索取号：O212.1/12.353:130902074 密级：公 开

# 南京师范大学

## 硕士学位论文

# 高维线性回归模型下的经验似然方法

# Empirical Likelihood Methods for High Dimensional Linear Regression Models

研 究 生：曾力立

指导教师：周秀轻 教授

培养单位：数学科学学院

一级学科：统计学

二级学科：统计学

完成时间：2016年3月7日

答辩时间：2016年5月24日

# 学位论文独创性声明

本人郑重声明：所提交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果。本论文中除引文外，所有实验、数据和有关材料均是真实的。本论文中除引文和致谢的内容外，不包含其他人或其它机构已经发表或撰写过的研究成果。其他同志对本研究所做的贡献均已在论文中作了声明并表示了谢意。

学位论文作者签名：[illegible]　　日　　期：2016年5月27日

# 学位论文使用授权声明

研究生在校攻读学位期间论文工作的知识产权单位属南京师范大学。学校有权保存本学位论文的电子和纸质文档，可以借阅或上网公布本学位论文的部分或全部内容，可以采用影印、复印等手段保存、汇编本学位论文。学校可以向国家有关机关或机构送交论文的电子和纸质文档，允许论文被查阅和借阅。（保密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论文，**密级**：公开　**保密期限**为_____年。

学位论文作者签名：[illegible]　　指导教师签名：[illegible]

日　　期：2016年5月27日　　期：2016年5月27日

# 目 录

# Contents

# 摘 要

高维数据在许多领域正受到越来越广泛的关注, 是目前统计学领域和应用研究中面临困难最多、挑战最严峻, 同时也是最有可能取得突破的研究领域之一. 在对高维数据进行分析时, 运用传统的多元统计分析方法处理实际数据时通常会碰到数据不符合正态分布或数据本身没有多少先验信息的情况. 所以, 处理时通常考虑非参数的方法. 而基于传统的经验似然所形成的方法现已被证明有一定效果. 这是因为经验似然既拥有参数方法的有效性又具有非参数方法的稳定性. 它在很多情形下比正态逼近方法精确, 特别是当数据来自非正态总体或方差估计不稳定时. 此外, 它有类似于 bootstrap 的抽样特性, 且相比之下更有其自身的优越性, 如域保持性、变换不变性以及 Bartlett 纠偏性. 跟 bootstrap 相比, 它的另一个好处是计算量小, 并且在一些给定的正则条件下, Wilks 定理成立.

在这篇论文里, 我们的主要目的是检验一个可能的高维线性回归模型的系数是否等于一个给定值. 通过把传统经验似然方法里面的高维约束条件巧妙地变换成与维数无关的低维情形, 我们首先提出了一个新奇且简单的经验似然检验方法. 这个方法保留了传统经验似然方法里的最优性准则并且拥有其他好的特性. 在计算偏经验似然函数的时候, 涉及到在约束条件下求极大值, 这也正是运用经验似然方法的一个关键步骤. 为了确保犯第一类错误的概率更加接近于给定显著性水平, 我们进一步在先前提出的方法里加入了伪观测值, 从而得到更优的检验统计量. 随后, 我们在几个不同的例子里做了相关模拟以评估我们所提出的方法的检验效果. 模拟结果显示我们所提出的检验方法都要优于前人所提出的方法. 而由于例子中回归系数取值的特殊性, 我们之后又提出了另一种可以用于检验更一般的系数的方法并给出相关定理及模拟结果.

本文的主要创新点在于:

(1)我们尝试着把传统的经验似然方法里与 $p$ 有关的高维约束条件巧妙地变换成低维情形, 以此构造出新的约束条件, 再利用经验似然的方法解决相关

假设检验问题.

(2)我们在前人的方法及我们所提出的简单经验似然方法里加入了伪观测值, 从而作出了一个新奇的调整. 调整后的经验似然方法保留了之前方法的所有最优性准则. 不仅如此, 该方法下的区间覆盖率更接近于置信水平, 而且还不需要 Bartlett 校正和 bootstrap 方法里那么复杂的程序.

(3)对于更一般地回归系数, 我们考虑了将原限制条件中的分量加权相加, 并且这个加权也是随机的, 从而解决了线性模型中回归系数更为一般情况下的假设检验问题.

(4)我们针对不同的维数, 有区别地加入了约束条件的个数, 一方面使得犯两类错误的概率令人满意, 另一方面也大大地节省了计算成本.

**关键词**: 高维数据; 经验似然; 线性模型; 假设检验; 伪观测值

# Abstract

High dimensional data is becoming more and more prevalent in many areas. The study of high dimensional data is now faced with many difficulties and challenges, but it is also the area where we can make progress. Multivariate statistical analysis methods in solving high dimensional data may come across such situations, for example, the data itself is not from normal distribution or it does not own enough prior information. Thus, nonparametric approaches are considered in general. Besides, the method based on traditional empirical likelihood (EL) has been shown as a powerful one, because it owns both the effectiveness of parametric likelihood methods and the reliability of nonparametric approaches. It is also more precise than normal approximation under many cases, especially when the data does not come from normal distribution or the estimate of variance is not stable; it also owns sampling properties like the bootstrap method, and its calculation method is simple. What's more, the nonparametric version of Wilks' theorem holds true under some regularity conditions.

In this paper, our focus is to test whether the regression coefficients in a possibly high dimensional linear model are equivalent to given values. By transforming the high dimensional estimating equation used in traditional EL approach into low dimensional case tactfully, we propose our novel but simple method which not only retains the optimal properties in traditional EL method but also owns other exciting results. We need to compute the maximum value limited by estimating equations when solving profile EL function, which is the key step of EL method. To ensure type I errors to be closer to the given nominal levels, we add a pseudo-observation in our method further to get relative test statistics. Simulations have also been conducted to assess the performance of the proposed methods in several different models. Moreover, because of the particularity of the regression coefficients, we come up with another EL method which is suitable for this situation. Simulation results are also presented.

Our main innovation points are as follows:

(1)We try to transform the high dimensional estimating equation used in traditional EL approach into low dimensional case tactfully so that new estimating equation which is not related to $p$ can be constructed, and then relative test problems could be solved by this EL method.

(2)Pseudo-observations have been added into both our method and the method proposed by predecessors so that a novel adjustment is gained. The adjusted method reserves all the properties. Moreover, the coverage ratios are closer to confidence levels, and its calculation procedure is not tedious.

(3)For the general regression coefficients, we weight the components of the original estimating equation, and then add them. Thus, hypothesis testing under this case is solved.

(4)Different numbers of estimating equations are used in different dimensions, which makes us satisfied in terms of both type I errors and type II errors, and the computational expense is saved to a large extent at the same time.

**Keywords:** High dimensional data; Empirical likelihood; Linear model; Hypothesis test; Pseudo-observations.

# 第 1 章　绪论

本章分为四节, 首先介绍经验似然的历史背景及其发展和相关的研究成果, 然后介绍调整的经验似然, 其次引入高维线性回归模型, 最后说明本文研究的主要内容和结构安排.

## 1.1 经验似然的研究背景

经验似然是一种非参的推断方法, 它基于以数据为导向的似然比函数, 同时具有参数方法和非参数方法的优点: 即既可以得到置信域, 且所构造的置信域的形状由数据自行决定, 又不需要知道数据的具体分布族; 通过构造的限制条件或是先验分布, 它又能直接地体现出边缘信息; 另外, 它还可以被应用于有偏样本和删失数据, 并且它具有非常好的渐近准则. 概括而言, 经验似然可以理解为是一种不需要再抽样的 bootstrap 方法, 也可以理解为是一种不需要作参数假设的似然方法.

经验似然的思想最早可追溯到文献 [1], 此文中作者利用经验似然的思想建立了在随机删失数据情形下生存概率的区间估计. 文献 [2] 和文献 [3] 系统地研究了经验似然方法, 作者在独立同分布的完全样本数据情形下提出了该方法来对未知参数进行非参数的统计推断. 文献 [4] 定义了经验似然比函数并以此构造参数向量的置信域, 研究表明, 当样本量小的时候, 该方法构造的置信域要优于基于渐近正态性所构造的置信域.

从那以后, 大量基于这一新奇概念的文章应运而生, 比如文献 [5]、文献 [6] 等. 较之其它一些经典的统计方法, 这一方法既拥有参数方法的有效性又具有非参数方法的稳定性, 例如, 在很多情形下比正态逼近方法精确, 特别是当数据来自非正态总体或方差估计不稳定时. 此外, 它有类似于 bootstrap 的抽样特性, 且相比之下更有其自身的优越性, 如域保持性, 变换不变性以及 Bartlett 纠偏性. 跟 bootstrap 比, 经验似然的另一个好处是计算量小. 并且在一

些给定的正则条件下, Wilks 定理成立. 正因为有这些优点, 经验似然方法备受青睐, 一整套完备的理论应运而生. 我们用完全数据下的参数回归作为例子详细说明.

在参数回归中, 考虑模型

$$Y_i = g(X_i;\beta) + \varepsilon_i, i = 1, \cdots, n.$$

其中 $\beta \in R^p$ 是已知回归函数 $g(x;\beta)$ 中的 $p$ $(p < n)$ 维未知参数, 误差项 $\varepsilon_i$ 是独立的随机变量, 满足 $E(\varepsilon_i|X_i) = 0$, 且 $Var(\varepsilon_i|X_i) = \sigma^2(X_i) < \infty$, 这里 $\sigma^2(X_i)$ 表明误差项可以是异方差的. 通过极小化

$$S_n(\beta) = \sum_{i=1}^{n}\{Y_i - g(X_i;\beta)\}^2$$

可以得到 $\beta$ 的最小二乘估计量. 进一步, 当 $g(x;\beta)$ 足够光滑时, 有 $\sum_{i=1}^{n}\frac{\partial g(X_i;\beta)}{\partial\beta}\{Y_i - g(X_i;\beta)\} = 0$ 成立. 假设 $\beta_0$ 是使得 $E[\frac{\partial g(X_i;\beta)}{\partial\beta}\{Y_i - g(X_i;\beta)\}|X_i] = 0$ 有唯一解的参数的真值. 令 $q_1, \cdots, q_n$ 为概率权重, 则根据 Owen 的经验似然思想, $\beta$ 的经验似然函数为

$$L_n(\beta) = \max\prod_{i=1}^{n} q_i, \tag{1.1.1}$$

其中约束条件为

$$\sum_{i=1}^{n} q_i = 1 \tag{1.1.2}$$

及

$$\sum_{i=1}^{n} q_i\frac{\partial g(X_i;\beta)}{\partial\beta}\{Y_i - g(X_i;\beta)\} = 0. \tag{1.1.3}$$

约束条件 (1.1.2) 暗示了 $q_i$ 是概率权重, (1.1.3) 则表明了参数的性质, 是经验似然定义里的一个核心环节. 通过拉格朗日乘子法可得

$$q_i = \frac{1}{n}\frac{1}{1 + \lambda^T\frac{\partial g(X_i;\beta)}{\partial\beta}\{Y_i - g(X_i;\beta)\}}, \tag{1.1.4}$$

其中由上述约束条件可知, $\lambda$ 需要满足

$$\sum_{i=1}^{n}\frac{\frac{\partial g(X_i;\beta)}{\partial\beta}\{Y_i-g(X_i;\beta)\}}{1+\lambda^T\frac{\partial g(X_i;\beta)}{\partial\beta}\{Y_i-g(X_i;\beta)\}}=0.$$

将 (1.1.4) 代入 (1.1.1) 得其经验似然函数为

$$L_n(\beta)=\prod_{i=1}^{n}\frac{1}{n}\frac{1}{1+\lambda^T\frac{\partial g(X_i;\beta)}{\partial\beta}\{Y_i-g(X_i;\beta)\}},$$

及其对数经验似然函数

$$l_n(\beta)=\log\{L_n(\beta)\}=-\sum_{i=1}^{n}\log\left\{1+\lambda^T\frac{\partial g(X_i;\beta)}{\partial\beta}\{Y_i-g(X_i;\beta)\}\right\}-n\log(n).$$

该最优化问题的求解可转化成相应对偶问题的求解, 见文献 [17]. 当所有的权重 $q_i$ 都等于 $n^{-1}$ 时, $i=1,\cdots,n$, 经验似然 $\prod_{i=1}^{n}q_i$ 达到最大值 $n^{-n}$. 因此, 如果存在某个 $\beta$, 不妨假定为 $\hat{\beta}$, 满足 $\sum_{i=1}^{n}\frac{\partial g(X_i;\beta)}{\partial\beta}\{Y_i-g(X_i;\beta)\}=0$, 那么经验似然函数在 $\hat{\beta}$ 处达到最大值 $L_n(\hat{\beta})=n^{-n}$.

遵循传统的标准参数似然比的定义方法, 上述模型的对数经验似然比被定义为

$$r_n(\beta)=-2\log\{L_n(\beta)/L_n(\hat{\beta})\}=2\sum_{i=1}^{n}\log\left\{1+\lambda^T\frac{\partial g(X_i;\beta)}{\partial\beta}\{Y_i-g(X_i;\beta)\}\right\}.$$

在给定正则条件下, Wilks 定理依然成立, 即

$$r_n(\beta_0)\xrightarrow{d}\chi_p^2, n\longrightarrow\infty.$$

由此可以构造 $\beta_0$ 的似然比置信域为

$$I_{1-\alpha}=\left\{\beta:r_n(\beta)\leq\chi_{p,1-\alpha}^2\right\},$$

其中 $\alpha$ 为给定显著性水平, $\chi_{p,1-\alpha}^2$ 是 $\chi_p^2$ 的 $1-\alpha$ 分位数.

