, q_p) - \Psi_p(q_1, \dots, q_p)$  的绝对值. 对于  $p < m, n$ , 很多文章已经估计了  $|F_p(q_1, \dots, q_p) - \Psi_p(q_1, \dots, q_p)|$  的大小, 例如 Sazonov(1968)<sup>[19]</sup>, Bhattacharya(1975)<sup>[20]</sup>, Nagaev S(1976)<sup>[21]</sup>, Sweeting T(1977)<sup>[22]</sup>, Sazonov V(1980)<sup>[23]</sup>, Sazonov V(1981)<sup>[24]</sup>, Bentkus(1986)<sup>[25]</sup> 和 Bentkus(2003)<sup>[26]</sup>. 对于  $p > m+n$ , 主要有 Chernozhukov(2013)<sup>[27]</sup>



和Chernozhukov(2017)<sup>[28]</sup> 估计过  $|F_p(q_1, \dots, q_p) - \Psi_p(q_1, \dots, q_p)|$  的大小. 所有关于估计  $|F_p(q_1, \dots, q_p) - \Psi_p(q_1, \dots, q_p)|$  的结果表明, 当  $m+n$  趋于  $\infty$  时, 不管  $p$  是大于还是小于  $m+n$ , 在一定条件下,  $|F_p(q_1, \dots, q_p) - \Psi_p(q_1, \dots, q_p)|$  渐近  $\Psi_p(q_1, \dots, q_p)$ .

### 2.3.2 stationary bootstrap算法的应用

基于定理1, 假设  $t_{s,1} = \frac{(\bar{X}_1 - \bar{Y}_1)^2}{\hat{\sigma}_{1:1,1}/m + \hat{\sigma}_{2:1,1}/n}, \dots, t_{s,p} = \frac{(\bar{X}_p - \bar{Y}_p)^2}{\hat{\sigma}_{1:p,p}/m + \hat{\sigma}_{2:p,p}/n}$  是渐近严平稳弱相关时间序列, 其均值在零假设下都渐近等于 1, 其中  $\bar{X}_1, \dots, \bar{X}_p$  为观察值  $\mathbf{X}^{(i)}$  ( $i = 1, \dots, m$ ) 各维的均值,  $\bar{Y}_1, \dots, \bar{Y}_p$  为观察值  $\mathbf{Y}^{(i)}$  ( $i = 1, \dots, n$ ) 各维的均值. 则基于 stationary bootstrap 算法的检验过程如下:

步骤1: 计算  $p$  个学生化列统计量的值  $t_{s,1} = \frac{(\bar{X}_1 - \bar{Y}_1)^2}{\hat{\sigma}_{1:1,1}/m + \hat{\sigma}_{2:1,1}/n}, \dots, t_{s,p} = \frac{(\bar{X}_p - \bar{Y}_p)^2}{\hat{\sigma}_{1:p,p}/m + \hat{\sigma}_{2:p,p}/n}$ , 并计算统计量的值, 记为

$$SBASS_0 = \frac{p(\bar{t}_s - 1)^2}{\bar{v}_s},$$

其中  $\bar{t}_s = \frac{1}{p} \sum_{i=1}^p t_{s,i}, \bar{v}_s = \frac{1}{p-1} \sum_{i=1}^p (t_{s,i} - \bar{t}_s)^2$ .

步骤2: 从步骤1 中利用 stationary bootstrap 方法抽出  $p$  个样本, 记为  $t_{s,1}^*, \dots, t_{s,p}^*$ , 并使用抽出来的样本计算 bootstrap 均值 ( $\bar{t}_{s,(1)}^*$ ) 和方差 ( $\bar{v}_{s,(1)}^*$ ), 接着计算 bootstrap 统计量的值, 记为

$$SBANSS_{(1)}^* = \frac{p(\bar{t}_{s,(1)}^* - \bar{t}_s)^2}{\bar{v}_{s,(1)}^*}.$$

步骤3: 重复步骤2  $B$  次, 得到  $B$  个 bootstrap 统计量值, 记为  $SBANSS_{(1)}^*, SBANSS_{(2)}^*, \dots, SBANSS_{(B)}^*$ .

步骤4: 计算  $P$  值,

$$P_{SB} = \frac{\sum_{b=1}^B I_{\{SBANSS_{(b)}^* > SBASS_0\}}}{B}.$$



## 第三章 统计模拟和实例研究

在这一部分, 将对以上理论进行上机模拟并结合实例应用, 对其结果进行讨论分析.

### 3.1 模拟及分析

在有限样本下, 通过大量模拟来比较 BS, CQ, GCT10(Parzen 滞后窗口长度  $L = 10$ ), GCT15(Parzen 滞后窗口长度  $L = 15$ ), GCT20(Parzen 滞后窗口长度  $L = 20$ ), SR 以及 SBASS 检验的效果.

#### 3.1.1 模拟

(1) 从  $p$  维正态分布中分别产生  $m$  个独立样本  $(X_{i1}, \dots, X_{ip})^T (i = 1, \dots, m)$ ,  $n$  个独立样本  $(Y_{i1}, \dots, Y_{ip})^T (i = 1, \dots, n)$ , 其均值分别为  $\boldsymbol{\mu}_1 = (\mu_{1,1}, \dots, \mu_{1,p})^T$  和  $\boldsymbol{\mu}_2 = (\mu_{2,1}, \dots, \mu_{2,p})^T$ , 协方差分别为  $\boldsymbol{\Sigma}_1$  和  $\boldsymbol{\Sigma}_2$  (以下简称正态分布);

(2) 从  $p$  维 Laplace 分布中分别产生  $m$  个独立样本  $(X_{i1}, \dots, X_{ip})^T (i = 1, \dots, m)$ ,  $n$  个独立样本  $(Y_{i1}, \dots, Y_{ip})^T (i = 1, \dots, n)$ , 其均值分别为  $\boldsymbol{\mu}_1 = (\mu_{1,1}, \dots, \mu_{1,p})^T$  和  $\boldsymbol{\mu}_2 = (\mu_{2,1}, \dots, \mu_{2,p})^T$ , 协方差分别为  $\boldsymbol{\Sigma}_1$  和  $\boldsymbol{\Sigma}_2$  (以下简称拉普拉斯分布);

(3) 从  $p$  维 Laplace 分布中产生  $m$  个独立样本  $(X_{i1}, \dots, X_{ip})^T (i = 1, \dots, m)$ , 其均值和协方差分别为  $\boldsymbol{\mu}_1 = (\mu_{1,1}, \dots, \mu_{1,p})^T$  和  $\boldsymbol{\Sigma}_1$ ; 从  $p$  维正态分布中产生  $n$  个独立样本  $(Y_{i1}, \dots, Y_{ip})^T (i = 1, \dots, n)$ , 其均值和协方差分别为  $\boldsymbol{\mu}_2 = (\mu_{2,1}, \dots, \mu_{2,p})^T$  和  $\boldsymbol{\Sigma}_2$  (以下简称混合分布).

针对经验水平, 令  $\boldsymbol{\mu}_1 = (\mu_{1,1}, \dots, \mu_{1,p})^T = \boldsymbol{\mu}_2 = (\mu_{2,1}, \dots, \mu_{2,p})^T = (0, \dots, 0)^T$ ; 针对经验功效, 令  $\boldsymbol{\mu}_1 = (\mu_{1,1}, \dots, \mu_{1,p})^T = (0, \dots, 0)^T$ ,  $\boldsymbol{\mu}_2 = (\mu_{2,1}, \dots, \mu_{2,p})^T$  有  $[p^{0.8}]$  个非零项, 这些非零项从  $\{1, 2, \dots, p\}$  中均匀随机抽取, 其中一半非零项设置为 0.23, 另一半非零项设置为 -0.23.

考虑一般性, 这里采用两种协方差结构. 令  $\boldsymbol{D} = (d_{a,s})_{q,s=1,\dots,p}$  为对角阵, 其对角元素  $d_{s,s} (s = 1, 2, \dots, p)$  产生于均匀分布 (1, 3). 两种协方差结构的设置如下:

模型1: 致密结构;  $\boldsymbol{\Sigma}_1 = \boldsymbol{D}^{1/2} \boldsymbol{R}_1 \boldsymbol{D}^{1/2}$ , 和  $\boldsymbol{\Sigma}_2 = \boldsymbol{D}^{1/2} \boldsymbol{R}_2 \boldsymbol{D}^{1/2}$ , 其中  $\boldsymbol{R}_1 = (\rho_1^{|q-s|})_{q,s=1,\dots,p}$  和  $\boldsymbol{R}_2 = (\rho_2^{|q-s|})_{q,s=1,\dots,p}$ ,  $(\rho_1, \rho_2 = 0, 0.1, 0.2, \dots, 0.5)$ ;

模型2: 稀疏结构;  $\boldsymbol{\Sigma}_1 = \boldsymbol{D}^{1/2} \boldsymbol{R}_1^{-1} \boldsymbol{D}^{1/2}$  和  $\boldsymbol{\Sigma}_2 = \boldsymbol{D}^{1/2} \boldsymbol{R}_2^{-1} \boldsymbol{D}^{1/2}$ , 其中  $\boldsymbol{R}_1 = (\rho_1^{|q-s|})_{q,s=1,\dots,p}$  和  $\boldsymbol{R}_2 = (\rho_2^{|q-s|})_{q,s=1,\dots,p}$ ,  $(\rho_1, \rho_2 = 0, 0.1, 0.2, \dots, 0.5)$ .



### 3.1.2 模拟结果及分析

令  $m = 60, n = 80, p = 300$ , 经验水平和经验功效也都基于  $\alpha = 0.05$  时重复 2000 次计算(代码见附录 1). 由于 GCT15, GCT20 的性能都比较差(见下文表 3.5 – 3.16), 所以在下文中的图 3.1, 表 3.1 – 3.4 剔除掉这两个统计量的经验水平和经验功效.

从图 3.1 中的 12 张图可以看出, 随着两样本变量间相关性的增大, GCT10 的变化范围在 0.06 – 0.08 之间, 有些膨胀甚至于控制不住, 而 SR 的变化范围大致在 0.03 – 0.045, 对第一类错误的控制过于保守. Bai, CQ 和 SBASS 统计量的经验水平控制都在合理范围内变动, 随着变量间相关性的增大, 其中 Bai 和 CQ 的第一类错误都在 0.05 左右, 而 SBASS 统计量的第一类错误控制稍微有所变动, 但在两样本变量间相关性增大时, 其经验功效比 Bai 和 CQ 的经验功效明显要好.