这里的回归函数 $g(x;\beta)$ 实际上包括了很多特殊的情况. 如: 若 $g(x;\beta)=x^T\beta$, $T$ 表示矩阵的转置, 则上述模型变为线性回归模型; 若$g(x;\beta)=G(x^T\beta)$ 且 $\sigma^2(x)=\sigma_0^2V\{G(x^T\beta)\}$, 其中 $G$ 是一个已知的相关函数, $V(\cdot)$ 为已知的方差函数, $\sigma_0^2>0$ 为未知常数, 则上述模型变为广义线性模型, 见文献 [7].

经过二十多年的发展, 完全数据下的经验似然的研究已经比较成熟, 并取得了丰硕的成果. 例如, 文献 [8]将经验似然应用于分位数回归的研究; 文献 [3], 文献 [9] 及文献 [10] 将经验似然应用于线性回归模型; 文献 [11]和文献 [12] 将经验似然应用于分位数的估计; 文献 [13] 发展了厚尾分布均值的经验似然方法; 文献 [14] 文章中的部分线形回归模型应用了经验似然的相关理论知识; 文献 [10] 和文献 [15] 将经验似然应用于抽样调查问题的研究; 文献[16] 将经验似然应用于经济模型的研究等. 2001 年, 文献 [17] 对经验似然作了详细地论述, 其中包括对经验似然方法效用及灵活性的阐述, 也包括对光滑性问题、时间序列、有偏样本等的研究.

随着统计方法的进一步完善及 Miller 估计, Buckley James 估计, Kaplan-Meier 估计等估计方法研究的成熟, 经验似然方法也被应用于删失数据, 如文献 [18] 探讨了右删失数据下的经验似然回归分析; 文献 [19] 用经验似然的方法构造了一个基于 B-J 估计量的检验, 并通过渐近卡方分布计算了 P-值; 文献 [20] 对于含有删失变量的中位数回归模型提出了一个新的且简单的估计式, 然后运用经验似然比统计量来构造参数的置信域; 文献 [21] 将经验似然方法应用于带有右删失数据的广义变换模型等等. 经验似然方法正以其自身的诸多优点被越来越广泛地应用到诸多领域并且成果显著.

## 1.2 调整的经验似然

在计算偏经验似然函数的时候, 涉及到在约束条件下求极大值, 这也正是运用经验似然方法里的一个关键步骤. 然而有些时候会出现相应的数值问题无解的情况. 这个时候, 传统的做法是在偏经验似然里面设置一个零值. 但是这样有两个缺陷. 其一, 数值上很难确定是否无解; 其二, 在偏经验似然被设为零值点处的参数上面, 该方法提供不了任何信息. 解决这一问题, 就需要考虑其他的调整方法.

我们仍然考虑 1.1 节中的参数回归模型. 为了便于书写, 下面我们记 (1.1.3) 中 $\frac{\partial g(X_i;\beta)}{\partial \beta}\{Y_i - g(X_i;\beta)\}$ 为 $m_i, i = 1, \ldots, n$. 于是, 该模型下的经验似然比函数可被定义为

$$L^{EL}(\beta) = \sup\left\{\prod_{i=1}^{n}(nq_i) : q_i \geq 0, \sum_{i=1}^{n} q_i = 1, \sum_{i=1}^{n} q_i m_i = 0\right\}, i = 1, \ldots, n.$$

相应地, 其对数经验似然比函数为

$$l^{EL}(\beta) = \sup\left\{\sum_{i=1}^{n}\log(nq_i) : q_i \geq 0, \sum_{i=1}^{n} q_i = 1, \sum_{i=1}^{n} q_i m_i = 0\right\}, i = 1, \ldots, n.$$

如文献 [22] 所指出的, 为了解决与 $l^{EL}(\beta)$ 相关的数值问题, 一个先决条件是 $\{m_i, i = 1, 2, \cdots, n\}$ 所组成的凸包必须包含 $\mathbf{0}$ 作为其内点. 按照该文章里的说法, 对于一个给定的 $\beta$, 我们不妨称 $l^{EL}(\beta)$ 拥有好的定义当且仅当 $\mathbf{0}$ 是 $\{m_i, i = 1, 2, \cdots, n\}$ 所组成的凸包的内点. 同时假定 $\Omega$ 为使得 $l^{EL}(\beta)$ 拥有好的定义的 $\beta$ 的定义域. 实际情况中, $\Omega$ 可能是空的. 通常的做法是对于那些不在 $\Omega$ 里的 $\beta$, 定义$l^{EL}(\beta) = -\infty$. 但是这样数值上很难确定是否无解, 此外对于任意两个不在 $\Omega$ 里的 $\beta_1$ 和 $\beta_2$, $l^{EL}(\beta_1)$ 和 $l^{EL}(\beta_2)$ 的值对于它们的相对可靠性提供不了任何信息, 即使它们之一足够接近 $\Omega$ 的边界而另一个足够远离. 也正是这一缺陷对于找出 $l^{EL}(\beta)$ 的极大值点提出了很大的挑战. 具体见文献 [22].

为了解决所谓的好的定义的问题, 文献 [22] 对于传统的经验似然作了一个新奇的调整, 使得调整后的经验似然函数对于所有的参数都拥有好的定义. 这样我们就不需要确定 $\beta$ 是不是在 $\Omega$ 之中, 由此找到调整后的 $l^{EL}(\beta)$ 的极大值点也会变得简单很多. 文献 [22] 定义 $m_{n+1} = -\frac{a_n}{n}\sum_{i=1}^{n} m_i = -a_n\bar{m}_n$, 其中 $a_n$ 为给定的正的常数. 相应地, 调整后的对数经验似然比函数为

$$l^{AEL}(\beta) = \sup\left\{\sum_{i=1}^{n+1}\log[(n+1)q_i] : q_i \geq 0, \sum_{i=1}^{n+1} q_i = 1, \sum_{i=1}^{n+1} q_i m_i = 0\right\},$$

$i = 1, \ldots, n+1$. 这样, 对于每一个给定的 $\beta$, 凸包 $\{m_i, i = 1, \ldots, n, n+1\}$ 都包含 $\mathbf{0}$ 作为其内点, 即$l_n^{AEL}(\beta)$ 对于所有 $\beta$ 拥有好的定义, 且原有经验似然的性质

依然被保留. 事实上, 如果 $\bar{m}_n = 0$, 则 $l^{AEL}(\beta) = l^{EL}(\beta) = 0$. 对于使得 $\bar{m}_n \approx 0$ 的那些 $\beta$, 我们有 $l^{EL}(\beta) \approx l^{AEL}(\beta)$. 当 $\beta$ 不属于 $\Omega$ 时, 则有 $l^{EL}(\beta) = -\infty$, 而此时 $l^{AEL}(\beta)$ 依然拥有好的定义. 因此, 对于极大化 $l^{EL}(\beta)$, $l^{AEL}(\beta)$ 的信息是充分的.

不仅如此, 调整后的基于渐近卡方分布的置信区间也会变大. 也就是说, 当样本量小时, 覆盖率会变大. 并且在模拟过程中, 并不需要像 Bartlett 校正和 bootstrap 校验那样复杂的程序. 事实上, 调整的经验似然的算法比未调整的算法收敛得更快. 算法收敛性的证明可参见文献 [23].

## 1.3 高维数据下的经验似然

随着科技的日渐发展, 人们在实际应用过程中常常会碰到许多类型的海量数据, 如多媒体图形图像视频数据、证券市场交易数据、互联网门户数据、生物特征数据、航天航空采集数据等, 这些数据通常在统计中被称为高维数据.

在分析高维数据过程中面临的一大挑战是维数的膨胀, 也就是通常所说的"维数灾难". 有研究表明, 当维数越来越大时, 分析和处理多维数据的成本与复杂度成指数级增长. 在对高维数据进行分析时，所需的空间样本数会随维数的增加而呈指数增长. 运用传统的多元统计分析方法处理实际数据时通常会碰到数据不符合正态分布或数据本身没有多少先验信息的情况. 所以, 处理时通常会考虑用非参数的方法去解决. 处理这类问题的非参数方法主要依赖大样本理论, 但会出现与空间的维数相比样本量显得比较少的情况. 因此, 有时大样本理论处理高维数据不适用. 另外, 许多经典的低维数据处理方法, 在处理高维数据时面临着难以解决的困难. 例如, 传统的数据处理方法在处理高维数据时不能满足稳健性要求; 高维导致空间的样本数变少, 从而使得一些统计上的渐近性难以实现; 维数的增加亦会导致数据的计算量迅速上升等等. 上述问题给高维数据处理中的模式识别带来了很大的挑战. 同样, 这种维数的膨胀, 给数学和数据分析造成了一定的困难. 因此在高维背景下研究传统统计方法的表现并建立新的方法显得尤为重要和紧迫.

传统的经验似然方法已经被应用到一些高维数据当中并且相关的渐近准

则也被仔细研究过. 如文献 [24] 在多元模型下研究了均值的渐近性质; 文献 [25] 基于 plug-in 估计对经验似然方法做了一个推广研究; 文献 [26] 将经验似然应用于高维变量选择等等.

我们考虑如下的高维线性回归模型, 即 1.1 节中 $g(x;\beta)=x^T\beta$ 的情况.

$$Y_i=\beta^T X_i+\varepsilon_i, i=1,\ldots,n, \tag{1.3.1}$$

其中 $\{X_i, i=1,\ldots,n\}$ 是独立同分布 (i.i.d.) 的随机向量, $\beta=(\beta_1,\ldots,\beta_p)^T\in R^p, p\geq 1$, 是一列未知系数向量, $\{\varepsilon_i, i=1,\ldots,n\}$ 是均值为 0 , 方差为 $\sigma^2>0$ 的 i.i.d. 随机变量. 同时我们假设 $\{X_i\}$ 和 $\{\varepsilon_i\}$ 相互独立.

对于线性模型 (1.3.1), 实际中最感兴趣的参数是 $\beta$, 例如有关 $\beta$ 的置信域或假设检验通常是人们所关注的统计问题. 当误差项 $\{\varepsilon_i, i=1,\ldots,n\}$ 的分布完全未知时, 经典的方法是利用 $\beta$ 的最小二乘估计的渐近正态性来构造 $\beta$ 的置信域. 但这种方法有两个缺陷: 一是基于渐近正态性导出的置信域是对称的, 而对称的置信域不一定合适, 特别是当总体分布为非对称时; 二是要估计渐近方差或者协方差, 这会给统计推断带来不便, 尤其是对于半参数或者非参数模型, 其方差估计往往很困难.

正如 1.1 节中所提, 近二十多年来经验似然广受关注, 被运用在了多种模型之中, 它是经典的参数似然比推断向非参数情形的自然推广. 同时其对数经验似然比统计量的分布仍然满足 Wilks 定理, 即渐近服从卡方分布. 此结论可用于构造置信域和假设检验. 在 $p$ 固定的时候, 可以基于 $M$ 估计量研究其相关性质, 见文献 [27]. 当 $p$ 与样本量 $n$ 有关, 甚至随着 $n\to\infty$ 而趋向于无穷的时候, 文献 [28] 和文献 [29] 研究了 $\beta$ 的 $M$ 估计量的一致性和渐近正态性, 而该研究需要 $p$ 与样本量 $n$ 相比不能太大.

然而在生物信息学中, 对于线性模型 (1.3.1) 的统计推断有时会出现 $p$ 是 $n$ 的指数阶的情况. 当 $\beta$ 的很多分量为零, 即稀疏时, 我们可以首先选出非零的分量, 然后对被选出来的非零分量做统计推断, 详见文献 [30]. 在稀疏变量的选择上类似于著名的 Lasso 估计量, 见文献 [31], 及其扩展变量, 见文献 [32], 文献 [33], 都是非常有效的. 文献 [34] 研究了高维线性回归模型中的变量选择问题.

而当 $\beta$ 的非零分量个数很多时, 简单地考虑变量选择是不够的, 我们需要

寻求新的方法. 对于模型 (1.3.1), 传统的经验似然比函数构造如下

$$L_n^T(\beta) = \sup\left\{\prod_{i=1}^n (nq_i) : q_i \geq 0, \sum_{i=1}^n q_i = 1, \sum_{i=1}^n q_i(Y_i - \beta^T X_i)X_i = 0\right\}, \quad (1.3.2)$$

相应地, 其对数经验似然比函数为

$$l_n^T(\beta) = \sup\left\{\sum_{i=1}^n \log(nq_i) : q_i \geq 0, \sum_{i=1}^n q_i = 1, \sum_{i=1}^n q_i(Y_i - \beta^T X_i)X_i = 0\right\} \quad (1.3.3)$$

其对数经验似然比可简单写为

$$r_n^T(\beta) = -2\log(\prod_{i=1}^n (nq_i)).$$

对于固定的 $p$, 根据文献 [3], 有 $r_n^T(\beta_0) \xrightarrow{d} \chi_p^2$, 当 $n \to \infty$ 时. 然而当 $p$ 很大的时候, 这就涉及到 1.2 节所说的好的定义的问题, 导致计算似然函数会出现其最大值非平凡甚至得不到的情况.

让我们欣喜的是, 文献 [35] 对于模型 (1.3.1) 提出了一个新的经验似然方法用于检验 $H_0 : \beta = \beta_0 \longleftrightarrow H_1 : \beta \neq \beta_0$, 我们简称为 PEL 方法, 这里 $\beta_0$ 是一个给定的值. 在他们的方法里, 约束条件与维数 $p$ 无关. 他们从估计量 $\vartheta = (\beta_0 - \beta)^T\Sigma^2(\beta_0 - \beta)$ 入手, 对于 $\Sigma = E(X_1X_1^T)$ 正定的情况, 易知上述假设检验问题等价于检验 $H_0 : \vartheta = 0 \longleftrightarrow H_1 : \vartheta \neq 0$. 所以他们的思路是找到这样的估计量 $\vartheta$.

于是他们把样本数据分为两部分. 具体为: 令 $m = [\frac{n}{2}]$, 即 $\frac{n}{2}$ 的整数部分, 并定义 $\tilde{X}_i = X_{m+i}, \tilde{Y}_i = Y_{m+i}, \tilde{\varepsilon}_i = \varepsilon_{m+i}$. 对于 $i = 1, \ldots, m$, 定义

$$W_i(\beta) = (Y_iX_i - X_iX_i^T\beta)^T(\tilde{Y}_i\tilde{X}_i - \tilde{X}_i\tilde{X}_i^T\beta).$$

这实际上相当于用每两个旧的观测值 $(X_i, Y_i)$ 和 $(X_{m+i}, Y_{m+i})$ 构造一个新的观测值 $W_i(\beta), i = 1, \ldots, m$. 易证明,

$$EW_i(\beta_0) = E[(X_iX_i^T(\beta_0 - \beta) + X_i\varepsilon_i)^T(\tilde{X}_i\tilde{X}_i^T(\beta_0 - \beta) + \tilde{X}_i\tilde{\varepsilon}_i)]$$

$$= \quad (\beta_0-\beta)^T\Sigma^2(\beta_0-\beta).$$

于是, 当 $\Sigma$ 正定的时候, 检验 $H_0:\beta=\beta_0 \longleftrightarrow H_1:\beta\neq\beta_0$ 等价于检验$H_0: EW_1(\beta_0)=0 \longleftrightarrow H_1: EW_1(\beta_0)\neq 0$. 注意到当 $\|\beta_0-\beta\|$ 很小的时候, $EW_i(\beta_0)=O\|\beta_0-\beta\|^2$ 而非 $\|\beta_0-\beta\|$, 其中 $\|\cdot\|$ 表示向量的 $L_2$ 范数. 这就导致直接对估计式 $EW_1(\beta_0)=0$ 应用经验似然的方法会出现功效较弱的情况. 为了提高功效, 他们希望能够找到另一个线性估计式 $EW_i^*(\beta_0)=0$, 而恰巧 $EW_i^*(\beta_0)$ 足够接近 $O(\|\beta_0-\beta\|_1)$, 其中 $\|\beta_0-\beta\|_1$ 是 $L_1$ 范数, 这样就可以抓住$\beta-\beta_0$的微小变化. 特别地, 定义

$$W_i^*(\beta)=(Y_iX_i-X_iX_i^T\beta)^T\mathbf{1}_p+(\tilde{Y}_i\tilde{X}_i-\tilde{X}_i\tilde{X}_i{}^T\beta)^T\mathbf{1}_p,$$

$i=1,\ldots,m$. 其中 $\mathbf{1}_p=(1,1,\ldots,1)^T\in R^p$.

基于此, 文献 [35] 定义了 $\beta$ 的经验似然比函数如下

$$L_n^P(\beta)=\sup\left\{\prod_{i=1}^m(mq_i): q_i\geq 0, \sum_{i=1}^m q_i=1, \sum_{i=1}^m q_iW_i(\beta)=0, \sum_{i=1}^m q_iW_i^*(\beta)=0\right\}.$$

如下定理表明对于上述经验似然方法, Wilks 定理依然成立.

**定理 1**[35]　令 $\beta_0$ 是参数 $\beta$ 的真值. 假设 $\Sigma$ 是正定的且存在某些 $\delta>0$ 满足

$$\frac{E|X_1^T\tilde{X}_1|^{2+\delta}}{\{\mathbf{tr}(\Sigma^2)\}^{(2+\delta)/2}}\left(\frac{E|\varepsilon_1|^{2+\delta}}{\sigma^{2+\delta}}\right)^2=o(m^{\delta/2}),$$

及

$$\frac{E|X_1^T\mathbf{1}_p|^{2+\delta}}{\{E(X_1^T\mathbf{1}_p)^2\}^{(2+\delta)/2}}\left(\frac{E|\varepsilon_1|^{2+\delta}}{\sigma^{2+\delta}}\right)=o(m^{\delta/2}),$$

其中 $\sigma^2=Var(\varepsilon_1)$, $\mathbf{tr}(\Sigma^2)$ 表示 $\Sigma^2$ 的迹. 则 $-2\log L_n^P(\beta_0)\xrightarrow{d}\chi_2^2$, 这里 "$\xrightarrow{d}$" 表示依分布收敛.

上面所提出的经验似然方法的一个优点是: 如果我们关于该假设检验问题能找到更多的信息, 那么这些信息就可以加入到估计式中, 或者也可以用其

他满足 $E\tilde{W}_1(\beta) = O(\|\beta_0 - \beta\|_1)$ 的 $\tilde{W}_1(\beta)$ 来代替$W_1^*(\beta)$. 此外, 文献 [35] 的模拟研究表明: 无论是从犯第一类错误还是功效函数的角度而言, 他们的方法对于假设检验问题 $H_0 : \beta = \beta_0 \longleftrightarrow H_1 : \beta \neq \beta_0$ 都拥有令人满意的结果.

## 1.4 本文思路及安排

受到 PEL 方法的启发, 我们的思路是利用已有的观测值去构造 $\omega_i(\beta)$, 构造出来的 $\omega_i(\beta)$ 需满足: (1) $E\omega_i(\beta_0) = 0$; (2) $E\omega_i(\beta)$ 非常接近于 $\beta_0 - \beta$ 的 $L_1$ 范数. 由此将经验似然的方法应用于估计式 $E\omega_i(\beta_0) = 0$, 从而解决上一节提到的 $\beta$ 的假设检验问题.

注意到 PEL 方法是将样本观测值分为两组, 由此我们很自然地想到也可以将样本观测值分为三组或者四组. 然而, 尽管加入更多的约束条件可以更好地提高检验功效, 但这会导致偏经验似然函数的计算变得更复杂. 后来我们尝试着把传统的经验似然方法里 (1.3.2) 与 $p$ 有关的高维约束条件巧妙地变换成一维情形, 即把 $(Y_i - \beta^T X_i)X_i^T$ 的所有分量相加, 以此构造出$\omega_i(\beta), i = 1, \ldots, n$. 用式子表示即为

$$\omega_i(\beta) = (Y_i - \beta^T X_i)X_i^T \mathbf{1}_p, i = 1, \ldots, n,$$

从而对于模型 (1.3.1) 提出了一个简单的经验似然方法. 在给定正则条件下, 我们也可以证明 Wilks 定理在这个简单的经验似然方法里成立, 于是基于该定理的假设检验问题可以继续研究. 为了对比这个方法和PEL方法的效用, 在模拟研究中我们所考虑的例 1 与文献 [35] 里的例子完全相同. 此外, 1.2 节所研究的调整的经验似然方法也将被应用于我们所提出的简单的经验似然方法及 PEL 方法以提升两者的表现. 具体的方法及模拟研究将在第 2 章中详细呈现.

另一个方面, 我们注意到文献 [35] 模拟研究的例子中回归系数取值的特殊性, 在他们的例子里 $\beta$ 和 $\beta_0$ 的每一个分量变化都是一致的. 而在实际运用中, $\beta$ 和 $\beta_0$ 分量之间可能存在的更复杂的关系. 进行相关模拟后, 我们发现 PEL 方法和我们之前所提出的简单的经验似然方法基本不影响犯第一类错误的概率, 但会导致犯第二类错误的概率变大很多. 这是由于在 $\omega_i(\beta)$ 中, 我们把

$(Y_i - \beta^T X_i)X_i^T$ 的所有分量直接相加, 这对 $\beta$ 和 $\beta_0$ 的每一个分量变化一致的情况没有影响, 而当其分量变化不一致的时候, 我们应该考虑将 $(Y_i - \beta^T X_i)X_i^T$ 的分量加权相加, 即将原来 $(Y_i - \beta^T X_i)X_i^T$ 后面所乘的 $\mathbf{1}_p$ 换为其他 $p$ 维随机向量, 且该向量的分量不能都相等. 于是我们构造了

$$\omega_i^k(\beta) = (Y_i - \beta^T X_i)X_i^T \alpha_k,$$

其中 $k = 1, 2, 3, 4, 5$, $\alpha_k$ 为相互独立的随机向量.

这里, 之所以构造 5 个估计式, 是因为我们发现随着维数的增加, 我们适当的多加入估计式, 会取得更好的效果. 而在维数不高的情况下, 减少估计式的个数会节约计算成本且对结果影响不会太大. 所以我们针对不同的维数, 有区别地加入了估计式的个数, 这也是本文的创新点之一. 具体的方法及模拟研究将在第3章中详细呈现.

本文内容安排如下：

第二章先介绍我们所提出的简单的经验似然方法及其调整, 并引出相关定理结论; 其次, 为了对比我们所提出的几个方法和 PEL 方法, 我们在几个不同的模型下就假设检验问题 $H_0 : \beta = \beta_0 \longleftrightarrow H_1 : \beta \neq \beta_0$ 做了相关模拟. 模拟结果表明：我们所提出的简单的经验似然方法总体上要优于 PEL 方法, 并且在很多情况下, 调整的方法能进一步降低犯第一类错误的概率和提高功效函数的表现.

第三章考虑 $\beta$ 和 $\beta_0$ 分量之间可能存在的更复杂的关系, 即并非 $\beta$ 和 $\beta_0$ 的每一个分量变化都是一致的情况, 对此提出了一般情形下的经验似然方法, 并给出相关定理及模拟研究. 模拟结果显示: 当样本量较小时, 该方法既能保证犯第一类错误的概率接近于给定的显著性水平, 也能将犯第二类错误的概率控制在一个可接受范围内; 当样本量足够大时, 该方法更能确保一个令人满意的效果. 同时, 由于不同的维数加入的约束条件的个数不同, 因此运算效率也能得到保证.

第四章对于前两章所涉及到的主要引理和定理进行相关证明.

第五章是对本文的总结和展望.

# 第 2 章 高维线性模型下的简单经验似然

高维数据, 即数据的维数 $p$ 很大, 在很多领域正受到越来越多的关注. 相关例子在超谱影像学、金融数据、计算生物学、基因组学、机器学习等领域很容易见到. 它使前人所研究的经典的统计方法受到了较大的冲击, 是目前统计学领域和应用研究中面临困难最多、挑战最严峻同时也是最有可能取得突破的研究领域之一. 近几年国内外的大型会议几乎都涉及到高维数据, 并且为此专题还开展过几次国际会议. 对高维数据的研究, 通常情况需要假定真实模型位于一个低维空间(至少协变量维数要比样本容量低), 也就是通常说的稀疏性假定. 然后再按照一定的规则选择变量进行降维. 概括而言, 目前常用的变量选择方法有: 预测误差方法、序贯方法、惩罚方法以及信息论方法, 当然它们可能中间有部分是相互重合的, 见文献 [36]. 通过变量选择, 将对因变量根本没有影响或者影响较小的变量剔除, 保留影响较大的变量, 从而建立简洁的模型.

然而在高维线性模型的变量选择中, 会面临两个比较突出的问题: 一是计算速度, 二是统计的准确性. 不仅如此, 目前变量选择的一些方法还是基于模型误差为正态分布, 并借助最小二乘或惩罚最小二乘的思想而建立. 这对于误差项分布未知的情况就不再适用. 此外, 对于 1.3 节所提到的假设检验问题中 $\beta$ 的许多分量不为零的情况, 变量选择的方法也不再合适. 于是针对此类问题, 我们想要寻求其他方法来解决.

在这一章里, 我们考虑线性模型 (1.3.1) 中 $\beta$ 的假设检验问题. 即检验 $H_0 : \beta = \beta_0 \longleftrightarrow H_1 : \beta \neq \beta_0$.