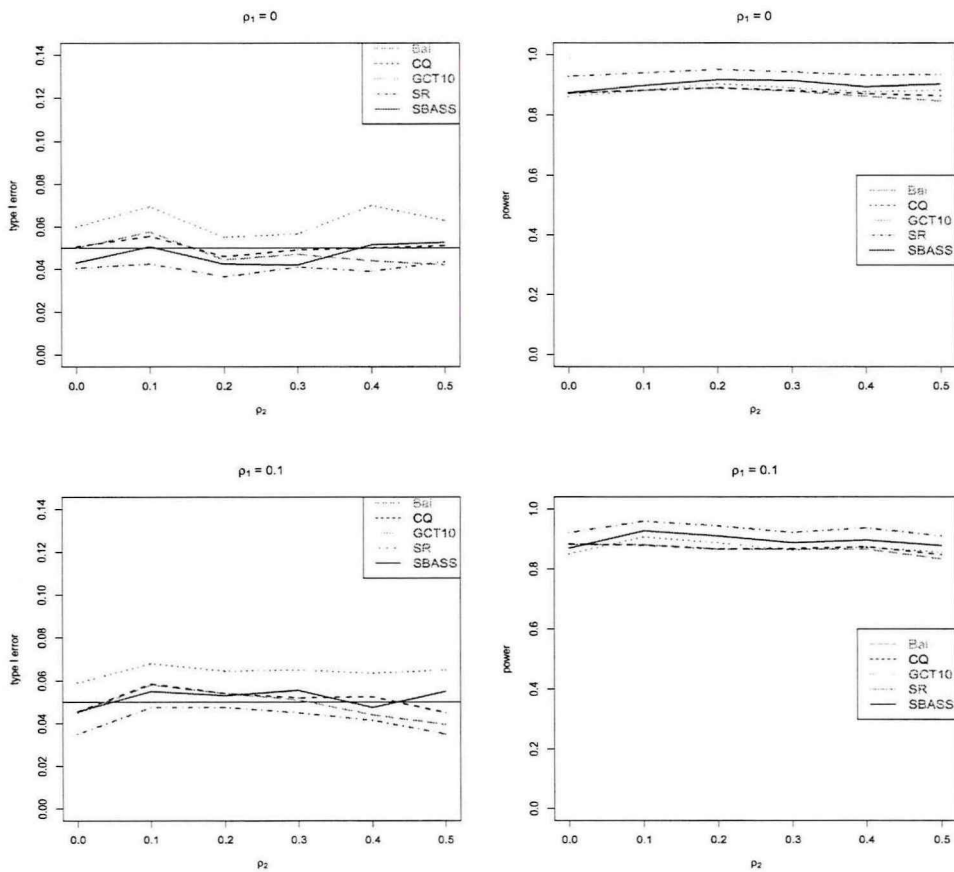


图 3.1 基于模型 1 在正态分布下各统计量的第一类型错误率和功效

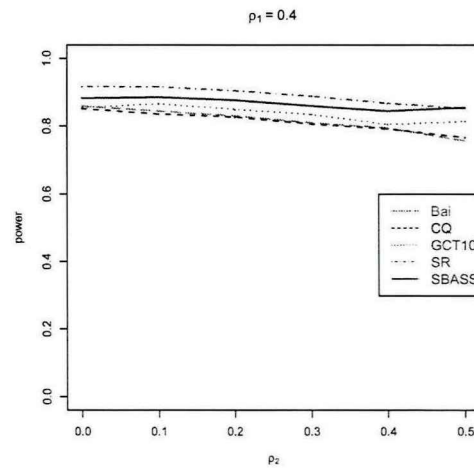
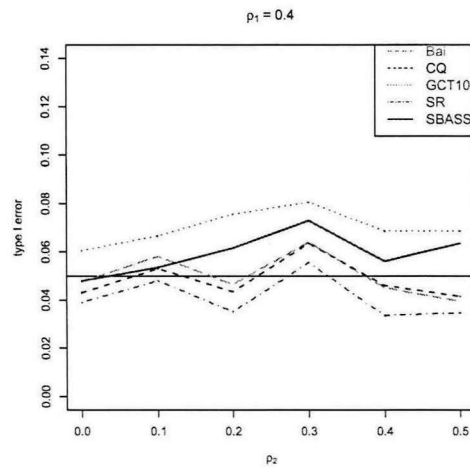
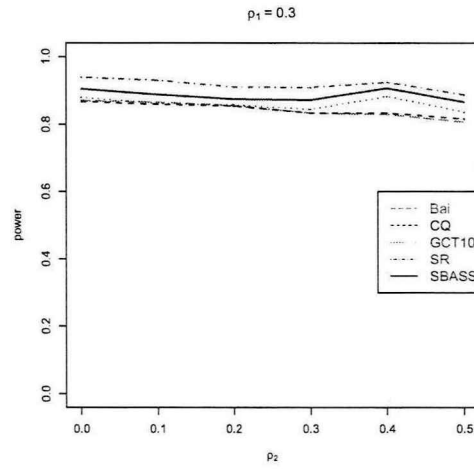
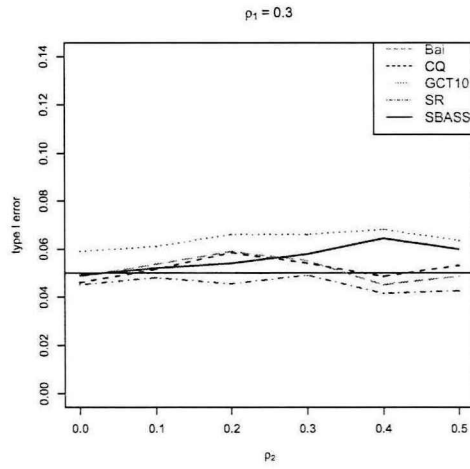
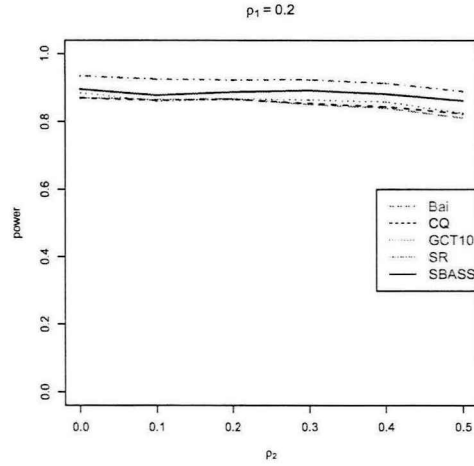
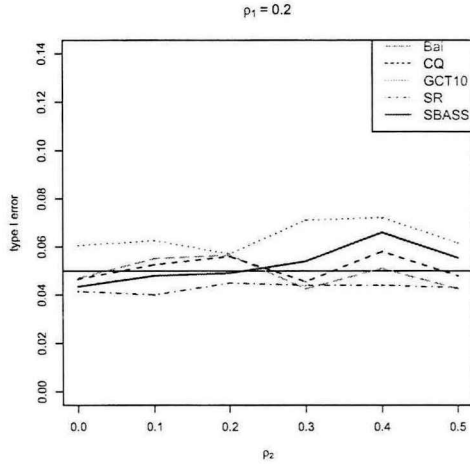


图 3.1 续图

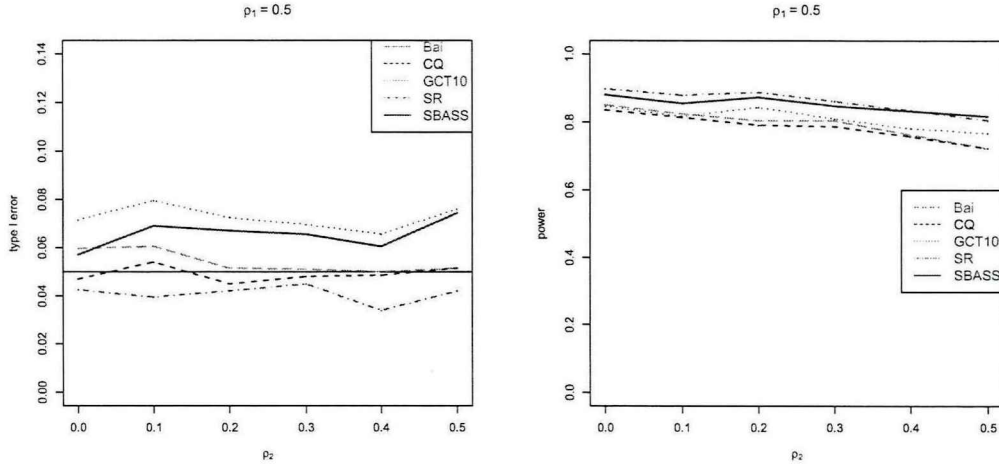


图 3.1 续图

表3.1 基于模型 1 下统计量的经验水平和经验功效

$\rho_1$	$\rho_2$	样本总体	size				
			Bai	CQ	GCT10	SR	SBASS
0.5	0	正态分布	0.0595	0.047	0.0715	0.0425	0.057
		拉普拉斯分布	0.06	0.0525	0.07	0.044	0.062
		混合分布	0.0545	0.046	0.073	0.0365	0.0625
$\rho_1$	$\rho_2$	样本总体	power				
			Bai	CQ	GCT10	SR	SBASS
0.5	0	正态分布	0.8505	0.836	0.8465	0.8985	0.881
		拉普拉斯分布	0.8635	0.8435	0.8865	0.9395	0.9195
		混合分布	0.8505	0.8385	0.8485	0.9055	0.8795
$\rho_1$	$\rho_2$	样本总体	size				
			Bai	CQ	GCT10	SR	SBASS
0	0.5	正态分布	0.042	0.051	0.063	0.0435	0.0525
		拉普拉斯分布	0.0355	0.0475	0.064	0.0335	0.0505
		混合分布	0.04	0.0705	0.0725	0.04	0.051
$\rho_1$	$\rho_2$	样本总体	power				
			Bai	CQ	GCT10	SR	SBASS
0	0.5	正态分布	0.8465	0.864	0.8825	0.934	0.902
		拉普拉斯分布	0.8345	0.858	0.8865	0.9375	0.9075
		混合分布	0.841	0.8605	0.8555	0.914	0.8885

从表 3.1 可知, 在总体分布是正态分布时, 各个统计量除了 GCT10 的经验水平控制较为膨胀之外, 其余的都比较好, 但总体分布为拉普拉斯分布和混合分布时, Bai, CQ 和 SR 的经验水平变动较大, 其中 Bai 和 SR 都出现了过于保守的情况, 而



GCT10 的经验水平仍旧过于膨胀; 相比较之下, SBASS 统计量在这三种总体假设下的经验水平和经验功效不受影响.

表 3.2: 基于模型 2 下统计量的经验水平和经验功效

$\rho_1$	$\rho_2$	样本总体	size				
			Bai	CQ	GCT10	SR	SBASS
0.5	0	正态分布	0.127	0.0455	0.0635	0.039	0.053
		拉普拉斯分布	0.134	0.0505	0.0675	0.04	0.0535
		混合分布	0.139	0.0455	0.063	0.0365	0.052
$\rho_1$	$\rho_2$	样本总体	power				
			Bai	CQ	GCT10	SR	SBASS
0.5	0	正态分布	0.8415	0.5965	0.573	0.681	0.62
		拉普拉斯分布	0.836	0.58	0.5995	0.7145	0.666
		混合分布	0.8515	0.616	0.604	0.711	0.6665
$\rho_1$	$\rho_2$	样本总体	size				
			Bai	CQ	GCT10	SR	SBASS
0	0.5	正态分布	0.0755	0.037	0.054	0.0305	0.044
		拉普拉斯分布	0.079	0.0525	0.076	0.042	0.0555
		混合分布	0.0795	0.054	0.069	0.0445	0.0545
$\rho_1$	$\rho_2$	样本总体	power				
			Bai	CQ	GCT10	SR	SBASS
0	0.5	正态分布	0.3005	0.6675	0.61	0.72	0.6505
		拉普拉斯分布	0.295	0.661	0.657	0.7885	0.719
		混合分布	0.326	0.7025	0.723	0.8285	0.759

从表 3.1, 3.2 可以看出, 不同的模型, 即不同协方差结构下, Bai 统计量的经验水平控制变化很大, 在第二种协方差结构下, 已经经验水平控制不住, 所以所用适用的情况相对有限; 其次 SR 统计量的受协方差结构的影响比较大, 在第二种协方差结构下, 虽然有不错的模拟功效, 但第一类错误控制不住; 另外就第一类错误的角度看, CQ, GCT10 和 SBASS 这三个统计量受协方差结构的影响并不是很大, 但 SBASS 统计量的模拟功效明显地比 CQ, GCT10 的要好一些.

从表 3.3 可以看出, Bai 统计量的经验水平随着维数的增加而变得越来越控制不住, GCT10 统计量的经验水平有一点膨胀而 SR 统计量偏保守, 就 CQ 和 SBASS 统计量的经验水平控制很好, 但 SBASS 统计量的功效比 CQ 表现得更好一些.

从表 3.4 可以知道, 随着样本量的减少, Bai 的经验水平越来越难以控制, SR 偏保守, GCT10 偏膨胀; 在功效上, SBASS 统计量要优于 CQ 的功效.





表 3.3: 基于模型 1 在正态分布下统计量的经验水平和经验功效

说明:  $m = 60, n = 80, S = 2000$

$\rho_1$	$\rho_2$	维数 $p$	size				
			Bai	CQ	GCT10	SR	SBASS
0.5	0	300	0.0595	0.047	0.0715	0.0425	0.057
		500	0.0495	0.0395	0.0585	0.035	0.0525
		700	0.0203	0.0435	0.055	0.0395	0.059
$\rho_1$	$\rho_2$	维数 $p$	power				
			Bai	CQ	GCT10	SR	SBASS
0.5	0	300	0.8505	0.836	0.8465	0.8985	0.881
		500	0.9335	0.9215	0.9325	0.9635	0.96
		700	0.974	0.81	0.8585	0.9095	0.915
$\rho_1$	$\rho_2$	维数 $p$	size				
			Bai	CQ	GCT10	SR	SBASS
0	0.5	300	0.042	0.051	0.063	0.0435	0.0525
		500	0.0435	0.0535	0.064	0.0445	0.058
		700	0.1265	0.047	0.052	0.0395	0.0515
$\rho_1$	$\rho_2$	维数 $p$	power				
			Bai	CQ	GCT10	SR	SBASS
0	0.5	300	0.8465	0.864	0.8825	0.934	0.902
		500	0.928	0.9395	0.9515	0.975	0.9695
		700	0.3485	0.8735	0.902	0.952	0.945

表 3.4: 基于模型 1 在正态分布下统计量的经验水平和经验功效

说明:  $m = 60, p = 300, S = 2000$

$\rho_1$	$\rho_2$	样本量 $n$	size				
			Bai	CQ	GCT10	SR	SBASS
0.5	0	80	0.0595	0.047	0.0715	0.0425	0.057
		60	0.0465	0.0465	0.0645	0.0405	0.0495
		40	0.0365	0.0445	0.0605	0.037	0.047
$\rho_1$	$\rho_2$	样本量 $n$	power				
			Bai	CQ	GCT10	SR	SBASS
0.5	0	80	0.8505	0.836	0.8465	0.8985	0.881
		60	0.751	0.751	0.747	0.836	0.7955
		40	0.5495	0.5875	0.5795	0.726	0.662
$\rho_1$	$\rho_2$	样本量 $n$	size				
			Bai	CQ	GCT10	SR	SBASS
0	0.5	80	0.042	0.051	0.063	0.0435	0.0525
		60	0.0345	0.045	0.0665	0.037	0.059
		40	0.2005	0.0555	0.062	0.0475	0.0535
$\rho_1$	$\rho_2$	样本量 $n$	power				
			Bai	CQ	GCT10	SR	SBASS
0	0.5	80	0.8465	0.864	0.8825	0.934	0.902
		60	0.486	0.522	0.51	0.6325	0.563
		40	0.7585	0.3515	0.3475	0.482	0.4405



表 3.5: 基于模型1 在正态分布下统计量的经验水平

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.05	0.0505	0.06	0.0635	0.0655	0.0405	0.043
0	0.1	0.0575	0.0555	0.0695	0.069	0.074	0.0425	0.0505
0	0.2	0.0445	0.046	0.055	0.061	0.0645	0.0365	0.0425
0	0.3	0.047	0.049	0.0565	0.0615	0.065	0.041	0.042
0	0.4	0.044	0.05	0.07	0.074	0.0785	0.0739	0.0515
0	0.5	0.042	0.051	0.063	0.0645	0.0665	0.0435	0.0525
0.1	0	0.045	0.0455	0.059	0.063	0.0645	0.035	0.0455
0.1	0.1	0.058	0.0585	0.068	0.073	0.076	0.0475	0.055
0.1	0.2	0.054	0.054	0.0645	0.071	0.075	0.0475	0.053
0.1	0.3	0.051	0.052	0.065	0.0675	0.067	0.045	0.0555
0.1	0.4	0.044	0.0525	0.0635	0.0685	0.074	0.0415	0.0475
0.1	0.5	0.0395	0.045	0.065	0.069	0.0725	0.035	0.055
0.2	0	0.047	0.0465	0.0605	0.066	0.067	0.0415	0.0435
0.2	0.1	0.055	0.0525	0.0625	0.065	0.0695	0.04	0.048
0.2	0.2	0.0565	0.056	0.057	0.062	0.0685	0.045	0.049
0.2	0.3	0.0425	0.0455	0.071	0.0715	0.074	0.044	0.054
0.2	0.4	0.051	0.058	0.072	0.0745	0.0775	0.044	0.066
0.2	0.5	0.0425	0.048	0.0615	0.0685	0.0705	0.043	0.0555
0.3	0	0.0485	0.046	0.059	0.0655	0.066	0.045	0.049
0.3	0.1	0.0535	0.0515	0.061	0.063	0.0675	0.048	0.052
0.3	0.2	0.059	0.0585	0.066	0.0715	0.075	0.0455	0.054
0.3	0.3	0.055	0.054	0.066	0.068	0.074	0.049	0.058
0.3	0.4	0.045	0.0485	0.068	0.069	0.0755	0.0415	0.0645
0.3	0.5	0.0485	0.053	0.0635	0.0645	0.066	0.0425	0.06
0.4	0	0.0475	0.043	0.0605	0.063	0.067	0.039	0.048
0.4	0.1	0.058	0.053	0.0665	0.0705	0.0735	0.048	0.0535
0.4	0.2	0.0465	0.0435	0.0755	0.0775	0.081	0.035	0.0615
0.4	0.3	0.064	0.0635	0.0805	0.085	0.0905	0.0555	0.073
0.4	0.4	0.045	0.046	0.0685	0.069	0.073	0.0335	0.056
0.4	0.5	0.0395	0.0415	0.0685	0.07	0.0735	0.0345	0.0635
0.5	0	0.0595	0.047	0.0715	0.0745	0.075	0.0425	0.057
0.5	0.1	0.0605	0.054	0.0795	0.08	0.0825	0.0395	0.069
0.5	0.2	0.0515	0.045	0.0725	0.074	0.0805	0.042	0.067
0.5	0.3	0.051	0.048	0.0695	0.073	0.075	0.045	0.0655
0.5	0.4	0.05	0.0485	0.0655	0.0655	0.066	0.034	0.0605
0.5	0.5	0.0515	0.0515	0.076	0.0765	0.079	0.042	0.0745



表 3.6: 基于模型1 在拉普拉斯分布下统计量的经验水平

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.0445	0.0495	0.063	0.069	0.0705	0.0415	0.045
0	0.1	0.0455	0.048	0.0655	0.0675	0.0715	0.0355	0.043
0	0.2	0.05	0.0525	0.066	0.0715	0.0755	0.046	0.05
0	0.3	0.044	0.052	0.067	0.0695	0.074	0.0385	0.0495
0	0.4	0.0545	0.061	0.071	0.074	0.0765	0.043	0.053
0	0.5	0.0355	0.0475	0.064	0.0695	0.0725	0.0335	0.0505
0.1	0	0.0515	0.054	0.0655	0.072	0.0765	0.047	0.0435
0.1	0.1	0.0475	0.0485	0.07	0.0715	0.0745	0.0455	0.0465
0.1	0.2	0.041	0.0435	0.06	0.066	0.0685	0.036	0.043
0.1	0.3	0.0435	0.047	0.07	0.0755	0.08	0.0405	0.0495
0.1	0.4	0.041	0.0515	0.0755	0.08	0.086	0.0425	0.058
0.1	0.5	0.036	0.042	0.07	0.073	0.077	0.038	0.0545
0.2	0	0.054	0.056	0.0725	0.077	0.0815	0.0445	0.058
0.2	0.1	0.044	0.046	0.0645	0.0675	0.0695	0.039	0.0465
0.2	0.2	0.0425	0.0445	0.059	0.064	0.069	0.0335	0.047
0.2	0.3	0.043	0.047	0.06	0.0635	0.067	0.0355	0.044
0.2	0.4	0.046	0.0495	0.0695	0.0745	0.081	0.041	0.0585
0.2	0.5	0.0415	0.0485	0.067	0.0695	0.0745	0.038	0.054
0.3	0	0.0495	0.0485	0.0715	0.0725	0.073	0.0395	0.056
0.3	0.1	0.0505	0.05	0.072	0.0785	0.08	0.034	0.056
0.3	0.2	0.0515	0.0515	0.084	0.0865	0.09	0.0395	0.0635
0.3	0.3	0.048	0.0495	0.067	0.069	0.073	0.038	0.053
0.3	0.4	0.046	0.0495	0.0675	0.071	0.0735	0.043	0.0605
0.3	0.5	0.041	0.051	0.0755	0.076	0.079	0.041	0.0655
0.4	0	0.058	0.0545	0.0655	0.065	0.0675	0.0395	0.0545
0.4	0.1	0.0565	0.053	0.0695	0.0725	0.077	0.0405	0.059
0.4	0.2	0.045	0.044	0.064	0.0665	0.074	0.036	0.053
0.4	0.3	0.0455	0.045	0.0635	0.067	0.0705	0.033	0.057
0.4	0.4	0.0525	0.057	0.0745	0.0795	0.0845	0.0445	0.0725
0.4	0.5	0.047	0.049	0.0775	0.079	0.079	0.033	0.0645
0.5	0	0.06	0.0525	0.07	0.0755	0.08	0.044	0.062
0.5	0.1	0.0565	0.0495	0.071	0.074	0.076	0.0425	0.061
0.5	0.2	0.049	0.0455	0.071	0.0735	0.0765	0.035	0.0585
0.5	0.3	0.0525	0.0485	0.0725	0.074	0.0765	0.04	0.074
0.5	0.4	0.05	0.0475	0.0705	0.0705	0.073	0.0365	0.061
0.5	0.5	0.0505	0.051	0.0725	0.0735	0.0775	0.0385	0.071