当 $p$ 固定的时候, 除了绪论中所提到的那些方法, 我们还可以基于霍特林 $T^2$ 检验统计量进行相关假设检验, 即

$$HT = \frac{1}{\hat{\sigma}^2}(\hat{\beta} - \beta_0)^T \left( \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T \right)^{-1} (\hat{\beta} - \beta_0), \tag{2.0.1}$$

其中 $\hat{\beta} = (\frac{1}{n}\sum_{i=1}^{n} X_i X_i^T)^{-1} \frac{1}{n} \sum_{i=1}^{n} Y_i X_i$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\beta}^T X_i)^2$. 易知当

$n \to \infty$ 时, $HT \xrightarrow{d} \chi_p^2$. 但是, 当 $p$ 比较大的时候, 寻找 (2.0.1) 中的逆矩阵会出现问题, 见文献 [35]. 此外, 我们在绪论中提到过, 文献 [3] 证明了 $r_n^T(\beta_0) \xrightarrow{d} \chi_p^2$. 显然, 当 $p$ 随 $n$ 同时趋于 $\infty$ 时, $r_n^T(\beta_0) \xrightarrow{d} \infty$. 也就是说, 利用传统经验似然所作的统计推断对高维协变量情形失效. 针对此问题, 文献 [37] 对于模型 (1.3.1) 作了相关研究. 在他们的文章中, 令 $V_n = \sum_{i=1}^n X_i X_i^T$. 显然 $V_n$ 是 $p \times p$ 的对称矩阵且 $V_n \geq 0$. 记 $\lambda_{1n}$ 和 $\lambda_{pn}$ 分别是 $V_n$ 的最小和最大特征根. 记 $M_n^2 = max_{1 \leq i \leq n} \|X_i\|^2$, $\|\cdot\|$ 为通常的欧氏模. 由矩阵理论, 有 $p\lambda_{1n} \leq \sum_{i=1}^n \|X_i\|^2 \leq nM_n^2$ 和 $\lambda_{pn} \leq \sum_{i=1}^n \|X_i\|^2 \leq nM_n^2$. 假定 $\|X_1\| \neq 0$, 并假设:

$(A_1)$ 对某 $r > 0$, 有 $E(|\varepsilon_1|^{2+r}) < \infty$.

$(A_2)$ 对某 $0 \leq \iota < \frac{r}{4+2r}$, 有 $\frac{nM_n^2}{p\lambda_{1n}} = O(n^\iota)$.

$(A_3)$ $p = p(n) \to \infty (n \to \infty)$, 且

$$p^{\frac{3}{2}} = \begin{cases} O(n^{\frac{r}{4+2r}-\iota}) & \text{若 } r < 1; \\ o(n^{\frac{1}{6}-\iota}) & \text{否则.} \end{cases}$$

则有如下定理成立.

**定理 2**[37] 假设 $(A_1) - (A_3)$ 成立, 则在真值 $H_0 : \beta = \beta_0$ 下, 有

$$\frac{r_n^T(\beta) - p}{\sqrt{2p}} \xrightarrow{d} N(0,1), n \to \infty.$$

因此, 当协变量维数 $p$ 以某种合理的速度趋于无穷大时, 由上述定理我们仍可以利用经验似然方法构造 $\beta$ 的置信域. 文献 [38] 和文献 [39] 中也给出了类似的定理或是推论. 不过此时有关临界值的确定依赖于正态分布而非卡方分布, 这会导致绪论中所提到的两个问题. 我们下面希望建立依赖于卡方分布的检验统计量.

## 2.1 高维数据的简单经验似然方法

正如绪论中所提到的那样, 对于高维线性回归模型 (1.3.1), 基于 (1.3.2)

中 $L_n^T(\beta)$ 的方法的主要问题在于：约束条件的维数会随着 $p$ 的增长而变大, 而这会使得相应似然函数的计算出现非平凡甚至得不到的情况. 为了解决这个问题, 同时也是受到 PEL 方法的启发, 我们的思路是去寻找另外的相互独立的 $\omega_i(\beta)$, 使得找到的 $\omega_i(\beta)$ 与维数 $p$ 无关, 同时又满足: (1) $E\omega_1(\beta_0) = 0$; (2) $E\omega_1(\beta) = O(\|\beta_0 - \beta\|_1)$, 以此来取代 $L_n^T(\beta)$ 中的约束条件 $\sum_{i=1}^n q_i(Y_i - \beta^T X_i)X_i = 0$. 从而运用经验似然的方法解决上面的假设检验问题 $H_0 : \beta = \beta_0 \longleftrightarrow H_1 : \beta \neq \beta_0$.

注意到 PEL 方法的思想是将样本观测值分组, 由此我们很自然的想到也可以将样本观测值分为三组或者四组. 然而, 尽管加入更多的估计式可以更好地提高检验功效, 但这会增加计算成本. 后来我们尝试着把传统的经验似然比函数 (1.3.2) 中与 $p$ 有关的高维约束条件巧妙地变换成一维情形, 即把 $(Y_i - \beta^T X_i)X_i^T$ 的所有分量相加, 以此构造出 $\omega_i(\beta), i = 1, \ldots, n$. 而非文献 [35] 方法里的先分组, 再相加. 于是我们定义

$$\omega_i(\beta) = (Y_i - \beta^T X_i)X_i^T \mathbf{1}_p,$$

i=1,...,n, 那么

$$E\omega_i(\beta_0) = E[(Y_i - \beta_0^T X_i)X_i^T \mathbf{1}_p] = (\beta_0 - \beta)^T \Sigma \mathbf{1}_p.$$

显然, 当 $\beta = \beta_0$ 成立时我们有 $E\omega_i(\beta_0) = 0$, 且 $E\omega_i(\beta) = O(\|\beta_0 - \beta\|_1)$.

进一步, 经验似然比函数可以很自然地被定义为

$$L_n^H(\beta) = \sup\left\{\prod_{i=1}^n (nq_i) : q_1 \geq 0, \ldots, q_n \geq 0, \sum_{i=1}^n q_i = 1, \sum_{i=1}^n q_i\omega_i(\beta) = 0\right\}.$$

通过拉格朗日乘子法, 我们有

$$q_i = \frac{1}{n}\frac{1}{1 + \lambda\omega_i(\beta)}, i = 1, \ldots, n, \tag{2.1.1}$$

其中 $\lambda$ 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\omega_i(\beta)}{1+\lambda\omega_i(\beta)}=0, \tag{2.1.2}$$

相应地, 其对数经验似然比为

$$-2\log L_n^H(\beta)=2\sum_{i=1}^{n}\log\{1+\lambda\omega_i(\beta)\}.$$

下面这个定理表明: 对于我们所提出的方法, Wilks 定理依然成立.

**定理 3** 假设 $E(X_iX_i^T)$ 是正定的并且$\beta_0$ 是参数 $\beta$ 的真值. 那么当 $n\to\infty$ 时, 我们有 $-2\log L_n^H(\beta_0)\xrightarrow{d}\chi_1^2$, 其中 "$\xrightarrow{d}$" 表示依分布收敛.

如绪论中所提, 对于一个给定的 $\beta$, 传统的经验似然比函数并不能经常被很好地定义. 为了解决这个问题, 文献 [22] 在传统的经验似然方法里加入了一个伪观测值, 从而作出了一个新奇的调整. 调整后的经验似然方法保留了之前方法的所有最优性准则, 不仅如此, 该方法下的区间覆盖率更接近于置性水平, 而且还不需要 Bartlett 校正和 bootstrap 方法里那么复杂的程序, 详见文献 [22]. 调整的方法已被许多作者应用在他们的文章里, 诸如文献 [40], 文献 [41] 及文献 [42] 等.

因此, 我们运用文献 [22] 里面所提出的加入伪观测值的方法来对我们提出的简单的经验似然方法做一些调整. 为了书写方便, 记 $\omega_i(\beta)$ 为 $\omega_i$, $i=1,\ldots,n$. 根据文献 [22], 我们用

$$\omega_{n+1}=-a_n\bar{\omega}_n=-a_nn^{-1}\sum_{i=1}^{n}\omega_i$$

作为一个伪观测值, 其中 $a_n$ 为大于 0 的某些值, 并重新定义 $\beta$ 的经验似然比函数为

$$L_n^{AH}(\beta)=\sup\left\{\prod_{i=1}^{n+1}[(n+1)q_i]:q_1\geq 0,\ldots,q_{n+1}\geq 0,\sum_{i=1}^{n+1}q_i=1,\sum_{i=1}^{n+1}q_i\omega_i(\beta)=0\right\}.$$

同样通过拉格朗日乘子法, 我们有

$$q_i = \frac{1}{n+1} \cdot \frac{1}{1+\lambda\omega_i(\beta)}, i = 1, \ldots, n+1, \tag{2.1.3}$$

其中 $\lambda$ 满足

$$\frac{1}{n+1}\sum_{i=1}^{n+1} \frac{\omega_i(\beta)}{1+\lambda\omega_i(\beta)} = 0, \tag{2.1.4}$$

且其对数经验似然比为

$$-2\log L_n^{AH}(\beta) = 2\sum_{i=1}^{n+1} \log\{1+\lambda\omega_i(\beta)\}. \tag{2.1.5}$$

这里, 只要 $a_n = o_p(n)$, 则 Wilks 定理依然成立.

**定理 4** 在定理 3 的假设条件下, 令 $a_n = o_p(n)$. 那么当 $n \to \infty$ 时, 我们有 $-2\log L_n^{AH}(\beta_0) \xrightarrow{d} \chi_1^2$.

**注 1.** 我们也可以用文献 [22] 里面伪观测值的方法对 PEL 方法做一些调整. 令伪观测值为 $(W_{m+1}, W_{m+1}^*)^T$, 其中

$$W_{m+1} = -a_m\bar{W}_m = -a_m m^{-1}\sum_{i=1}^{m} W_i(\beta),$$

$$W_{m+1}^* = -a_m\bar{W}_m^* = -a_m m^{-1}\sum_{i=1}^{m} W_i^*(\beta).$$

类似地, 我们可以得到 $L_n^{AP}(\beta)$, 并且在一些给定的正则条件下, 有 $-2\log L_n^{AP}(\beta_0) \xrightarrow{d} \chi_2^2$. 用调整的 PEL 方法所做的模拟也将在 2.2 节中给出.

**注 2.** 为了应用文献 [22] 中调整的方法, $a_n$ 需要被预先确定为某个值. 由于文献 [22] 中指出令 $a_n = \max(1, \log(n)/2)$ 在很多例子中都起到好的效果, 在本文的模拟中我们就采用这个值. 第三部分中的模拟研究表明使用 $a_n = \max(1, \log(n)/2)$ 作为调整因子确实能提高经验似然方法的效果.

## 2.2 模拟研究及结果

我们在这一节里做了一些模拟研究以评估我们所提出的方法的效用, 并将它们与 PEL 方法作比较. 为此, 我们考虑下面的高维线性回归模型

$$Y = \beta^T X + \varepsilon,$$

其中回归系数 $\beta = \beta_0 + \frac{\Delta}{\sqrt{n}}\mathbf{1}_p$ 对于某些$\beta_0 \in R^p$ 及 $\Delta \geq 0$, 且$\varepsilon_i \sim t(8)$, $i = 1, \ldots, n$. 我们将观测值表示为 $\{(X_i, Y_i), i = 1, \ldots, n\}$. 在下面的例子中, 我们检验 $H_0 : \beta = \beta_0 \longleftrightarrow H_1 : \beta \neq \beta_0$ (显著性水平 $\alpha$ 为0.05). 为了便于比较我们的方法与 PEL 方法, 其中 *例1* 即为文献 [35] 中的例子.

*例1.* $\beta_0 = \mathbf{1}_p$, $\{X_i\}$ 为服从 $N(\mathbf{0}, \Sigma_N)$ 的独立同分布的随机变量, 其中 $\Sigma_N = (0.5^{(|i-j|)})_{1 \leq i,j \leq p}$.

*例2.* $\beta_0 = \mathbf{1}_p$, $X_1, \ldots, X_{0.7n}$ 为服从$N(0.5. * \mathbf{1}_p, \Sigma_N)$ 的独立同分布的随机变量, 这里 $\Sigma_N = (0.8^{(|i-j|)})_{1 \leq i,j \leq p}$, 又 $X_{0.7n+1}, \ldots, X_n$ 为服从自由度为 5 的多元 t 分布, 其相关阵为 $\Sigma_t = (0.4^{(|i-j|)})_{1 \leq i,j \leq p}$.

*例3.* $\{X_i\}$ 为服从 $N(\mathbf{0}, \Sigma_N)$ 的独立同分布的随机变量, 其中 $\Sigma_N = (0.5^{(|i-j|)})_{1 \leq i,j \leq p}$. 而 $\beta_0 = (\beta_{01}^T, \beta_{02}^T, \beta_{03}^T, \beta_{04}^T, \beta_{05}^T)^T$, 其中 $\beta_{0j} = j. * \mathbf{1}_{p/5}$, $j = 1, \ldots, 5$, 这里的 $.*$ 表示向量 $\mathbf{1}_{p/5}$ 的每个元素都乘上数值 $j$.

*例4.* $\{X_i\}$ 的生成方式与 *例2* 中相同, $\beta_0$ 与 *例3* 中的一致.

对于上述每一个例子, 我们用不同的方法分别计算了其相应地犯第一类错误的概率 ($\Delta = 0$) 和功效函数 ($\Delta = 0.3$). 这些计算结果都是在样本大小 $n$ ($n = 200$ 或者 $n = 1000$) 的情况下, 随机重复 10000 次模拟所得. 模拟结果展示在表 2-1 至 2-8 中, 其中 HEL 表示我们所提出的简单的经验似然方法, PEL 是文献 [35] 中提出的方法, 而 AHEL 和 APEL 分别表示调整的 HEL 和 PEL 方法.

**注 3.** 为了比较几种方法的效果, 在本章中我们所考虑的回归模型与文献 [35] 中的一致. 值得注意的是, 在备择假设下, 该模型中的回归系数向量 $\beta$ 是 $\beta_0$ 的每一个分量做相同的变换所得, 而实际情况中可能会出现 $\beta$ 与 $\beta_0$ 的一些甚至所有分量变换不一致的情况. 这部分内容将在下一章给出讨论.

表 2-1 例*1* 中四类方法犯第一类错误的概率 ($\alpha = 0.05$).

| p | HEL | | PEL | | AHEL | | APEL | |
|---|---|---|---|---|---|---|---|---|
| | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 |
| 5 | 0.0563 | 0.0500 | 0.0894 | 0.0620 | 0.0496 | 0.0479 | 0.0733 | 0.0564 |
| 10 | 0.0568 | 0.0508 | 0.0872 | 0.0574 | 0.0502 | 0.0478 | 0.0774 | 0.0539 |
| 20 | 0.0575 | 0.0496 | 0.0829 | 0.0587 | 0.0495 | 0.0467 | 0.0739 | 0.0505 |
| 30 | 0.0564 | 0.0459 | 0.0822 | 0.0583 | 0.0559 | 0.0497 | 0.0722 | 0.0590 |
| 40 | 0.0569 | 0.0500 | 0.0820 | 0.0592 | 0.0495 | 0.0488 | 0.0709 | 0.0534 |
| 50 | 0.0521 | 0.0506 | 0.0814 | 0.0539 | 0.0524 | 0.0562 | 0.0726 | 0.0558 |
| 60 | 0.0537 | 0.0476 | 0.0837 | 0.0567 | 0.0517 | 0.0540 | 0.0744 | 0.0558 |
| 70 | 0.0569 | 0.0523 | 0.0831 | 0.0597 | 0.0548 | 0.0514 | 0.0743 | 0.0551 |
| 80 | 0.0575 | 0.0522 | 0.0812 | 0.0575 | 0.0500 | 0.0471 | 0.0737 | 0.0573 |
| 90 | 0.0521 | 0.0495 | 0.0779 | 0.0549 | 0.0521 | 0.0529 | 0.0705 | 0.0547 |
| 100 | 0.0541 | 0.0523 | 0.0813 | 0.0576 | 0.0518 | 0.0504 | 0.0703 | 0.0548 |

表 2-2 例*1* 中四类方法的功效函数 ($\alpha = 0.05$).

| p | HEL | | PEL | | AHEL | | APEL | |
|---|---|---|---|---|---|---|---|---|
| | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 |
| 5 | 0.1405 | 0.1369 | 0.1572 | 0.1279 | 0.1355 | 0.1390 | 0.1425 | 0.1216 |
| 10 | 0.2735 | 0.2605 | 0.2534 | 0.2141 | 0.2638 | 0.2613 | 0.2310 | 0.2067 |
| 20 | 0.4901 | 0.4872 | 0.4330 | 0.3882 | 0.4855 | 0.4886 | 0.4176 | 0.3992 |
| 30 | 0.6648 | 0.6614 | 0.6025 | 0.5728 | 0.6608 | 0.6645 | 0.5648 | 0.5597 |
| 40 | 0.7826 | 0.7974 | 0.7194 | 0.7041 | 0.7822 | 0.7902 | 0.6988 | 0.6977 |
| 50 | 0.8681 | 0.8721 | 0.8082 | 0.8041 | 0.8625 | 0.8751 | 0.7851 | 0.7973 |
| 60 | 0.9135 | 0.9262 | 0.8751 | 0.8757 | 0.9095 | 0.9295 | 0.8536 | 0.8749 |
| 70 | 0.9494 | 0.9568 | 0.9202 | 0.9224 | 0.9462 | 0.9557 | 0.9034 | 0.9200 |
| 80 | 0.9689 | 0.9753 | 0.9432 | 0.9509 | 0.9654 | 0.9757 | 0.9318 | 0.9509 |
| 90 | 0.9785 | 0.9866 | 0.9634 | 0.9707 | 0.9804 | 0.9869 | 0.9561 | 0.9709 |
| 100 | 0.9888 | 0.9922 | 0.9770 | 0.9826 | 0.9871 | 0.9925 | 0.9745 | 0.9806 |

为了与 PEL 方法作对比, 表 2-1 和 2-2 使用了与文献 [35] 中完全一样的例子. 从表中可以看出: 当样本量 $n = 200$ 时, PEL 方法下犯第一类错误的概率比较稳定, 除了两个值略高, 其他几个都维持在 0.084 以下, 但是距离显著性水平 $\alpha = 0.05$ 还有一定的差距. 而此时相对应的功效函数随着维数的增加逐渐变大, 且最后有接近于 1 的趋势. 当样本量 $n = 1000$ 时, 犯第一类错误的概率有明显的下降, 绝大部分都降到了 0.06 以下, 并保持稳定, 这与显著性水平 $\alpha = 0.05$ 已经非常接近, 此时的功效函数在 $p = 50$ 之前, 比 $n = 200$ 时的值略小, 但有越来越接近的趋势, 随着维数的进一步增大, 功效函数也开始大于 $n = 200$ 时对应的值, 当 $p = 100$ 时, 功效函数比样本量小的情况更接近于 1. 总

体而言, PEL 方法对于上述假设检验问题有一定的效果, 特别是当样本量较大的时候.

相对应的, 在 APEL 方法下, 无论是样本量 $n = 200$ 还是 $n = 1000$, 犯第一类错误的概率都比 PEL 方法下的值要小. 在 $n = 200$ 的时候, 犯第一类错误的概率稳定在 0.078 以下, 且对于每个 $p$, APEL 方法下犯第一类错误的概率比 PEL 方法下相对应的值至少小了 0.007, 对于 $p = 5$, 甚至小了 0.0161, 这相对于显著性水平 $\alpha = 0.05$ 来说, 是一个比较显著的降低. 在 $n = 1000$ 的时候, 犯第一类错误的概率稳定在 0.059 及其以下, 比之前未调整方法下的值也有所降低. 功效函数方面, APEL 方法比 PEL 方法下的值略有降低, 但幅度非常小, 对于有的 $p$, 甚至持平或略有提高. 所以整体而言, APEL 方法要优于 PEL 方法, 特别是从犯第一类错误的概率而言, 而该层面也正是假设检验问题中所主要考虑的层面.

表 2-3 例2 中四类方法犯第一类错误的概率 ($\alpha = 0.05$).

| p | HEL | | PEL | | AHEL | | APEL | |
|---|---|---|---|---|---|---|---|---|
| | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 |
| 5 | 0.0558 | 0.0527 | 0.0970 | 0.0611 | 0.0518 | 0.0524 | 0.0865 | 0.0623 |
| 10 | 0.0563 | 0.0518 | 0.0944 | 0.0653 | 0.0569 | 0.0534 | 0.0790 | 0.0614 |
| 20 | 0.0562 | 0.0527 | 0.0900 | 0.0610 | 0.0531 | 0.0490 | 0.0804 | 0.0608 |
| 30 | 0.0548 | 0.0501 | 0.0847 | 0.0604 | 0.0557 | 0.0491 | 0.0768 | 0.0614 |
| 40 | 0.0564 | 0.0506 | 0.0912 | 0.0575 | 0.0547 | 0.0494 | 0.0786 | 0.0614 |
| 50 | 0.0610 | 0.0526 | 0.0869 | 0.0614 | 0.0539 | 0.0481 | 0.0778 | 0.0582 |
| 60 | 0.0529 | 0.0504 | 0.0899 | 0.0584 | 0.0484 | 0.0475 | 0.0725 | 0.0557 |
| 70 | 0.0555 | 0.0494 | 0.0860 | 0.0596 | 0.0487 | 0.0474 | 0.0755 | 0.0628 |
| 80 | 0.0572 | 0.0594 | 0.0908 | 0.0599 | 0.0508 | 0.0496 | 0.0727 | 0.0593 |
| 90 | 0.0560 | 0.0585 | 0.0871 | 0.0565 | 0.0502 | 0.0498 | 0.0738 | 0.0570 |
| 100 | 0.0539 | 0.0482 | 0.0887 | 0.0628 | 0.0540 | 0.0532 | 0.0796 | 0.0566 |

再来看 HEL 方法, 我们欣喜地发现: 当样本量 $n = 200$ 的时候, 除了对于极个别的 $p$, 其他犯第一类错误的概率就都低于 0.058, 且其中有四个值低于 0.055, 而相同情况下, 在 PEL 方法里每个值都高于 0.077, 显然从犯第一类错误的概率层面, 这是一个很大的提升, 而此时的功效函数亦是随着维数的增加逐渐变大, 除了 $p = 5$ 的时候, HEL 方法下的功效函数低于 PEL 方法, 其他值都优于相应 PEL 方法下的值. 当样本量 $n = 1000$ 的时候, HEL 方法下犯第一类错误的概率比之前样本量较小的情况更低, 足够接近于 0.05, 甚至有超过一

表 2-4 例2 中四类方法下的功效函数 ($\alpha = 0.05$).

| p | HEL | | PEL | | AHEL | | APEL | |
|---|---|---|---|---|---|---|---|---|
| | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 |
| 5 | 0.2335 | 0.2304 | 0.2420 | 0.1933 | 0.2386 | 0.2279 | 0.1885 | 0.2149 |
| 10 | 0.5573 | 0.5631 | 0.5138 | 0.4713 | 0.5524 | 0.5514 | 0.4834 | 0.4667 |
| 20 | 0.9306 | 0.9396 | 0.8958 | 0.9043 | 0.9275 | 0.9428 | 0.8841 | 0.8975 |
| 30 | 0.9947 | 0.9977 | 0.9904 | 0.9946 | 0.9942 | 0.9978 | 0.9863 | 0.9936 |
| 40 | 0.9996 | 1.0000 | 0.9996 | 0.9999 | 0.9996 | 1.0000 | 0.9991 | 1.0000 |
| 50 | 0.9999 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 60 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 70 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 80 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 90 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 1.0000 |
| 100 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

半的值低于或等于 0.05, 而 PEL 方法没有达到这样的效果. 此时的功效函数在 $p = 40$ 及之后就开始高于对应状态下 $n = 200$ 的值, 当 $p = 100$ 时, 功效函数为 0.9922, 已经相当接近于 1. 也就是说随着样本量的增大, HEL 方法的检验效果也更优, 这一点与 PEL 方法是一致的. 不仅如此, 无论是从犯第一类错误的概率层面还是从功效函数的层面, HEL 方法在 $n = 200$ 时的结果就已经优于 PEL 方法在 $n = 1000$ 时的结果.

表 2-5 例3 中四类方法犯第一类错误的概率 ($\alpha = 0.05$).

| p | HEL | | PEL | | AHEL | | APEL | |
|---|---|---|---|---|---|---|---|---|
| | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 |
| 5 | 0.0538 | 0.0488 | 0.0846 | 0.0597 | 0.0551 | 0.0512 | 0.0777 | 0.0634 |
| 10 | 0.0537 | 0.0495 | 0.0882 | 0.0611 | 0.0500 | 0.0525 | 0.0789 | 0.0577 |
| 20 | 0.0523 | 0.0529 | 0.0867 | 0.0593 | 0.0517 | 0.0473 | 0.0755 | 0.0580 |
| 30 | 0.0580 | 0.0519 | 0.0859 | 0.0571 | 0.0536 | 0.0505 | 0.0716 | 0.0570 |
| 40 | 0.0504 | 0.0522 | 0.0812 | 0.0576 | 0.0509 | 0.0497 | 0.0704 | 0.0556 |
| 50 | 0.0576 | 0.0521 | 0.0832 | 0.0588 | 0.0557 | 0.0490 | 0.0730 | 0.0540 |
| 60 | 0.0511 | 0.0506 | 0.0809 | 0.0550 | 0.0515 | 0.0473 | 0.0726 | 0.0572 |
| 70 | 0.0547 | 0.0498 | 0.0786 | 0.0561 | 0.0516 | 0.0560 | 0.0718 | 0.0579 |
| 80 | 0.0564 | 0.0543 | 0.0757 | 0.0541 | 0.0533 | 0.0526 | 0.0657 | 0.0545 |
| 90 | 0.0548 | 0.0561 | 0.0790 | 0.0549 | 0.0541 | 0.0496 | 0.0764 | 0.0525 |
| 100 | 0.0573 | 0.0493 | 0.0789 | 0.0541 | 0.0514 | 0.0508 | 0.0722 | 0.0578 |

相对应的, 在 AHEL 方法下, 也有与 APEL 方法类似的结论. 在 $n = 200$ 的时候, 犯第一类错误的概率稳定在 0.056 以下, 且有半数的值都非常接近于

表 2-6 例*3* 中四类方法下的功效函数 ($\alpha = 0.05$).

| p | HEL | | PEL | | AHEL | | APEL | |
|---|---|---|---|---|---|---|---|---|
| | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 |
| 5 | 0.1438 | 0.1457 | 0.1635 | 0.1211 | 0.1382 | 0.1371 | 0.1459 | 0.1182 |
| 10 | 0.2635 | 0.2590 | 0.2489 | 0.2115 | 0.2629 | 0.2554 | 0.2282 | 0.2118 |
| 20 | 0.4929 | 0.5003 | 0.4375 | 0.4067 | 0.4836 | 0.4882 | 0.4039 | 0.3990 |
| 30 | 0.6620 | 0.6675 | 0.5919 | 0.5737 | 0.6525 | 0.6733 | 0.5657 | 0.5648 |
| 40 | 0.7916 | 0.7896 | 0.7159 | 0.7131 | 0.7785 | 0.7867 | 0.6903 | 0.7044 |
| 50 | 0.8643 | 0.8769 | 0.8082 | 0.8071 | 0.8632 | 0.8813 | 0.7921 | 0.8076 |
| 60 | 0.9215 | 0.9274 | 0.8686 | 0.8749 | 0.9124 | 0.9251 | 0.8580 | 0.8732 |
| 70 | 0.9469 | 0.9560 | 0.9123 | 0.9204 | 0.9456 | 0.9576 | 0.9041 | 0.9229 |
| 80 | 0.9667 | 0.9758 | 0.9395 | 0.9559 | 0.9657 | 0.9758 | 0.9367 | 0.9510 |
| 90 | 0.9796 | 0.9874 | 0.9631 | 0.9704 | 0.9790 | 0.9868 | 0.9604 | 0.9712 |
| 100 | 0.9873 | 0.9933 | 0.9783 | 0.9853 | 0.9854 | 0.9929 | 0.9732 | 0.9793 |

0.05, 在 $n = 1000$ 的时候, 犯第一类错误的概率基本稳定在 0.054 及其以下, 比之前未调整方法下的值也有所降低. 功效函数方面, AHEL 方法比 HEL 方法下的值略有降低, 但幅度非常小, 对于有的 $p$, 甚至持平或略有提高.