表 3.7: 基于模型1 在混合分布下统计量的经验水平

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.04	0.0425	0.054	0.0555	0.0595	0.033	0.0355
0	0.1	0.048	0.049	0.066	0.0655	0.072	0.0435	0.0505
0	0.2	0.057	0.061	0.0725	0.072	0.076	0.0465	0.06
0	0.3	0.04	0.044	0.0595	0.0645	0.0655	0.0425	0.0465
0	0.4	0.0435	0.0485	0.057	0.064	0.067	0.037	0.0415
0	0.5	0.04	0.047	0.0705	0.0725	0.0755	0.04	0.051
0.1	0	0.049	0.0475	0.062	0.065	0.0685	0.0435	0.0475
0.1	0.1	0.0455	0.046	0.061	0.065	0.0695	0.0355	0.0475
0.1	0.2	0.042	0.044	0.0725	0.077	0.079	0.039	0.05
0.1	0.3	0.044	0.048	0.0525	0.055	0.0585	0.039	0.0425
0.1	0.4	0.0405	0.0465	0.068	0.072	0.0755	0.034	0.047
0.1	0.5	0.0325	0.0375	0.0645	0.0635	0.067	0.033	0.0495
0.2	0	0.0505	0.049	0.074	0.0765	0.0815	0.037	0.0515
0.2	0.1	0.0665	0.066	0.081	0.0865	0.0895	0.05	0.065
0.2	0.2	0.046	0.047	0.072	0.072	0.077	0.0415	0.056
0.2	0.3	0.0555	0.0575	0.07	0.075	0.0775	0.044	0.061
0.2	0.4	0.0485	0.053	0.066	0.0695	0.0765	0.0405	0.0565
0.2	0.5	0.038	0.042	0.079	0.078	0.082	0.037	0.054
0.3	0	0.04	0.04	0.066	0.0685	0.0745	0.033	0.0435
0.3	0.1	0.065	0.063	0.077	0.079	0.083	0.049	0.062
0.3	0.2	0.0395	0.041	0.0685	0.0685	0.072	0.032	0.05
0.3	0.3	0.0425	0.044	0.0645	0.069	0.0715	0.0325	0.0525
0.3	0.4	0.057	0.062	0.086	0.0915	0.095	0.0515	0.0765
0.3	0.5	0.0515	0.056	0.075	0.08	0.0865	0.044	0.0705
0.4	0	0.05	0.0445	0.0775	0.0825	0.0875	0.0405	0.0595
0.4	0.1	0.0555	0.0515	0.065	0.0665	0.071	0.0405	0.055
0.4	0.2	0.052	0.0495	0.067	0.0695	0.074	0.031	0.0565
0.4	0.3	0.052	0.051	0.0715	0.078	0.0705	0.0415	0.0625
0.4	0.4	0.044	0.0435	0.068	0.071	0.077	0.0325	0.0555
0.4	0.5	0.05	0.052	0.0755	0.079	0.0835	0.044	0.0715
0.5	0	0.0545	0.046	0.073	0.073	0.0755	0.0365	0.0625
0.5	0.1	0.055	0.048	0.065	0.069	0.076	0.047	0.064
0.5	0.2	0.0585	0.05	0.0755	0.0785	0.083	0.0395	0.068
0.5	0.3	0.0575	0.0535	0.075	0.075	0.0815	0.039	0.0639
0.5	0.4	0.049	0.046	0.0745	0.076	0.079	0.038	0.0655
0.5	0.5	0.0545	0.055	0.081	0.083	0.09	0.042	0.0715



表 3.8: 基于模型2 在正态分布下统计量的经验水平

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.0425	0.0425	0.054	0.056	0.0575	0.0335	0.043
0	0.1	0.0455	0.0465	0.061	0.0635	0.066	0.0385	0.0435
0	0.2	0.043	0.046	0.0635	0.068	0.0705	0.039	0.0455
0	0.3	0.0495	0.057	0.074	0.079	0.082	0.0425	0.0515
0	0.4	0.059	0.047	0.0665	0.067	0.0705	0.0365	0.0495
0	0.5	0.0755	0.037	0.054	0.0575	0.0645	0.0305	0.044
0.1	0	0.0575	0.0545	0.0675	0.074	0.074	0.0455	0.0515
0.1	0.1	0.043	0.0425	0.059	0.063	0.0665	0.042	0.0425
0.1	0.2	0.04	0.0445	0.058	0.0655	0.07	0.0435	0.0465
0.1	0.3	0.0505	0.0575	0.0775	0.0795	0.0835	0.0535	0.057
0.1	0.4	0.062	0.0485	0.0755	0.0775	0.0825	0.042	0.0575
0.1	0.5	0.0675	0.0525	0.065	0.073	0.0765	0.044	0.06
0.2	0	0.048	0.046	0.0665	0.0695	0.074	0.0415	0.049
0.2	0.1	0.061	0.0575	0.0655	0.067	0.0705	0.045	0.0505
0.2	0.2	0.05	0.049	0.0615	0.065	0.07	0.0405	0.048
0.2	0.3	0.042	0.041	0.061	0.063	0.068	0.041	0.048
0.2	0.4	0.0475	0.0425	0.0665	0.0735	0.0745	0.0345	0.0505
0.2	0.5	0.067	0.0495	0.071	0.072	0.073	0.0395	0.057
0.3	0	0.069	0.053	0.0625	0.067	0.0705	0.036	0.044
0.3	0.1	0.065	0.048	0.0555	0.0585	0.0625	0.046	0.0435
0.3	0.2	0.0585	0.049	0.06	0.0635	0.071	0.0425	0.055
0.3	0.3	0.055	0.053	0.0715	0.0745	0.077	0.0425	0.064
0.3	0.4	0.0495	0.051	0.0645	0.067	0.068	0.042	0.0535
0.3	0.5	0.06	0.051	0.081	0.084	0.0885	0.048	0.069
0.4	0	0.091	0.042	0.0615	0.0625	0.0675	0.039	0.0445
0.4	0.1	0.084	0.047	0.0695	0.075	0.077	0.042	0.0555
0.4	0.2	0.081	0.045	0.0585	0.0625	0.0655	0.04	0.047
0.4	0.3	0.0555	0.0455	0.062	0.066	0.0695	0.0355	0.051
0.4	0.4	0.0475	0.0475	0.0775	0.0815	0.0845	0.0365	0.0565
0.4	0.5	0.0475	0.052	0.0675	0.072	0.0755	0.037	0.059
0.5	0	0.127	0.0455	0.0635	0.067	0.0705	0.039	0.053
0.5	0.1	0.1155	0.0495	0.065	0.069	0.071	0.0425	0.0585
0.5	0.2	0.1185	0.046	0.067	0.0685	0.0705	0.0455	0.0575
0.5	0.3	0.094	0.0525	0.074	0.0765	0.082	0.0395	0.0675
0.5	0.4	0.068	0.05	0.0715	0.075	0.078	0.046	0.062
0.5	0.5	0.045	0.0455	0.0735	0.0755	0.076	0.0465	0.0655



表 3.9: 基于模型2 在拉普拉斯分布下统计量的经验水平

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.043	0.0465	0.068	0.0725	0.075	0.0365	0.046
0	0.1	0.04	0.043	0.0595	0.066	0.068	0.039	0.044
0	0.2	0.0445	0.046	0.064	0.0685	0.076	0.0315	0.0445
0	0.3	0.049	0.0495	0.063	0.066	0.0705	0.037	0.046
0	0.4	0.051	0.0425	0.0615	0.068	0.0725	0.035	0.047
0	0.5	0.079	0.0525	0.076	0.0775	0.0775	0.042	0.0555
0.1	0	0.0405	0.0415	0.065	0.0665	0.0675	0.033	0.0435
0.1	0.1	0.0385	0.0405	0.0605	0.065	0.0685	0.035	0.0415
0.1	0.2	0.0445	0.05	0.067	0.0695	0.0725	0.041	0.048
0.1	0.3	0.0475	0.054	0.079	0.0845	0.0885	0.0345	0.0565
0.1	0.4	0.048	0.049	0.0665	0.066	0.0705	0.043	0.053
0.1	0.5	0.0615	0.045	0.0675	0.073	0.0785	0.044	0.0575
0.2	0	0.057	0.0515	0.075	0.0775	0.0795	0.04	0.054
0.2	0.1	0.056	0.052	0.081	0.0855	0.087	0.043	0.055
0.2	0.2	0.049	0.051	0.072	0.073	0.077	0.039	0.057
0.2	0.3	0.047	0.049	0.0605	0.0645	0.07	0.037	0.047
0.2	0.4	0.059	0.0515	0.077	0.084	0.089	0.0365	0.0635
0.2	0.5	0.0555	0.05	0.066	0.0675	0.072	0.035	0.0495
0.3	0	0.068	0.049	0.065	0.072	0.0755	0.045	0.0505
0.3	0.1	0.0625	0.0485	0.0715	0.0755	0.08	0.0395	0.054
0.3	0.2	0.0575	0.0485	0.0705	0.074	0.0735	0.0455	0.05
0.3	0.3	0.043	0.045	0.068	0.07	0.075	0.036	0.0505
0.3	0.4	0.048	0.053	0.069	0.0705	0.0725	0.04	0.055
0.3	0.5	0.0515	0.062	0.0705	0.0775	0.0825	0.0415	0.064
0.4	0	0.087	0.0495	0.064	0.0675	0.0725	0.0375	0.051
0.4	0.1	0.095	0.0555	0.069	0.072	0.0755	0.043	0.0535
0.4	0.2	0.072	0.0535	0.075	0.076	0.0785	0.0415	0.0565
0.4	0.3	0.0635	0.058	0.07	0.075	0.081	0.0405	0.0645
0.4	0.4	0.042	0.044	0.067	0.0695	0.0725	0.0355	0.055
0.4	0.5	0.0385	0.044	0.065	0.0685	0.073	0.026	0.0485
0.5	0	0.134	0.0505	0.0675	0.069	0.0735	0.04	0.0535
0.5	0.1	0.129	0.0515	0.0685	0.0715	0.0735	0.043	0.0505
0.5	0.2	0.116	0.0495	0.0715	0.0685	0.0685	0.0455	0.056
0.5	0.3	0.089	0.0505	0.0745	0.081	0.0845	0.0385	0.066
0.5	0.4	0.0665	0.0555	0.065	0.068	0.072	0.037	0.0595
0.5	0.5	0.0385	0.0405	0.0635	0.066	0.0695	0.033	0.0535



表 3.10: 基于模型2 在混合分布下统计量的经验水平

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.0475	0.0515	0.0745	0.0775	0.081	0.041	0.0495
0	0.1	0.0425	0.0425	0.0595	0.062	0.065	0.0385	0.0395
0	0.2	0.0475	0.0475	0.067	0.0705	0.072	0.0365	0.0455
0	0.3	0.0505	0.055	0.0685	0.073	0.076	0.045	0.055
0	0.4	0.061	0.051	0.071	0.0725	0.073	0.0405	0.054
0	0.5	0.0795	0.054	0.069	0.0765	0.079	0.0445	0.0545
0.1	0	0.0465	0.0435	0.0645	0.0675	0.0715	0.0355	0.047
0.1	0.1	0.042	0.045	0.0645	0.0655	0.068	0.038	0.0455
0.1	0.2	0.048	0.0515	0.0635	0.0645	0.0675	0.0375	0.048
0.1	0.3	0.047	0.0475	0.0665	0.067	0.07	0.038	0.0515
0.1	0.4	0.0525	0.0455	0.07	0.0725	0.0805	0.04	0.0555
0.1	0.5	0.074	0.0545	0.067	0.0705	0.073	0.0465	0.06
0.2	0	0.058	0.049	0.075	0.079	0.081	0.043	0.054
0.2	0.1	0.0555	0.049	0.06	0.0645	0.0665	0.0435	0.048
0.2	0.2	0.048	0.0495	0.063	0.069	0.0745	0.033	0.0505
0.2	0.3	0.045	0.0495	0.0685	0.0695	0.0775	0.037	0.0505
0.2	0.4	0.053	0.056	0.0725	0.0745	0.0795	0.043	0.053
0.2	0.5	0.061	0.0435	0.0675	0.071	0.0755	0.0315	0.0505
0.3	0	0.0665	0.0475	0.0625	0.0635	0.0675	0.0335	0.046
0.3	0.1	0.066	0.055	0.074	0.0775	0.083	0.0465	0.0565
0.3	0.2	0.059	0.0465	0.0645	0.0695	0.073	0.0415	0.0525
0.3	0.3	0.055	0.056	0.0755	0.077	0.0825	0.044	0.057
0.3	0.4	0.0485	0.0485	0.0705	0.075	0.079	0.0435	0.0605
0.3	0.5	0.0515	0.0555	0.074	0.076	0.0795	0.0415	0.061
0.4	0	0.092	0.0395	0.06	0.062	0.069	0.035	0.0455
0.4	0.1	0.078	0.043	0.074	0.0775	0.0815	0.0425	0.053
0.4	0.2	0.071	0.049	0.069	0.066	0.0695	0.036	0.0545
0.4	0.3	0.0515	0.0415	0.072	0.0755	0.082	0.0335	0.058
0.4	0.4	0.0525	0.055	0.0715	0.0775	0.0815	0.0435	0.06
0.4	0.5	0.0525	0.0585	0.0795	0.083	0.088	0.0435	0.068
0.5	0	0.139	0.0455	0.063	0.0685	0.07	0.0365	0.052
0.5	0.1	0.1155	0.046	0.0665	0.0715	0.0785	0.031	0.05
0.5	0.2	0.106	0.0465	0.0675	0.067	0.074	0.034	0.0535
0.5	0.3	0.0885	0.049	0.07	0.0705	0.076	0.035	0.0555
0.5	0.4	0.0645	0.0485	0.0695	0.0745	0.0765	0.041	0.065
0.5	0.5	0.045	0.0465	0.065	0.071	0.0745	0.0375	0.057



表 3.11: 基于模型1 在正态分布下统计量的经验功效

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.8725	0.8725	0.862	0.863	0.8655	0.928	0.8745
0	0.1	0.8825	0.882	0.882	0.8895	0.8935	0.9395	0.8975
0	0.2	0.891	0.8905	0.903	0.9035	0.903	0.951	0.916
0	0.3	0.879	0.882	0.8905	0.8915	0.8945	0.9425	0.914
0	0.4	0.8625	0.871	0.8775	0.878	0.8765	0.932	0.8935
0	0.5	0.8465	0.864	0.8825	0.8835	0.8855	0.934	0.902
0.1	0	0.8815	0.8835	0.8505	0.8535	0.8585	0.9215	0.87
0.1	0.1	0.881	0.8795	0.907	0.907	0.908	0.96	0.928
0.1	0.2	0.8665	0.867	0.8875	0.885	0.886	0.944	0.9105
0.1	0.3	0.866	0.8685	0.863	0.865	0.8705	0.922	0.8875
0.1	0.4	0.8665	0.8745	0.8715	0.869	0.8705	0.937	0.8965
0.1	0.5	0.8335	0.848	0.8565	0.8525	0.8545	0.9105	0.878
0.2	0	0.871	0.87	0.885	0.8845	0.8865	0.9355	0.8965
0.2	0.1	0.8655	0.8635	0.8605	0.863	0.8695	0.926	0.8785
0.2	0.2	0.8665	0.866	0.867	0.874	0.879	0.9225	0.888
0.2	0.3	0.8505	0.8535	0.8645	0.858	0.857	0.9235	0.893
0.2	0.4	0.84	0.8445	0.8585	0.8555	0.86	0.9135	0.8825
0.2	0.5	0.811	0.8235	0.825	0.8295	0.8295	0.8895	0.8625
0.3	0	0.8695	0.8665	0.8795	0.881	0.887	0.9385	0.9045
0.3	0.1	0.865	0.858	0.861	0.857	0.862	0.9295	0.8875
0.3	0.2	0.854	0.853	0.8575	0.8585	0.8635	0.909	0.8745
0.3	0.3	0.8325	0.8325	0.843	0.8475	0.8505	0.9085	0.8715
0.3	0.4	0.8285	0.8325	0.882	0.878	0.88	0.924	0.907
0.3	0.5	0.806	0.8155	0.836	0.831	0.883	0.8865	0.8655
0.4	0	0.859	0.851	0.854	0.8575	0.8545	0.9165	0.883
0.4	0.1	0.844	0.835	0.8655	0.868	0.8705	0.916	0.886
0.4	0.2	0.8305	0.8265	0.849	0.8475	0.849	0.905	0.877
0.4	0.3	0.8095	0.8055	0.8345	0.8385	0.838	0.8885	0.86
0.4	0.4	0.794	0.792	0.805	0.8045	0.8075	0.8675	0.8445
0.4	0.5	0.7565	0.766	0.8135	0.8155	0.819	0.852	0.8555
0.5	0	0.8505	0.836	0.8465	0.85	0.853	0.8985	0.881
0.5	0.1	0.824	0.814	0.817	0.821	0.826	0.8795	0.8555
0.5	0.2	0.804	0.7905	0.8435	0.8425	0.8435	0.8875	0.874
0.5	0.3	0.804	0.787	0.8095	0.814	0.8125	0.8615	0.848
0.5	0.4	0.763	0.7575	0.7815	0.785	0.7795	0.835	0.833
0.5	0.5	0.7225	0.723	0.7675	0.7625	0.762	0.805	0.8175





表 3.12: 基于模型1 在拉普拉斯分布下统计量的经验功效

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.875	0.879	0.877	0.8805	0.88	0.951	0.9035
0	0.1	0.888	0.8905	0.8835	0.883	0.8835	0.9445	0.904
0	0.2	0.882	0.8875	0.893	0.8935	0.8955	0.9575	0.9265
0	0.3	0.8585	0.8675	0.885	0.89	0.892	0.9585	0.922
0	0.4	0.8375	0.846	0.8575	0.8555	0.858	0.9185	0.88
0	0.5	0.8345	0.858	0.8865	0.888	0.8895	0.9375	0.9075
0.1	0	0.878	0.881	0.9235	0.929	0.928	0.9695	0.9435
0.1	0.1	0.8805	0.881	0.913	0.916	0.916	0.957	0.9315
0.1	0.2	0.8645	0.8675	0.8845	0.8875	0.893	0.943	0.908
0.1	0.3	0.863	0.8725	0.8895	0.863	0.8845	0.9515	0.9165
0.1	0.4	0.838	0.854	0.9085	0.91	0.9135	0.95	0.929
0.1	0.5	0.816	0.8355	0.84	0.842	0.841	0.9045	0.8795
0.2	0	0.879	0.8795	0.9065	0.9105	0.916	0.958	0.925
0.2	0.1	0.877	0.879	0.884	0.887	0.8885	0.9485	0.9105
0.2	0.2	0.8655	0.865	0.8825	0.887	0.889	0.945	0.909
0.2	0.3	0.854	0.8615	0.9075	0.906	0.9065	0.9555	0.9295
0.2	0.4	0.845	0.854	0.872	0.8745	0.875	0.929	0.9025
0.2	0.5	0.8025	0.8235	0.87	0.872	0.876	0.9265	0.9045
0.3	0	0.8645	0.859	0.902	0.9045	0.906	0.945	0.929
0.3	0.1	0.849	0.8495	0.8645	0.8675	0.8725	0.93	0.893
0.3	0.2	0.836	0.8365	0.855	0.8545	0.858	0.924	0.8875
0.3	0.3	0.8365	0.8425	0.881	0.881	0.878	0.932	0.9055
0.3	0.4	0.825	0.83	0.861	0.8675	0.8705	0.922	0.901
0.3	0.5	0.7845	0.8075	0.858	0.857	0.8565	0.9	0.8965
0.4	0	0.866	0.858	0.8985	0.9015	0.9055	0.95	0.929
0.4	0.1	0.8465	0.837	0.8455	0.8435	0.8445	0.91	0.878
0.4	0.2	0.826	0.8235	0.8715	0.867	0.8675	0.923	0.901
0.4	0.3	0.8175	0.8175	0.854	0.853	0.855	0.9025	0.8825
0.4	0.4	0.7865	0.7865	0.8045	0.8015	0.8065	0.8655	0.8545
0.4	0.5	0.766	0.7745	0.843	0.8405	0.8425	0.889	0.888
0.5	0	0.863	0.8435	0.8865	0.889	0.8865	0.9395	0.9195
0.5	0.1	0.8435	0.8275	0.8735	0.869	0.871	0.9225	0.906
0.5	0.2	0.8305	0.816	0.86	0.8595	0.86	0.9065	0.893
0.5	0.3	0.788	0.78	0.807	0.806	0.8085	0.8685	0.8535
0.5	0.4	0.7535	0.7505	0.8065	0.807	0.8115	0.8485	0.8475
0.5	0.5	0.7285	0.731	0.8205	0.8155	0.8175	0.859	0.8715



表 3.13: 基于模型1 在混合分布下统计量的经验功效

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.8765	0.88	0.864	0.8645	0.8685	0.944	0.8935
0	0.1	0.884	0.886	0.9055	0.9055	0.9085	0.953	0.9285
0	0.2	0.8865	0.8875	0.9095	0.902	0.907	0.9645	0.9315
0	0.3	0.885	0.89	0.9125	0.916	0.916	0.966	0.9355
0	0.4	0.8625	0.8715	0.8865	0.887	0.888	0.9405	0.901
0	0.5	0.841	0.8605	0.8555	0.86	0.863	0.914	0.8885
0.1	0	0.8815	0.879	0.881	0.8785	0.8795	0.951	0.9115
0.1	0.1	0.888	0.888	0.889	0.889	0.8885	0.9455	0.9125
0.1	0.2	0.881	0.885	0.8795	0.883	0.884	0.936	0.9025
0.1	0.3	0.873	0.877	0.871	0.8665	0.87	0.9415	0.902
0.1	0.4	0.853	0.8655	0.88	0.884	0.8885	0.9405	0.9075
0.1	0.5	0.8205	0.8385	0.8475	0.8465	0.848	0.9085	0.8805
0.2	0	0.887	0.886	0.9185	0.9195	0.9205	0.959	0.933
0.2	0.1	0.8745	0.8775	0.8835	0.884	0.891	0.9415	0.9075
0.2	0.2	0.8835	0.884	0.877	0.8765	0.8755	0.9345	0.903
0.2	0.3	0.8625	0.866	0.873	0.8755	0.8745	0.9375	0.9035
0.2	0.4	0.839	0.849	0.8505	0.852	0.853	0.913	0.88
0.2	0.5	0.8195	0.8365	0.871	0.8675	0.873	0.919	0.9035
0.3	0	0.891	0.8845	0.888	0.8915	0.892	0.9455	0.9145
0.3	0.1	0.869	0.8655	0.901	0.904	0.9025	0.955	0.925
0.3	0.2	0.853	0.857	0.8655	0.8665	0.8665	0.927	0.896
0.3	0.3	0.8395	0.837	0.846	0.851	0.8485	0.9145	0.8805
0.3	0.4	0.8255	0.8315	0.861	0.859	0.8605	0.9125	0.891
0.3	0.5	0.803	0.8125	0.8705	0.869	0.8705	0.9225	0.9135
0.4	0	0.8815	0.8555	0.8925	0.8955	0.8885	0.9395	0.915
0.4	0.1	0.8465	0.8365	0.855	0.856	0.86	0.9135	0.8825
0.4	0.2	0.8565	0.852	0.8825	0.8535	0.882	0.928	0.91
0.4	0.3	0.821	0.817	0.839	0.833	0.831	0.897	0.8795
0.4	0.4	0.79	0.7925	0.8345	0.845	0.8425	0.8825	0.8705
0.4	0.5	0.763	0.769	0.8095	0.806	0.809	0.8535	0.8465
0.5	0	0.8505	0.8385	0.8485	0.848	0.851	0.9055	0.8795
0.5	0.1	0.846	0.829	0.8365	0.838	0.836	0.8975	0.8755
0.5	0.2	0.842	0.8295	0.8595	0.863	0.8625	0.908	0.885
0.5	0.3	0.802	0.791	0.8255	0.8285	0.834	0.8715	0.871
0.5	0.4	0.765	0.7575	0.809	0.799	0.8025	0.8455	0.8495
0.5	0.5	0.7345	0.7365	0.73	0.726	0.73	0.7805	0.7815