同时, 我们很明显地可以看出, APEL 方法下犯第一类错误的概率依然要高于 HEL 方法下相对应的值, 功效函数的值也比 HEL 方法下的略低. 也就是说, HEL 方法不仅优于 PEL 方法, 甚至优于调整后的 PEL 方法, 即 APEL 方法. 显然, 对于模型 1, AHEL 方法效果最好, HEL 方法次之, 再者为 APEL 方法, PEL 方法的效果相对最弱.

*例2* 、 *例3* 、 *例4* 是为了表明 HEL 方法、APEL 方法和 AHEL 方法不仅对于文献 [35] 中的例子适用, 对于其他更一般地情形也都适用. 从表 2-3-表 2-8可以看出, 在这些例子中, 四中方法对于犯第一类错误的概率和功效函数, 其表现与表 2-1、表 2-2 一致, 且我们所提出的三种方法一如既往地优于 PEL 方法.

概括而言: 当样本量 $n = 200$ 时, 尽管 PEL 方法下犯第一类错误的概率一直保持稳定, 即并不随着 $p$ 的增加有明显变化, 但是它们与给定的显著性水平 $\alpha = 0.05$ 还是有一定差距. 而相比较而言, HEL 方法则在很大程度上降低了犯第一类错误的概率, 使得它们更加接近于显著性水平. 当样本量 $n$ 增加到 1000 时, PEL 方法和 HEL 方法的效果都有提升. 然而, 我们可以注意到, HEL 方法下的计算结果大部分都接近于显著性水平并且将近一半是低于显著性水平的,

表 2-7 例*4* 四类方法犯第一类错误的概率 ($\alpha = 0.05$).

| p | HEL | | PEL | | AHEL | | APEL | |
|---|---|---|---|---|---|---|---|---|
| | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 |
| 5 | 0.0612 | 0.0495 | 0.0998 | 0.0615 | 0.0537 | 0.0500 | 0.0879 | 0.0592 |
| 10 | 0.0537 | 0.0500 | 0.0960 | 0.0630 | 0.0504 | 0.0490 | 0.0797 | 0.0582 |
| 20 | 0.0566 | 0.0487 | 0.0890 | 0.0636 | 0.0500 | 0.0518 | 0.0787 | 0.0630 |
| 30 | 0.0536 | 0.0520 | 0.0886 | 0.0580 | 0.0544 | 0.0496 | 0.0795 | 0.0606 |
| 40 | 0.0525 | 0.0477 | 0.0925 | 0.0618 | 0.0492 | 0.0501 | 0.0787 | 0.0582 |
| 50 | 0.0540 | 0.0503 | 0.0969 | 0.0628 | 0.0521 | 0.0501 | 0.0833 | 0.0549 |
| 60 | 0.0503 | 0.0526 | 0.0876 | 0.0591 | 0.0500 | 0.0500 | 0.0795 | 0.0573 |
| 70 | 0.0510 | 0.0487 | 0.0863 | 0.0647 | 0.0516 | 0.0470 | 0.0749 | 0.0586 |
| 80 | 0.0553 | 0.0479 | 0.0834 | 0.0669 | 0.0515 | 0.0501 | 0.0796 | 0.0595 |
| 90 | 0.0559 | 0.0487 | 0.0861 | 0.0620 | 0.0538 | 0.0478 | 0.0712 | 0.0586 |
| 100 | 0.0562 | 0.0527 | 0.0892 | 0.0544 | 0.0519 | 0.0483 | 0.0716 | 0.0603 |

表 2-8 例*4* 中四类方法下的功效函数 ($\alpha = 0.05$).

| p | HEL | | PEL | | AHEL | | APEL | |
|---|---|---|---|---|---|---|---|---|
| | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 | n=200 | n=1000 |
| 5 | 0.2369 | 0.2301 | 0.2370 | 0.1860 | 0.2298 | 0.2222 | 0.2136 | 0.1833 |
| 10 | 0.5602 | 0.5585 | 0.5075 | 0.4730 | 0.5489 | 0.5603 | 0.4861 | 0.4608 |
| 20 | 0.9314 | 0.9431 | 0.9023 | 0.8984 | 0.9287 | 0.9422 | 0.8868 | 0.9012 |
| 30 | 0.9954 | 0.9976 | 0.9899 | 0.9943 | 0.9953 | 0.9975 | 0.9863 | 0.9934 |
| 40 | 0.9995 | 1.0000 | 0.9991 | 0.9996 | 0.9997 | 0.9999 | 0.9988 | 1.0000 |
| 50 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 1.0000 |
| 60 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 1.0000 |
| 70 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 80 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 90 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 100 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

而 PEL 方法显然达不到这样一个好的效果. 此外, 两种方法的功效函数都随着 $p$ 的增大而增大, 直至接近或等于 1, 但是总的来说, HEL 方法要优于 PEL 方法.

正如我们所期望的那样, AHEL 方法和 APEL 方法都优于它们未调整之前的方法. 当 $n = 1000$ 时, 调整的方法下相当一部分犯第一类错误的概率值都接近于显著性水平 $\alpha = 0.05$, 并且相应的功效函数也令人满意. 总体而言, 无论是犯第一类错误的概率还是功效函数 APEL 方法依然比 HEL 方法要差一些.

# 第 3 章 高维线性模型下的经验似然

从前面的讨论中可以看出 HEL 的方法要优于 PEL, 但同时我们也注意到了注 3 中所提出的 $\beta$ 在选取上的特殊性. 事实上, 在第 2 章的模拟研究中, 由于 $\beta$ 与 $\beta_0$ 每一个分量的变化都一致, PEL 方法与我们所提出的 HEL 方法检验的是 $H_0 : \beta = \beta_0 \longleftrightarrow H_1 : \beta = \beta_0 + \frac{\Delta}{\sqrt{n}}\mathbf{1}_p$, 其中 $\Delta$ 为非零实数. 而实际情况中 $\beta$ 和 $\beta_0$ 之间可能存在更复杂的关系, 例如 $\beta$ 和 $\beta_0$ 的部分分量变化不一致, 而其它分量变化相同, 甚至 $\beta$ 和 $\beta_0$ 两者之间所有的分量变化都不一致, 这对犯第一类错误的概率几乎没有影响, 但会影响到前面四种方法下犯第二类错误的概率, 具体见表 3-1. 为此我们需要其他方法来解决这一问题.

## 3.1 一般情形下的经验似然

同样地, 对于高维线性回归模型 (1.3.1), 当 $\beta$ 和 $\beta_0$ 之间存在更复杂的关系时, 即当备择假设下 $\beta$ 可取$\beta = \beta_0 + \zeta$ 时, 这里的 $\zeta$ 可以是任意与 $\beta_0$ 维数相同的向量, 我们可以尝试着将$(Y_i - \beta^T X_i)X_i^T$ 的分量加权相加, 从而解决更为一般的情形下的假设检验问题 $H_0 : \beta = \beta_0 \longleftrightarrow H_1 : \beta \neq \beta_0$.

这里我们定义

$$\omega_i^k(\beta) = (Y_i - \beta^T X_i)X_i^T \alpha_k,$$

其中 $k = 1, 2, 3, 4, 5$, $\alpha_k$ 为相互独立的随机向量. 相应地, $\beta$ 的经验似然比函数为

$$L_n^K(\beta) = \sup\left\{\prod_{i=1}^{n}(nq_i) : q_i \geq 0 \sum_{i=1}^{n} q_i = 1, \sum_{i=1}^{n} q_i \omega_i^j(\beta) = 0, j = 1, 2, \cdots, k\right\},$$

$i=1,\ldots,n$. 记 $U_i(\beta)$ 为由 $\omega_i^j(\beta), j=1,2,\cdots,k$, 构成的 $k$ 行 1 列的列向量. 通过拉格朗日乘子法, 则有

$$q_i=\frac{1}{n}\frac{1}{1+\lambda^T U_i(\beta)}, i=1,\ldots,n, \tag{3.1.1}$$

其中 $\lambda$ 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{U_i(\beta)}{1+\lambda^T U_i(\beta)}=0, \tag{3.1.2}$$

其对数经验似然比为

$$-2\log L_n^K(\beta)=2\sum_{i=1}^{n}\log\{1+\lambda^T U_i(\beta)\}. \tag{3.1.3}$$

我们称该方法为 KEL 方法. 进一步, 如下定理成立.

**定理 5** 假设 $E(X_iX_i^T)$ 是正定的并且 $\beta_0$ 是参数 $\beta$ 的真值. 那么当 $n\to\infty$ 时, 我们有 $-2\log L_n^K(\beta_0)\xrightarrow{d}\chi_k^2$.

**注 4.** $k$ 的不同取值实际上代表着加入了不同的约束条件个数, 这是因为我们发现随着维数的增加, 适当地多加入约束条件, 会取得更好的效果. 也正是基于这个考虑, 3.2 节的模拟研究中我们没有取 $k=1$. 而在维数不高的情况下, 减少约束条件的个数会节约计算成本且对结果影响不会太大. 实际情况中我们可以根据需要选取不同的 $k$.

**注 5.** 在后面的模拟研究中, 我们取 $\alpha_k\sim U(1,1.5)$ 且相互独立, $k=1,2,3,4,5$. 实际上 $\alpha_k$ 也可以来自其它随机分布, 但需要相互独立, 比如正态分布.

## 3.2 模拟研究及结果

在这一节里, 我们对于 3.1 节中所提出的方法进行模拟研究. 同样考虑高

维线性回归模型

$$Y = \beta^T X + \varepsilon, \tag{3.2.1}$$

我们将观测值表示为 $\{(X_i, Y_i), i = 1, \ldots, n\}$. 其中 $\{X_i\}$ 为服从 $N(\mathbf{0}, \Sigma_N)$ 的独立同分布的随机变量, 这里 $\Sigma_N = (0.5^{(|i-j|)})_{1\le i,j\le p}$, 回归系数 $\beta = \beta_0 + \zeta, \beta_0 \in R^p, \zeta_i \sim N(0, \frac{3}{n^3})$, 且 $\varepsilon_i \sim t(8), i = 1, \ldots, n$. 需要指出的是 $\zeta$ 的不同取值代表 $\beta$ 与 $\beta_0$ 之间的不同关系, 并不仅仅局限于某一种分布. 我们仅是以此作为例子来检验 KEL 方法的效果. 下面考虑假设检验问题 $H_0 : \beta = \beta_0 \longleftrightarrow H_1 : \beta \neq \beta_0$ (显著性水平 $\alpha$ 为0.05). HEL 和 PEL 方法下犯第二类错误的概率见表 3-1.

表 3-1 HEL 和 PEL 方法下犯第二类错误的概率 ($\alpha = 0.05$).

| p | HEL | | PEL | |
|---|---|---|---|---|
| | n=200 | n=1000 | n=200 | n=1000 |
| 5 | 0.1976 | 0.1357 | 0.1727 | 0.1224 |
| 10 | 0.2334 | 0.1410 | 0.1594 | 0.1301 |
| 20 | 0.2967 | 0.1639 | 0.1847 | 0.1249 |
| 30 | 0.3453 | 0.1878 | 0.2222 | 0.1432 |
| 40 | 0.3832 | 0.1953 | 0.2699 | 0.1498 |
| 50 | 0.4206 | 0.2172 | 0.3139 | 0.1572 |
| 60 | 0.4492 | 0.2350 | 0.3598 | 0.1611 |
| 70 | 0.4723 | 0.2536 | 0.3943 | 0.1702 |
| 80 | 0.5059 | 0.2589 | 0.4084 | 0.1779 |
| 90 | 0.5204 | 0.2699 | 0.4396 | 0.1854 |
| 100 | 0.5503 | 0.2906 | 0.4659 | 0.2001 |

这里需要说明的是: 第二章中所有的表格都呈现功效函数值是为了与文献 [35] 中的做法保持一致, 便于对比. 而犯第二类错误的概率等于用 1 减去相应的功效函数, 在这一章里为了结果的直观性, 我们呈现的是犯第二类错误的概率而非前一章的功效函数.

由于原假设依然是 $H_0 : \beta = \beta_0$, 所以 $\zeta$ 的不同选取基本上不会影响HEL和PEL方法犯第一类错误的概率, 这由前面的定理 3 可以保证, 但从表

3-1 中不难发现, 无论是 PEL 方法还是我们提出的 HEL 方法犯第二类错误的概率都比较大, 且都随着维数递增. 下面针对同一模型, 我们做相关模拟研究以验证 KEL 方法的效用.