表 3.14: 基于模型2 在正态分布下统计量的经验功效

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.849	0.8495	0.814	0.815	0.8165	0.8915	0.837
0	0.1	0.8425	0.8515	0.84	0.839	0.8395	0.9045	0.855
0	0.2	0.7885	0.826	0.83	0.829	0.8295	0.9055	0.8475
0	0.3	0.6975	0.8005	0.8275	0.8275	0.8275	0.9005	0.8395
0	0.4	0.5195	0.766	0.7755	0.776	0.778	0.8615	0.8045
0	0.5	0.3005	0.6675	0.61	0.6145	0.6185	0.72	0.6505
0.1	0	0.851	0.842	0.8225	0.8265	0.8285	0.893	0.841
0.1	0.1	0.848	0.8485	0.8575	0.8555	0.856	0.9195	0.879
0.1	0.2	0.7895	0.825	0.83	0.8325	0.833	0.906	0.855
0.1	0.3	0.681	0.7665	0.7765	0.7795	0.7795	0.866	0.805
0.1	0.4	0.5295	0.7365	0.7055	0.7145	0.7235	0.7965	0.737
0.1	0.5	0.2875	0.643	0.5915	0.6	0.6025	0.7095	0.635
0.2	0	0.8695	0.835	0.8375	0.841	0.844	0.905	0.855
0.2	0.1	0.847	0.814	0.7985	0.798	0.7995	0.887	0.842
0.2	0.2	0.796	0.795	0.7795	0.782	0.785	0.874	0.8095
0.2	0.3	0.679	0.739	0.727	0.731	0.7375	0.8185	0.768
0.2	0.4	0.519	0.693	0.734	0.7415	0.7445	0.822	0.763
0.2	0.5	0.3185	0.6045	0.543	0.548	0.5535	0.649	0.588
0.3	0	0.8315	0.7535	0.7325	0.7385	0.74	0.8335	0.771
0.3	0.1	0.8415	0.7605	0.7815	0.7865	0.795	0.872	0.8115
0.3	0.2	0.7875	0.7395	0.7295	0.731	0.7315	0.8425	0.762
0.3	0.3	0.6985	0.694	0.731	0.732	0.742	0.8025	0.7705
0.3	0.4	0.5375	0.629	0.643	0.649	0.6565	0.737	0.687
0.3	0.5	0.3375	0.553	0.544	0.5465	0.5485	0.6445	0.585
0.4	0	0.86	0.703	0.7125	0.7145	0.7145	0.8005	0.7425
0.4	0.1	0.8295	0.6865	0.6955	0.7	0.698	0.796	0.7365
0.4	0.2	0.7935	0.676	0.6945	0.698	0.703	0.786	0.7345
0.4	0.3	0.7035	0.616	0.604	0.6095	0.6145	0.706	0.6575
0.4	0.4	0.5515	0.5525	0.5115	0.5165	0.524	0.619	0.5685
0.4	0.5	0.364	0.4915	0.401	0.4065	0.4155	0.522	0.4655
0.5	0	0.8415	0.5965	0.573	0.5745	0.579	0.681	0.62
0.5	0.1	0.814	0.559	0.545	0.5455	0.547	0.6525	0.588
0.5	0.2	0.7855	0.5505	0.546	0.556	0.565	0.6445	0.597
0.5	0.3	0.6965	0.5095	0.5	0.5115	0.519	0.608	0.556
0.5	0.4	0.573	0.465	0.4625	0.464	0.473	0.5793	0.5185
0.5	0.5	0.412	0.4085	0.4015	0.4125	0.422	0.507	0.4605



表 3.15: 基于模型2 在拉普拉斯分布下统计量的经验功效

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.8645	0.869	0.889	0.8935	0.8955	0.947	0.9125
0	0.1	0.8475	0.86	0.888	0.8915	0.8895	0.9475	0.9065
0	0.2	0.796	0.8405	0.842	0.8435	0.848	0.9185	0.8685
0	0.3	0.683	0.799	0.787	0.7845	0.787	0.878	0.8205
0	0.4	0.535	0.758	0.783	0.7835	0.788	0.8855	0.828
0	0.5	0.295	0.661	0.657	0.6625	0.656	0.7885	0.719
0.1	0	0.8585	0.8525	0.843	0.841	0.8425	0.9285	0.8735
0.1	0.1	0.867	0.872	0.8655	0.869	0.872	0.94	0.8915
0.1	0.2	0.7985	0.8335	0.869	0.8705	0.8715	0.934	0.894
0.1	0.3	0.6955	0.7935	0.7875	0.789	0.785	0.875	0.8235
0.1	0.4	0.5235	0.7475	0.775	0.78	0.78	0.867	0.814
0.1	0.5	0.3045	0.661	0.6545	0.6585	0.6665	0.775	0.7115
0.2	0	0.8665	0.8285	0.8445	0.8475	0.849	0.9255	0.8805
0.2	0.1	0.842	0.8235	0.8415	0.846	0.85	0.9225	0.872
0.2	0.2	0.811	0.8165	0.8395	0.8405	0.844	0.906	0.8735
0.2	0.3	0.688	0.75	0.7575	0.7545	0.755	0.8545	0.8075
0.2	0.4	0.5175	0.6905	0.676	0.675	0.6765	0.78	0.723
0.2	0.5	0.315	0.63	0.6305	0.6435	0.653	0.743	0.685
0.3	0	0.8695	0.7885	0.8385	0.839	0.8445	0.915	0.8745
0.3	0.1	0.845	0.7815	0.779	0.7775	0.779	0.8765	0.83
0.3	0.2	0.811	0.7615	0.791	0.7945	0.7975	0.88	0.842
0.3	0.3	0.6835	0.6865	0.6625	0.6675	0.6725	0.778	0.7175
0.3	0.4	0.5545	0.6495	0.673	0.6745	0.6825	0.775	0.7295
0.3	0.5	0.324	0.5545	0.529	0.531	0.538	0.6485	0.6005
0.4	0	0.8555	0.735	0.737	0.7435	0.742	0.833	0.7855
0.4	0.1	0.834	0.6855	0.663	0.6695	0.674	0.7815	0.727
0.4	0.2	0.7875	0.6485	0.659	0.661	0.6645	0.759	0.715
0.4	0.3	0.704	0.6225	0.661	0.667	0.6665	0.765	0.726
0.4	0.4	0.5715	0.575	0.5755	0.5825	0.5915	0.6985	0.6535
0.4	0.5	0.3755	0.506	0.5095	0.5165	0.5245	0.6315	0.591
0.5	0	0.836	0.58	0.5995	0.603	0.6035	0.7145	0.666
0.5	0.1	0.8185	0.568	0.5825	0.5885	0.599	0.7055	0.662
0.5	0.2	0.7785	0.5575	0.5705	0.5725	0.579	0.689	0.6435
0.5	0.3	0.703	0.522	0.4865	0.4955	0.5005	0.607	0.572
0.5	0.4	0.5805	0.4795	0.461	0.471	0.484	0.582	0.5455
0.5	0.5	0.4195	0.424	0.403	0.4125	0.4205	0.5135	0.4735



表 3.16: 基于模型2 在混合分布下统计量的经验功效

$\rho_1$	$\rho_2$	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
0	0	0.876	0.876	0.8625	0.8685	0.8715	0.932	0.8855
0	0.1	0.8795	0.8885	0.8855	0.8875	0.8855	0.9495	0.912
0	0.2	0.843	0.8785	0.89	0.8915	0.889	0.9575	0.9185
0	0.3	0.7325	0.8255	0.8395	0.8415	0.843	0.915	0.861
0	0.4	0.568	0.7805	0.798	0.8035	0.813	0.892	0.8315
0	0.5	0.326	0.7025	0.723	0.72	0.72	0.8285	0.759
0.1	0	0.8865	0.8785	0.8755	0.877	0.874	0.938	0.901
0.1	0.1	0.8535	0.8585	0.8755	0.8785	0.8805	0.9405	0.9005
0.1	0.2	0.879	0.8575	0.8655	0.8695	0.8715	0.9335	0.894
0.1	0.3	0.7335	0.82	0.815	0.8155	0.8615	0.9055	0.851
0.1	0.4	0.5555	0.766	0.832	0.8315	0.833	0.9005	0.868
0.1	0.5	0.3317	0.69	0.686	0.694	0.6975	0.7985	0.737
0.2	0	0.892	0.8655	0.888	0.89	0.8905	0.949	0.9085
0.2	0.1	0.8675	0.8455	0.8575	0.852	0.854	0.925	0.8895
0.2	0.2	0.82	0.8195	0.835	0.834	0.839	0.9145	0.8715
0.2	0.3	0.726	0.7885	0.85	0.853	0.853	0.919	0.8865
0.2	0.4	0.549	0.7105	0.723	0.7255	0.7305	0.8265	0.7715
0.2	0.5	0.323	0.626	0.6575	0.656	0.6635	0.7705	0.703
0.3	0	0.8905	0.829	0.863	0.8635	0.8685	0.93	0.8915
0.3	0.1	0.8695	0.8105	0.8485	0.852	0.85	0.9175	0.886
0.3	0.2	0.8225	0.7875	0.7775	0.778	0.7765	0.867	0.823
0.3	0.3	0.743	0.7445	0.744	0.7415	0.744	0.838	0.7865
0.3	0.4	0.5675	0.6585	0.652	0.657	0.664	0.7695	0.7075
0.3	0.5	0.393	0.6115	0.599	0.606	0.614	0.703	0.6585
0.4	0	0.857	0.718	0.6275	0.6815	0.6865	0.795	0.741
0.4	0.1	0.861	0.7245	0.764	0.7635	0.769	0.8585	0.8145
0.4	0.2	0.8085	0.6785	0.7275	0.7295	0.7295	0.832	0.7825
0.4	0.3	0.7465	0.661	0.635	0.6385	0.644	0.751	0.699
0.4	0.4	0.583	0.585	0.615	0.6175	0.6245	0.7075	0.6625
0.4	0.5	0.3965	0.5265	0.5325	0.539	0.5515	0.652	0.5955
0.5	0	0.8515	0.616	0.604	0.6105	0.6115	0.711	0.6665
0.5	0.1	0.8345	0.6115	0.6105	0.6185	0.6235	0.7315	0.6785
0.5	0.2	0.8065	0.576	0.6045	0.608	0.6135	0.7135	0.6715
0.5	0.3	0.739	0.556	0.592	0.591	0.5995	0.7095	0.6595
0.5	0.4	0.602	0.49	0.5	0.502	0.5095	0.6285	0.5865
0.5	0.5	0.4445	0.4495	0.4485	0.454	0.4595	0.575	0.527



从表 3.5, 表 3.6, 表 3.7, 表 3.8, 表 3.9, 表 3.10 总体上可以看出, 在模型一中 BS 和 CQ 的经验水平均在 0.05 左右, 这是因为 BS 和 CQ 本质上是样本均值向量差的平方总和, 但在模型二下, BS 的经验水平的控制情况在两个协方差维数之间相关系数较大时明显比 CQ 的差, 这可能是因为这两个统计量的协方差估计有一定程度的区别; 而从这六个表可以看出 GCT10, GCT15, GCT20 这三个检验的经验水平都有所膨胀, 其水平都接近 0.07; SR 统计量的经验水平都比较保守, 接近 0.04; 统计量 SBASS 的经验水平在 0.05 左右. 从表 3.11, 表 3.12, 表 3.13, 表 3.14, 表 3.15, 表 3.16 总体上可以看出, BS 和 CQ 的经验功效基本相同, 而其他 5 项统计量的经验功效比较接近, 但在总体上其他 5 项统计量的功效比 BS, CQ 的要略高一些, 这是因为各维方差不同, 范围在 1 到 3 之间, 而 GCT, GCT15, GCT20, SR, SBASS 这 5 项采用了方差的信息; 但在模型二中, 因各维之间的相关系数较强时, BS 的经验功效有些突出, 时而高时而低, 这可能是其统计量的协方差估计不是很精确.

综合而言, 新提出的检验方法 SBASS 统计量具有很好性能, 因为与 BS, CQ, GCT10, GCT15, GCT20, SR 等统计量相比, 它的经验水平的控制良好, 功效的表现也相对很好. GCT10, GCT15, GCT20 统计量的经验水平控制都有些膨胀, 而 SR 统计量的经验水平控制都有些保守; 虽然 BS 和 CQ 统计量的经验水平控制较好, 但其功效都比 SBASS 统计量的经验功效较差, BS 统计量的经验功效在模型二下的表现更是不好.

## 3.2 实例及分析

R包“datamicroarray”提供了一组数据集脚本, 有关该数据集的详细信息, 请查看网址: <https://github.com/ramhiser/datamicroarray>. 本文在该包中下载了关于乳腺癌的数据集——Gravier et al.(2010)<sup>[29]</sup> 研究的数据集. Gravier et al.(2010)<sup>[29]</sup> 认为无腋窝淋巴结转移的小侵袭性导管癌 (T1T2N0) 可以预测小淋巴结阴性乳腺癌的转移. 他们使用比较基因组杂交技术对 168 名患者进行了为期 5 年的研究. 111 例诊断后没什么情况的患者为良性, 57 例早期转移患者为恶性. 2905 基因的表达水平已通过  $\log_2$  转换归一化.

首先, 检验 168 名“好”病人和“差”病人的所有基因变量是否存在差异. BS, CQ, GCT10, GCT15, GCT20, SR, SBASS 的  $P$  值分别为: 0, 0, 0,  $1.11e-14$ ,  $5.68e-12$ ,  $4.07e-05$ ,  $5.00e-05$ . 由这些  $P$  值可以看出这 7 项统计量的检验都能发现显著性差异(可能由于 BS, CQ, GCT10 等统计量的计算精度的问题, 无法具体精确出数值, 因而其  $P$  值表现为 0). 然后将 2905 个基因变量分成 6 个组, 前 5 组每组 500 个, 最后一组 405 个. 从表 3.17 可以看出, 检验统计量 SBASS 可以显著



的识别出六组之间的所有差异,而其它六个统计量不能显著识别出其中的一些差异,甚至是得出相反的结论(假设检验水平为0.05).例如:第4组,BS, CQ, GCT10, GCT15, GCT20, SR, SBASS 的  $P$  值分别为:  $3.29e-03$ ,  $3.14e-02$ ,  $2.36e-03$ ,  $7.29e-03$ ,  $1.31e-02$ ,  $1.10e-01$ ,  $1.45e-02$ . 所以检验统计量 SBASS 的检验是很敏锐的.

表 3.17: 7 项统计量每组500维的  $p$  值

组	个	BS	CQ	GCT10	GCT15	GCT20	SR	SBASS
1	500	1.97e-01	2.87e-01	3.53e-05	2.94e-04	8.89e-04	1.24e-01	1.13e-02
2	500	2.22e-16	8.44e-12	4.67e-07	1.83e-05	1.27e-04	2.49e-04	1.28e-02
3	500	0.00e-00	0.00e-00	2.00e-07	1.25e-05	1.08e-04	8.48e-12	3.89e-02
4	500	2.95e-04	5.37e-03	1.52e-03	3.77e-03	5.65e-03	1.83e-01	2.7e-03
5	500	3.29e-03	3.14e-02	2.36e-03	7.29e-03	1.31e-02	1.10e-01	1.45e-02
6	405	5.18e-03	1.73e-02	3.87e-05	4.38e-04	1.60e-03	8.24e-02	1.97e-02



## 第四章 总结与展望

在本文中,提出了一种新的检验方法来检验高维两样本均值向量差异比较密集的问题.该方法采用 stationary bootstrap 方法,得到维数较大时的 bootstrap  $P$  值.本文借助模拟结果,从下面几个角度去分析 SBASS 统计量的好处(举例见附表):

(1)变量间的相关性.随着变量间相关性的变大, GCT10 统计量的经验水平控制有些膨胀甚至控制不住, SR 过于保守,只有 Bai, CQ 和 SBASS 统计量的经验水平在正常范围内,但三者中的 SBASS 的功效优胜于其它两个统计量.

(2)样本量.随着样本量的变少, Bai 的经验水平越来越难以控制, SR 仍旧偏保守, GCT10 偏膨胀, SBASS 和 CQ 的经验水平都在 0.05 左右波动,但就功效而言, SBASS 的功效优于 CQ.

(3)维数.维数的增加使得 Bai 的经验水平越来越控制不住,而 SR 的经验水平偏保守,但 SBASS, CQ 和 GCT10 的经验水平在正常范围内波动,其中数 SBASS 统计量的经验水平控制的表现最好;而在功效上, SBASS 也明显地比 CQ, GCT10 的好一些.

附表:

$\rho_1$	$\rho_2$	$n$	$p$	总体	模型	size				
						Bai	CQ	GCT10	SR	SBASS
0	0.5	80	300	正态分布	模型1	0.042	0.051	0.063	0.0435	0.0525
0	0.4	80	300	正态分布	模型1	0.044	0.05	0.07	0.0739	0.0515
0	0.5	40	300	正态分布	模型1	0.2005	0.0555	0.062	0.0475	0.0535
0	0.5	80	700	正态分布	模型1	0.1265	0.047	0.052	0.0395	0.0515
0	0.5	80	300	拉普拉斯分布	模型1	0.0355	0.0475	0.064	0.0335	0.0505
0	0.5	80	300	正态分布	模型2	0.0755	0.037	0.054	0.0305	0.044
$\rho_1$	$\rho_2$	$n$	$p$	总体	模型	power				
						Bai	CQ	GCT10	SR	SBASS
0	0.5	80	300	正态分布	模型1	0.8465	0.864	0.8825	0.934	0.902
0	0.4	80	300	正态分布	模型1	0.8625	0.871	0.8775	0.932	0.8935
0	0.5	40	300	正态分布	模型1	0.7585	0.3515	0.3475	0.482	0.4405
0	0.5	80	700	正态分布	模型1	0.3485	0.8735	0.902	0.952	0.945
0	0.5	80	300	拉普拉斯分布	模型1	0.8345	0.858	0.8865	0.9375	0.9075
0	0.5	80	300	正态分布	模型2	0.3005	0.6675	0.61	0.72	0.6505

(4)样本总体假设.在本文中随机作了三个样本总体假设,从这个模拟结果可以得出: SR, Bai 和 CQ 等统计量经验水平的控制波动较大,因此受样本总体假设影响比较大,而 SBASS 和 GCT10 统计量的经验水平从数据上看相对较为稳定,所以基本上不受样本总体假设的影响,但 GCT10 的经验水平基本上膨胀,而且其功效相





对 SBASS 统计量较差. 因此基本上可以确定 SBASS 统计量不受样本总体假设的影响, 且在作检验时能够得出较为准确的判定.

(5)协方差结构. 为得出 SBASS 统计量的检验在一般协方差结构下也能有效的结论, 本文采用两种协方差结构. 从模拟结果可以得出, 在第二种协方差结构下 Bai 和 SR 的经验水平已经出现控制不住的情形, 而虽 CQ, GCT10 和 SBASS 统计量的经验水平都得到控制, 但从总体上看, GCT10 统计量的经验水平的控制还是偏膨胀, 相比较而言, SBASS 的经验水平控制情况最为理想, 基本都在 0.05 左右, 其次 SBASS 统计量的功效都要明显地比其它两个优胜.

最后从大量模拟结果可以看出, 由于高维数据各列的方差不同, 现有方法中基于学生化列统计量平方和(SR, GCT10, GCT15 和 GCT20)比基于非学生化列统计量平方和(Bai 和 CQ)的功效更好; 但 SR 的经验水平基本上都有一定的保守性, 而 GCTs 的经验水平有时会有些膨胀甚至失去控制. 而我本文中提出的基于 stationary bootstrap 方法的统计量 SBASS 无论是在何种情况下, 其第一错误的控制水平基本上都在合理范围内, 且其经验功效也有良好的表现. 另外从理论上, 现有的基于学生化和非学生化列统计量平方和方法的一个难点是需要估计一个非常复杂的协方差, 而协方差的估计在很大程度上影响了整体的经验水平和功效. 本文中提出的 stationary bootstrap 方法避免了对协方差的估计, 并且可以在计算机的帮助下完成一些可行的计算. 在大量的模拟中, 该方法具有正常的经验水平和理想的功效. 在实例中也可以看出本文提出的检验 SBASS 的优点. 综合而言, 基于 stationary bootstrap 方法的统计量 SBASS 在作高维数据均值向量检验时拥有良好的性能, 因此在针对高维假设检验问题, 即维数  $p$  很大时, 笔者建议采用本文提出统计量 SBASS 进行检验.

当然该方法也是有一定的局限性: 第一要求数据是严平稳弱相关的时间序列, 而现实中有些数据并不满足该条件; 第二计算时间相对比较长, 因为它需要多次抽样, 多次计算重抽样的统计量来减少抽样带来的误差.



## 参考文献

- [1] Hotelling H. The generalization of Student' s ratio[M]. Breakthroughs in s-tatistics. Springer, New York, NY, 1992: 54-65.
- [2] Donoho D L. High-dimensional data analysis: The curses and blessings of di-mensionality[J]. AMS math challenges lecture, 2000, 1(2000): 32.
- [3] Fan J, Li R. Statistical challenges with high dimensionality: Feature selection in knowledge discovery[J]. arXiv preprint math/0602133, 2006.
- [4] Dempster, A. P. A high dimenstional two sample significance test[J]. The An-nals of Mathematical Statistics, 1958, 29,995-1010.
- [5] Dempster, A. P. A significance test for the separation of two highly multivariate small samples[J]. Biometrics, 1960, 16,41-50.
- [6] Bai Z, Saranadasa H. Effect of high dimension: by an example of a two sample problem[J]. Statistica Sinica, 1996: 311-329.
- [7] Chen S X, Qin Y L. A two-sample test for high-dimensional data with applica-tions to gene-set testing[J]. The Annals of Statistics, 2010, 38(2): 808-835.
- [8] Srivastava M S, Katayama S, Kano Y. A two sample test in high dimensional data[J]. Journal of Multivariate Analysis, 2013, 114: 349-358.
- [9] Gregory K B, Carroll R J, Baladandayuthapani V, et al. A two-sample test for equality of means in high dimension[J]. Journal of the American Statistical Association, 2015, 110(510): 837-849.
- [10] Cai T, Liu W, Xia Y. Two-sample covariance matrix testing and support re-covery in high-dimensional and sparse settings[J]. Journal of the American S-tatistical Association, 2013, 108(501): 265-277.
- [11] Cai T T, Liu W, Xia Y. Two-sample test of high dimensional means under dependence[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2014, 76(2): 349-372.