这一部分, 我们分别取 $k = 2, 3, 4, 5$, 看对于模型 (3.2.1) 犯第二类错误的概率 ($II$) 是否有一定的改进作用. 同时, 我们也给出了其犯一类错误的概率 ($I$), 以便从整体上分析 KEL 方法的效果.

表 3-2 k=2 和 3, 样本量 n=200 (1000) 时, 犯两类错误的概率 ($\alpha = 0.05$).

| p | k=2 | | k=3 | |
|---|---|---|---|---|
| | $I$ | $II$ | $I$ | $II$ |
| 5 | 0.0654(0.0554) | 0.0746(0.0341) | 0.0677(0.0544) | 0.0273(0.0071) |
| 10 | 0.0601(0.0541) | 0.0921(0.0365) | 0.0628(0.0492) | 0.0351(0.0099) |
| 20 | 0.0589(0.0527) | 0.1428(0.0450) | 0.0645(0.0520) | 0.0639(0.0119) |
| 30 | 0.0620(0.0493) | 0.1873(0.0607) | 0.0636(0.0519) | 0.0964(0.0153) |
| 40 | 0.0564(0.0540) | 0.2373(0.0701) | 0.0665(0.0483) | 0.1408(0.0234) |
| 50 | 0.0575(0.0515) | 0.2680(0.0769) | 0.0659(0.0549) | 0.1733(0.0264) |
| 60 | 0.0629(0.0527) | 0.2984(0.0914) | 0.0654(0.0578) | 0.2022(0.0349) |
| 70 | 0.0588(0.0524) | 0.3352(0.0978) | 0.0588(0.0524) | 0.2386(0.0445) |
| 80 | 0.0606(0.0531) | 0.3684(0.1116) | 0.0648(0.0511) | 0.2657(0.0446) |
| 90 | 0.0576(0.0529) | 0.3902(0.1199) | 0.0698(0.0502) | 0.3061(0.0484) |
| 100 | 0.0641(0.0524) | 0.4195(0.1370) | 0.0677(0.0537) | 0.3204(0.0615) |

表 3-2 是加入的约束条件为两个和三个的情况. 从表中可以看出: 犯第二类错误的概率随着维数的增加而变大. 当 $n = 200$ 时, 对于 $k = 2$ 的情况, 在 $p = 5$ 和 $p = 10$ 的时候, 犯第二类错误的概率低于 0.1, 相对来说令人满意, 而当 $p$ 继续变大, 其犯第二类错误的概率也显得比较大, 特别是当 $p = 50$ 及其以上, 其犯第二类错误的概率已不在可接受范围之内; 对于 $k = 3$ 的情况, 在 $p = 5, 10, 20, 30$ 的时候, 犯第二类错误的概率也低于 0.1, 当 $p$ 继续变大, 特别是当 $p = 60$ 及其以上, 其犯第二类错误的概率已大于 0.2, 不在可接受范围之内. 当 $n = 1000$ 时, 除了对于 $k = 2$ 的情况下少数高的维数, 其他犯第二类错误的概率都低于 0.1, 比较令人满意. 所以在 KEL 的方法中加入三个约束条件, 对于大样本的情形, 犯第二类错误的概率已可以做到基本令人满意.

表 3-3 k=4 和 5, 样本量 n=200 (1000) 时,犯两类错误的概率 ($\alpha = 0.05$).

| p | k=4 | | k=5 | |
|---|---|---|---|---|
| | $I$ | $II$ | $I$ | $II$ |
| 5 | 0.0728(0.0534) | 0.0119(0.0049) | 0.0694(0.0504) | 0.0170(0.0069) |
| 10 | 0.0737(0.0565) | 0.0181(0.0034) | 0.0699(0.0489) | 0.0339(0.0202) |
| 20 | 0.0754(0.0530) | 0.0324(0.0035) | 0.0721(0.0561) | 0.0252(0.0081) |
| 30 | 0.0742(0.0516) | 0.0505(0.0048) | 0.0777(0.0535) | 0.0403(0.0076) |
| 40 | 0.0740(0.0543) | 0.0791(0.0069) | 0.0779(0.0513) | 0.0520(0.0080) |
| 50 | 0.0716(0.0491) | 0.1062(0.0115) | 0.0750(0.0541) | 0.0704(0.0069) |
| 60 | 0.0730(0.0556) | 0.1328(0.0137) | 0.0799(0.0526) | 0.0994(0.0087) |
| 70 | 0.0738(0.0571) | 0.1669(0.0158) | 0.0747(0.0553) | 0.1192(0.0086) |
| 80 | 0.0725(0.0552) | 0.1847(0.0200) | 0.0770(0.0514) | 0.1471(0.0116) |
| 90 | 0.0737(0.0526) | 0.2129(0.0260) | 0.0777(0.0528) | 0.1718(0.0126) |
| 100 | 0.0766(0.0576) | 0.2496(0.0270) | 0.0765(0.0522) | 0.1903(0.0138) |

表 3-3 是加入的约束条件为四个和五个的情况. 与表 3-2 类似, 犯第二类错误的概率也是随着维数的增加而变大. 当 $n = 200$ 时, 对于 $k = 4$ 的情况, 在 $p = 5$ 到 $p = 70$, 犯第二类错误的概率最高为 0.1669, 对于维数相对较高的情况, 这样的结果令人满意; 对于 $k = 5$ 的情况, 对所有的 $p$, 犯第二类错误的概率也都低于 0.2, 且是在维数较高, 即 $p = 70, 80, 90, 100$ 时, 犯第二类错误的概率才开始大于 0.1. 这与表 3-1 中的值相比较, 降幅明显. 当样本量 $n = 1000$ 时, 无论是 $k = 4$ 还是 $k = 5$, 结果都令人满意.

概括而言: 对于更一般的假设检验问题, KEL 方法显然要优于 PEL 和 HEL 方法, 且随着 $k$ 的变大, KEL 方法犯第二类错误的概率有明显的减小, 特别是当样本量 $n = 1000$ 时, KEL 方法无论是犯第一类错误的概率还是犯第二类错误的概率都令人满意. 但是, $k$ 的变大意味着约束条件的增加, 这会大大降低运算效率, 增加运算成本. 进一步分析表 3-2 和表 3-3, 我们发现当 $p$ 和 $k$ 呈现如下表关系时, 犯两类错误的概率和运算效率都可以得到好的保证.

表 3-4 $p$ 和 $k$ 的对应关系

| p | (5,10) | (20,30,40) | (50,60,70) | (80,90,100) |
|---|---|---|---|---|
| k | 2 | 3 | 4 | 5 |

**注 6.** 表 3-4 是通过很多额外的模拟所得出的, 表里直观地呈现了对于不同的 $p, k$ 所对应的取值. 当 $k$ 与 $p$ 为这样的对应关系时, 本章所提方法的模拟结果是令人满意的. 当然, 取更大的 $k$ 会得到更好的效果, 但同时也会增加计算时间和计算难度, 这就需要我们具体问题具体分析. 从表 3-5 可以看出, 对

表 3-5 $p$ 和 $k$ 如表 3-4 对应后, 犯两类错误的概率 ($\alpha = 0.05$).

| p | k | $I$ | | $II$ | |
|---|---|---|---|---|---|
| | | n=200 | n=1000 | n=200 | n=1000 |
| 5 | 2 | 0.0654 | 0.0554 | 0.0746 | 0.0341 |
| 10 | 2 | 0.0601 | 0.0541 | 0.0921 | 0.0365 |
| 20 | 3 | 0.0645 | 0.0520 | 0.0639 | 0.0119 |
| 30 | 3 | 0.0636 | 0.0519 | 0.0964 | 0.0153 |
| 40 | 3 | 0.0665 | 0.0483 | 0.1408 | 0.0234 |
| 50 | 4 | 0.0716 | 0.0491 | 0.1062 | 0.0115 |
| 60 | 4 | 0.0730 | 0.0556 | 0.1328 | 0.0137 |
| 70 | 4 | 0.0738 | 0.0571 | 0.1669 | 0.0158 |
| 80 | 5 | 0.0770 | 0.0514 | 0.1471 | 0.0116 |
| 90 | 5 | 0.0777 | 0.0528 | 0.1718 | 0.0126 |
| 100 | 5 | 0.0765 | 0.0522 | 0.1903 | 0.0138 |

于 $\beta$ 与 $\beta_0$ 的每个分量变化不一致的情况, 当样本量较小时 ($n = 200$), KEL 方法即能保证犯第一类错误的概率接近于显著性水平 0.05, 也能将犯第二类错误的概率控制在一个可接受范围内; 当样本量足够大 ($n = 1000$) 时, 该方法能大大降低犯第二类错误的概率, 确保一个令人满意的效果. 同时, 运算效率也能得到保证.

另外, 通过前面的讨论我们知道, 在经验似然方法里加入调整因子后, 都可以得到比之前方法更好的区间覆盖率, 所以这里我们不再呈现调整的 KEL 方法的模拟结果.

# 第 4 章 相关证明

为了书写方便, 我们在证明定理 3 和定理 4 的过程中采用下列记号:

$(N_1)$ $Z_n^* = \max_{1\le i\le n} |\omega_i(\beta_0)|$;

$(N_2)$ $S = \frac{1}{n}\sum_{i=1}^n \omega_i^2(\beta_0)$;

$(N_3)$ $\bar{\omega}_n(\beta_0) = \frac{1}{n}\sum_{i=1}^n \omega_i(\beta_0)$.

我们首先给出下列引理.

**引理 1** 在定理 3 的假设下, 我们有

$$Var(\omega_i(\beta_0)) < \infty; \tag{4.0.1}$$

$$Z_n^* = o_p(n^{\frac{1}{2}}); \tag{4.0.2}$$

$$S = O_p(1); \tag{4.0.3}$$

$$\bar{\omega}_n(\beta_0) = O_p(n^{-\frac{1}{2}}). \tag{4.0.4}$$

**证明** 因为 $X_i$ 和 $\varepsilon_i$ 相互独立, 易知

$$\begin{aligned} Var(\omega_i(\beta_0)) &= Var(\varepsilon_i X_i^T \mathbf{1}_p) \\ &= E\varepsilon_i^2 \cdot E(\mathbf{1}_p^T X_i X_i^T \mathbf{1}_p) \\ &\le p\lambda_p \sigma^2. \end{aligned}$$

其中 $\sigma^2 > 0$ 是 $\varepsilon_i$ 的方差, 且 $0 < \lambda_p < \infty$ 是 $E(X_i X_i^T)$ 的最大的特征值. 因此, (4.0.1) 得证. 再由 (4.0.1) 和文献 [17] 的引理 11.2 , 可知 (4.0.2) 成立. 根据大数定律和中心极限定理, 可知 (4.0.3) 和 (4.0.4) 显然成立.

**定理 3 的证明** 注意到 (2.1.2) 式, 我们有

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n}\frac{\omega_i(\beta_0)[1+\lambda\omega_i(\beta_0)]-\lambda\omega_i^2(\beta_0)}{1+\lambda\omega_i(\beta_0)} \\ &= \frac{1}{n}\sum_{i=1}^{n}\omega_i(\beta_0)-\frac{\lambda}{n}\sum_{i=1}^{n}\frac{\omega_i^2(\beta_0)}{1+\lambda\omega_i(\beta_0)}. \end{aligned}$$

又由 (2.1.1) 式知 $1+\lambda\omega_i(\beta_0)>0$ 成立. 通过 (4.0.4) 式及上式可知

$$0 \le O_p(n^{-\frac{1}{2}})-\frac{\lambda}{n}\sum_{i=1}^{n}\frac{\omega_i^2(\beta_0)}{1+\lambda Z_n^*} = O_p(n^{-\frac{1}{2}})-\frac{\lambda S}{1+\lambda Z_n^*}.$$

再结合引理 1 中的 (4.0.2) 式和 (4.0.3) 式, 可得

$$\lambda = O_p(n^{-\frac{1}{2}}). \tag{4.0.5}$$

于是, 进一步可得到

$$\max_{1\le i\le n}|\lambda\omega_i(\beta_0)| = O_p(n^{-\frac{1}{2}})\cdot o_p(n^{\frac{1}{2}}) = o_p(1). \tag{4.0.6}$$

另一方面, (2.1.2) 式可变形为

$$0 = \frac{1}{n}\sum_{i=1}^{n}\omega_i(\beta_0)-\frac{\lambda}{n}\sum_{i=1}^{n}\omega_i^2(\beta_0)+\frac{\lambda^2}{n}\sum_{i=1}^{n}\frac{\omega_i^3(\beta_0)}{1+\lambda\omega_i(\beta_0)}. \tag{4.0.7}$$

通过 (4.0.2), (4.0.3), (4.0.5) 和 (4.0.6) 式, 可知上式右边的第三项有界, 即

$$\frac{\lambda^2}{n}\sum_{i=1}^{n}|\omega_i(\beta_0)|^3|1+\lambda\omega_i(\beta_0)|^{-1} = O_p(n^{-1})\cdot o_p(n^{\frac{1}{2}})\cdot O_p(1) = o_p(n^{-\frac{1}{2}}).$$

因此 (4.0.7) 式又可写成

$$0 = \bar{\omega}_n(\beta_0)-\lambda S+o_p(n^{-\frac{1}{2}}).$$

所以, 我们有

$$\lambda = S^{-1}\bar{\omega}_n(\beta_0) + o_p(n^{-\frac{1}{2}}). \tag{4.0.8}$$

通过 (4.0.3) 式, (4.0.5) 式, (4.0.6) 式和泰勒展式, 现在我们可以记

$$\begin{aligned} -2\log L_n^H(\beta_0) &= 2\sum_{i=1}^{n}\log(1+\lambda\omega_i(\beta_0)) \\ &= 2\sum_{i=1}^{n}\lambda\omega_i(\beta_0) - \sum_{i=1}^{n}\lambda^2\omega_i^2(\beta_0) + o_p(1). \end{aligned}$$

再结合 (4.0.8) 式, 可得

$$-2\log L_n^H(\beta_0) = n\bar{\omega}_n^2(\beta_0)S^{-1} + o_p(1).$$

通过中心极限定理, 易知当 $n \to \infty$ 时, $-2\log L_n^H(\beta_0) \xrightarrow{d} \chi_1^2$ . 这就证明了定理 3.