- [12] Chang J, Zheng C, Zhou W X, et al. Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity[J]. *Biometrics*, 2017, 73(4): 1300-1310.
- [13] Dimitris N. Politis, Joseph P. Romano. The stationary bootstrap[J]. *Journal of the American Statistical Association*, 1994, 89(428):1303-1313.
- [14] Brockwell P J, Davis R A, Fienberg S E. Time series: theory and methods: theory and methods[M]. Springer Science and Business Media, 1991.
- [15] Dimitris N, Politis, et al. Bias-corrected nonparametric spectral estimation[J]. *Journal of Time Series Analysis*, 1995, 16(1):67-103.
- [16] White H, Politis D N. Automatic block-length selection for the dependent bootstrap[J]. *Econometric Reviews*, 2004, 23(1):53-70.
- [17] Politis D N. Adaptive bandwidth choice[J]. *Journal of Nonparametric Statistics*, 2001,15(4):517-533
- [18] Bilodeau M, Brenner D. Theory of multivariate statistics[M]. Springer, 1999.
- [19] Sazonov V V. On the multi-dimensional central limit theorem[J]. *Sankhyā: The Indian Journal of Statistics, Series A*, 1968: 181-204.
- [20] Bhattacharya R N. On errors of normal approximation[J]. *The Annals of Probability*, 1975: 815-828.
- [21] Nagaev S V. An estimate of the remainder term in the multidimensional central limit theorem[C]. *Proceedings of the Third Japan—USSR Symposium on Probability Theory*. Springer, Berlin, Heidelberg, 1976: 419-438.
- [22] Sweeting T J. Speeds of convergence for the multidimensional central limit theorem[J]. *The Annals of Probability*, 1977, 5(1): 28-41.
- [23] Senatov V V. Several estimates of the rate of convergence in the multidimensional CLT[J]. *Doklady Akademii nauk Soiuza Sovetskikh Sotsialisticheskikh Respublik*, 1980, 254,809-812.
- [24] Sazonov V V. Normal approximation: some recent advances[M]. Berlin: Springer, 1981.



- [25] Bentkus V. Dependence of the Berry-Esseen estimate on the dimension[J]. Lithuanian Mathematical Journal, 1986, 26(2): 110-114.
- [26] Bentkus V. On the dependence of the Berry - Esseen bound on dimension[J]. Journal of Statistical Planning and Inference, 2003, 113(2): 385-402.
- [27] Chernozhukov V, Chetverikov D, Kato K. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors[J]. The Annals of Statistics, 2013, 41(6): 2786-2819.
- [28] Chernozhukov V, Chetverikov D, Kato K. Central limit theorems and bootstrap in high dimensions[J]. The Annals of Probability, 2017, 45(4): 2309-2352.
- [29] Gravier E, Pierron G, Vincent-Salomon A, et al. A prognostic DNA signature for T1T2 node-negative breast cancer patients[J]. Genes, chromosomes and cancer, 2010, 49(12): 1125-1134.



## 附 录

1, 有关模拟的 r code:

```
rm(list=ls())
#install.packages("highD2pop")
library(highD2pop)
#install.packages('MASS')
library(MASS)
#install.packages("forecast")
library(forecast)
#install.packages("np")
library(np)
#####sampling SBboot sample
SBoot = function(x,b){
X = rep(x,times=2)
n = length(x)
J = rgeom(n,1/b)+1
tt=(which(cumsum(J)>=n))[1]
I = sample(1:n,tt,replace=TRUE)
Block = c()
for(j in 1:tt)
{
Block = c(Block,I[j):(I[j]+J[j]-1))
}
N = Block[1:n]
return(X[N])
}
#####sampling SBboot sample

#####以下是定义BS
BS=function(X,Y)
{
m = nrow(X)
n = nrow(Y)
```



```
p = ncol(X)
N = n+m-2
tau1 = (n+m)/(n*m)

Xbar = colMeans(X)
Ybar = colMeans(Y)
Sn = ((m-1)*cov(X)+(n-1)*cov(Y))/N
Bn = sqrt(N^2/((N+2)*(N-1))*( sum(Sn^2)-1/N*( sum(diag(Sn)) )^2 ))
Z = (1/tau1*sum((Xbar-Ybar)^2)-sum(diag(Sn)) )/(sqrt(2*(N+1)/N)*Bn)
pvaluesbs1 = 2*(1-pnorm(abs(Z)))

results = pvaluesbs1
return(results)
}
#####以上是定义BS

#####以下是定义SR
SR=function(X,Y)
{
m = nrow(X)
n = nrow(Y)
p = ncol(X)
xmean = colMeans(X)
ymean = colMeans(Y)
Sx = cov(X)
Sy = cov(Y)
D_hat = diag(diag(Sx)/m + diag(Sy)/n)
R = solve(sqrt(D_hat)) %*% (Sx/m+Sy/n) %*% solve(sqrt(D_hat))
Cpn = 1 + sum(R^2)/sqrt(p^3)

Var_hat=2*sum(R^2)/p-2*(sum(diag(solve(D_hat)%*% Sx)))^2/(p*(m-1)*m^2)
- 2*(sum(diag(solve(D_hat)%*% Sy)))^2/(p*(n-1)*n^2)

Ts =(t(xmean-ymean)%*%solve(D_hat)%*%(xmean-ymean)-p)
```



```
1/sqrt(p*abs(Var_hat)*Cpn)
```

```
pvMs = 2*(1-pnorm(abs(Ts)))  
results = pvMs  
return(results)  
}
```

```
#####以上是定义SR
```

```
#####以下是定义SBASS
```

```
SBASS=function(X,Y)  
{  
m = nrow(X)  
n = nrow(Y)  
p = ncol(X)  
xmean = colMeans(X)  
ymean = colMeans(Y)  
X0 = (xmean-ymean)^2 /diag(cov(X) /m + cov(Y)/n)  
T0 = p*(mean(X0) - 1)^2/var(X0)  
b = b.star(X0,round = TRUE)[1]  
#####  
stf = function(y)  
{  
    ( p*(mean(y) - mean(X0))^2/var(y) )  
}  
#####  
az = c()  
for(i in 1:3000)  
{  
boot = SBoot(X0,b)  
T = STF(boot)  
az=c(az,T)  
}  
pvMSs = mean( az > T0 )  
results = pvMSs
```



```
return(results)
}
#####以上是定义SBASS

#####
covMed=function(rho,p)
{
  M=matrix(0,p,p)
  for(i in 1:p)
  {
    for(j in 1:p)
      {M[i,j]=rho^{abs(i-j)}}
  }
  return(M)
}

#####
#####
rmultLapLace <- function (n, mu, vmat, tol = 1e-10)
{
  p <- ncol(vmat)
  if (length(mu) != p)
    stop(paste("mu vector is the wrong length:", length(mu)))
  vs <- La.svd(vmat)
  vsqrt <- t(t(vs$vt) %*% (t(vs$u) * sqrt(vs$d)))
  ## The beginning of the modified code
  x <- runif(n*p,min=-0.5,max=0.5); x[x>0] <- 1; x[x<0] <- -1;
  y <- (1/sqrt(2))*x*rexp(n*p);
  ans <- matrix(y, nrow = n) %*% vsqrt
  ## The end of the modified code
  ans <- sweep(ans, 2, mu, "+")
  dimnames(ans) <- list(NULL, dimnames(vmat)[[2]])
  ans
}
#####
```





```
S = 2000
n1 = 60
n2 = 80
p = 300
alpha = 0.05
DD = diag(sqrt(runif(p,1,3)))
for (rho1 in seq(0,0.5,by=0.1))
{
names(rho1) = "rho1"
print(rho1)
VV1 = DD %*% solve(covMed(rho1,p)) %*% DD
for (rho2 in seq(0,0.5,by=0.1))
{
names(rho2) = "rho2"
print(rho2)
VV2 = DD %*% solve(covMed(rho2,p)) %*% DD
#n0s = round(p^(0.8),0)
#n0d = sample(1:p,n0s,replace=F)
mu1 = rep(0,p)
mu2 = rep(0,p)
#mu2[n0d[1:round(n0s/2,0)]] = 0.23
#mu2[n0d[(1+round(n0s/2,0)):n0s]] = -0.23
pvs = c()
print(date())
for(i in 1:S)
{
if(i %% 100 == 0) print(i)
#print(i)
#X = rmultLapLace(n1, mu=mu1, vmat=VV1, tol = 1e-10)
#Y = rmultLapLace(n2, mu=mu2, vmat=VV2, tol = 1e-10)
X = mvrnorm(n1, mu1, VV1)
Y = mvrnorm(n2, mu2, VV2)

T1 = BS(X,Y)
```



```
T2 = ChenQin.test(X=X, Y=Y)
T3 = GCT.test(X=X, Y=Y,r=10,smoother="parzen",ntoorderminus=1)
T4 = GCT.test(X=X, Y=Y,r=15,smoother="parzen",ntoorderminus=1)
T5 = GCT.test(X=X, Y=Y,r=20,smoother="parzen",ntoorderminus=1)
T6 = SR(X,Y)
T7 = SBASS(X,Y)

pv = c(T1,T2[[2]],T3[[2]],T4[[2]],T5[[2]],T6,T7)
pvs = rbind(pvs,pv)
}
sizes = colMeans(pvs<=alpha)
para = c(n1, n2, p, rho1, rho2, S, sizes)

names(para) = c('m', 'n', 'p', 'rho1', 'rho2', 'S', 'Bai',
               'CQ', 'GCT10', 'GCT15', 'GCT20', 'SR', 'SBASS')

print(para)

write.table(t(as.matrix(para)),file=" model 2 of mvnorm sizes.txt",
            sep = '\t',col.names=TRUE,row.names=FALSE,
            quote = FALSE, append=TRUE)

}
}
```



## 致 谢

在华中师范大学的三年研究生求学生涯即将结束之际,我想对我的老师,家人和朋友表示感谢,感谢他们在这一路上陪伴着我鼓舞着我。

首先感谢我的指导老师左国新老师和李正帮老师,本文从头到尾是在左国新老师和李正帮老师的指导和帮助下完成的。在这三年间这两位老师经常在学习和生活上给予我热诚的帮助。左国新老师是一位学识渊博,教学态度严谨,对学生要求严格的教师,在他的身上我学习到了对待学术必须要做到一丝不苟的品质;李正帮老师是一位博览群书,诲人不倦,待人真诚的教师,在他身上我学习到了待人真诚厚道的品德。无论何时有问题请教这两位老师,他们都会及时给予我指导性的意见和建议,为我写论文和修改论文指出明确又正确的方向,所以本文的完成离不开这两位老师的谆谆教导,很感谢在华师能遇到两位亦师亦友的老师。

其次感谢数统学院所有任教的教师,是他们的辛勤付出让我学习和掌握到了统计学的专业知识。还有感谢数统学院行政部门的老师,尤其是宋红文老师和李东老师,他们在我的研究生生活中给予了我不可磨灭的帮助。

然后感谢父母在物质上默默地给予我支持,感谢奶奶每次在我回家时都留一些她喜欢的吃食给我吃,我很幸运出生在一个和谐又温馨的家庭。

接下来要感谢我的女朋友邱小姐,感谢对我缺点的包容,并给予我精神上的支持和肯定,她的支持和肯定无疑是在我读研期间自信心的来源,感谢她出现在我的生命里,让我成长为一个真正有担当的男子汉。另外还要感谢身边所有关心和帮助过我的朋友,他们的关心和帮助让我能够在跌倒时快速的爬起来。

最后,感谢参与评审本篇硕士毕业论文的老师,祝所有的人民教师身体健康,阖家幸福。

曾奎杰

2020年4月