**定理 4 的证明** 定理4的证明大部分与定理 3 类似, 所以这里我们只写出其中关键的步骤而省略掉其他推导.

(2.1.4) 式两边同乘 $\frac{n+1}{n}\lambda$, 得

$$\begin{aligned} 0 &= \frac{\lambda}{n}\sum_{i=1}^{n+1}\frac{\omega_i(\beta_0)}{1+\lambda\omega_i(\beta_0)} \\ &= \frac{\lambda}{n}\sum_{i=1}^{n+1}\omega_i(\beta_0) - \frac{\lambda^2}{n}\sum_{i=1}^{n+1}\frac{\omega_i^2(\beta_0)}{1+\lambda\omega_i(\beta_0)} \\ &\le \lambda\bar{\omega}_n(\beta_0)(1-a_n/n) - \frac{\lambda^2}{n}\sum_{i=1}^{n}\frac{\omega_i^2(\beta_0)}{1+\lambda\omega_i(\beta_0)}. \end{aligned}$$

因为 $1+\lambda\omega_i(\beta_0)>0$, 于是可得

$$\frac{\lambda^2}{n}\sum_{i=1}^{n}\frac{\omega_i^2(\beta_0)}{1+|\lambda|Z_n^*}\leq\lambda\bar{\omega}_n(\beta_0)(1-a_n/n). \tag{4.0.9}$$

因为 $a_n=o_p(n)$, 同时将引理 1 中的 (4.0.2) 式、(4.0.3) 式、(4.0.4) 式代入 (4.0.9) 式, 可得

$$O_p(1)\frac{\lambda^2}{1+|\lambda|o_p(n^{\frac{1}{2}})} \leq \lambda O_p(n^{-\frac{1}{2}}),$$

上式即表明 $\lambda=O_p(n^{-\frac{1}{2}})$.

另一方面，由于

$$\begin{aligned}\frac{1}{n}\frac{\omega_{n+1}(\beta_0)}{1+\lambda\omega_{n+1}(\beta_0)} &= -\frac{1}{n}\frac{a_n\bar{\omega}_n(\beta_0)}{1-\lambda a_n\bar{\omega}_n(\beta_0)}\\ &= \frac{1}{n}\frac{o_p(n)O_p(n^{-\frac{1}{2}})}{1-O_p(n^{-\frac{1}{2}})o_p(n)O_p(n^{-\frac{1}{2}})}\\ &= o_p(n^{-\frac{1}{2}}),\end{aligned}$$

(2.1.4) 式可被记为

$$0 = \frac{1}{n}\sum_{i=1}^{n}\frac{\omega_i(\beta_0)}{1+\lambda\omega_i(\beta_0)}+o_p(n^{-\frac{1}{2}}).$$

通过与定理 3 中相同的证明方法, 我们有

$$\lambda=S^{-1}\bar{\omega}_n(\beta_0)+o_p(n^{-\frac{1}{2}}).$$

最后，我们注意到

$$\log\{1+\lambda\omega_{n+1}(\beta_0)\}=\log\{1+o_p(1)\}.$$

因此 (2.1.5) 式可被记为

$$-2\log L_n^{AH}(\beta_0) = 2\sum_{i=1}^{n}\log\{1+\lambda\omega_i(\beta_0)\} + o_p(1).$$

再结合定理 3 的证明, 我们有 $-2\log L_n^{AH}(\beta_0) \xrightarrow{d} \chi_1^2$, 当 $n \to \infty$ 时. 于是, 这就证明了定理 4.

下面我们给出定理 5 的证明. 在此之前, 首先证明下列引理. 为书写方便, 下面 $U_i$ 即表示 $U_i(\beta)$, 其中 $U_i(\beta)$ 为由 $\omega_i^j(\beta)$, $j = 1, 2, \cdots, k$, 构成的 $k$ 行 1 列的列向量.

**引理 2** 记 $S^* = \frac{1}{n}\sum_{i=1}^{n} U_i U_i^T$. 若 $\alpha_1, \cdots, \alpha_k$ 相互独立, 则 $R(S^*) = k$.

**证明** $S^*$ 可以展开来写成下式

$$\begin{aligned} S^* &= \frac{1}{n}\sum_{i=1}^{n} U_i U_i^T = \frac{1}{n}\sum_{i=1}^{n}(\alpha_1^T, \cdots, \alpha_k^T)^T X_i (Y_i - \beta^T X_i)^2 X_i^T (\alpha_1, \cdots, \alpha_k) \\ &= (\alpha_1^T, \cdots, \alpha_k^T)^T \frac{1}{n}\sum_{i=1}^{n}(Y_i - \beta^T X_i)^2 X_i X_i^T (\alpha_1, \cdots, \alpha_k), \end{aligned}$$

再由大数定律, 可知

$$\begin{aligned} S^* &= (\alpha_1, \cdots, \alpha_k)^T E\left[(Y_i - \beta^T X_i)^2 X_i X_i^T\right](\alpha_1, \cdots, \alpha_k) \\ &= \sigma^2(\alpha_1^T, \cdots, \alpha_k^T)^T E(X_i X_i^T)(\alpha_1, \cdots, \alpha_k). \end{aligned}$$

设$A = E(X_i X_i^T)$, 由定理 5 中给出的条件知 $A$ 正定, 于是存在可逆阵 $P$, 使得 $A = PP^T$, 所以

$$\begin{aligned} R(S^*) &= R\left[(\alpha_1^T, \cdots, \alpha_k^T)^T PP^T(\alpha_1, \cdots, \alpha_k)\right] \\ &= R\left[(\alpha_1^T, \cdots, \alpha_k^T)^T P\right] \end{aligned}$$

$$= R(\alpha_1, \cdots, \alpha_k)$$

$$= k.$$

这便证明了引理 2.

**定理 5 的证明** 假设 $0 < \lambda_1 \leq \cdots \leq \lambda_k < \infty$ 为 $Var(U_i)$ 的特征根. 令 $\lambda = \|\lambda\|\theta$, 这里 $\theta \in \Theta$ 是一个单位向量. 令 $Y_i^* = \lambda^T U_i$; $U_n^* = \max_{1\leq i\leq n} \|U_i\|$.

(3.1.2) 式两边同乘 $\theta^T$ 得

$$\frac{1}{n}\sum_{i=1}^{n} \theta^T U_i \frac{1}{1+\lambda^T U_i} = \frac{1}{n}\sum_{i=1}^{n} \theta^T U_i (1 - \frac{\lambda^T U_i}{1+\lambda^T U_i}) = 0,$$

化简后可得

$$\theta^T \bar{U} = \|\lambda\| \theta^T \tilde{S} \theta,$$

其中 $\tilde{S} = \frac{1}{n}\sum_{i=1}^{n} \frac{U_i U_i^T}{1+\lambda^T U_i}$, $\bar{U} = \frac{1}{n}\sum_{i=1}^{n} U_i$.

因为任意 $q_i \geq 0$, 所以 $1 + \lambda^T U_i > 0$, 所以

$$\begin{aligned} \|\lambda\|\theta^T S^* \theta &\leq \|\lambda\|\theta^T \tilde{S}\theta (1 + \max_{i=1}^{n} |Y_i^*|) \\ &\leq \|\lambda\|\theta^T \tilde{S}\theta (1 + \|\lambda\| U_n^*) \\ &= \theta^T \bar{U} (1 + \|\lambda\| U_n^*) \end{aligned}$$

于是 $\|\lambda\|(\theta^T S^* \theta - U_n^* \theta^T \bar{U}) \leq \theta^T \bar{U}$. 因为 $0 < \lambda_1 \leq \cdots \leq \lambda_k < \infty$为$Var(U_i)$的特征根, 所以 $\lambda_1 + o_p(1) \leq \theta^T S^* \theta \leq \lambda_k + o_p(1)$, 同时也由文献 [17] 中的引理 11.2 知 $U_n^* = o_p(n^{\frac{1}{2}})$. 又由中心极限定理可证 $\theta^T \bar{U} = O_p(n^{-\frac{1}{2}})$. 所以 $\|\lambda\|(\theta^T S^* \theta + o_p(1)) = O_p(n^{-\frac{1}{2}})$. 所以 $\|\lambda\| = O_p(n^{-\frac{1}{2}})$. 所以 $\max_{i=1}^{n} |Y_i^*| =$

$O_p(n^{-\frac{1}{2}})o_p(n^{\frac{1}{2}}) = o_p(1)$. 再由 (3.1.2) 式, 得

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n} U_i(1 - Y_i^* + \frac{Y_i^{*2}}{1+Y_i^*}) \\ &= \bar{U} - \frac{1}{n}\sum_{i=1}^{n} U_i\lambda^T U_i + \frac{1}{n}\sum_{i=1}^{n}\frac{U_i Y_i^{*2}}{1+Y_i^*} \\ &= \bar{U} - S^*\lambda + \frac{1}{n}\sum_{i=1}^{n}\frac{U_i Y_i^{*2}}{1+Y_i^*}, \end{aligned}$$

又 $\frac{1}{n}\sum_{i=1}^{n}\|U_i\|^3 \cdot \|\lambda\|^2 \cdot |1+Y_i^*|^{-1} = o_p(n^{\frac{1}{2}}) \cdot O_p(n^{-1}) \cdot O_p(1) = o_p(n^{-\frac{1}{2}})$, 所以

$$\lambda = S^{*-1} \cdot \bar{U} + o_p(n^{-\frac{1}{2}}). \tag{4.0.10}$$

对 (3.1.3) 式作泰勒展开, 再将 (4.0.10) 式代入, 可得

$$\begin{aligned} -2\log L_n^K(\beta) &= 2\sum_{i=1}^{n} Y_i^* - \sum_{i=1}^{n} Y_i^{*2} + o_p(1) \\ &= 2\sum_{i=1}^{n}\lambda^T U_i - \sum_{i=1}^{n}\lambda^T U_i U_i^T\lambda + o_p(1) \\ &= 2n\lambda^T\bar{U} - n\lambda^T S^*\lambda + o_p(1) \\ &= n\bar{U}^T S^{*-1}\bar{U} + o_p(1). \end{aligned}$$

当 $n \to \infty$ 时, $n\bar{U}^T S^{*-1}\bar{U} \to \chi_k^2$, 所以 $-2\log L_n^K(\beta_0) \xrightarrow{d} \chi_k^2$. 这就证明了定理 5.

# 第 5 章 总结与展望

对于高维线性回归模型

$$Y_i = \beta^T X_i + \varepsilon_i, i = 1, \ldots, n, \tag{5.0.1}$$

其中 $\{X_i, i = 1, \ldots, n\}$ 是独立同分布 (i.i.d.) 的随机向量, $\beta = (\beta_1, \ldots, \beta_p)^T \in R^p, p \geq 1$, 是一列未知系数向量, $\{\varepsilon_i, i = 1, \ldots, n\}$ 是均值为 0, 方差为 $\sigma^2 > 0$ 的 i.i.d. 随机变量. 同时我们假设 $\{X_i\}$ 和 $\{\varepsilon_i\}$ 相互独立. 本文考虑了该模型的假设检验问题: $H_0 : \beta = \beta_0 \longleftrightarrow H_1 : \beta \neq \beta_0$. 我们尝试着把传统的经验似然方法里 (1.3.2) 与 $p$ 有关的高维约束条件巧妙地变换成一维情形, 即把 $(Y_i - \beta^T X_i)X_i^T$ 的所有分量相加, 以此构造出 $\omega_i(\beta), i = 1, \ldots, n$. 而非前人分组的方法. 由此我们得到了一个新的且简单的经验似然方法来解决该假设检验问题. 不仅如此, 我们还在 HEL 方法和 PEL 方法里各自加入了伪观测值, 从而作出了一个新奇的调整. 调整后的经验似然方法保留了之前方法的所有最优性准则, 且该方法下的区间覆盖率更接近于显著性水平, 而且还不需要 Bartlett 校正和 bootstrap 方法里那么复杂的程序. 模拟研究表明: 对于 $\beta$ 和 $\beta_0$ 的每一个分量变化都是一致的情况, 我们所提出的方法无论是从犯第一类错误的概率还是从功效函数的角度, 都有很好的检验效果, 且都优于 PEL 方法.

另外, 我们还考虑了 $\beta$ 和 $\beta_0$ 分量之间可能存在的更复杂的关系的情况. 我们同样运用经验似然的方法, 将 $(Y_i - \beta^T X_i)X_i^T$ 的分量加权相加, 即将原来 $(Y_i - \beta^T X_i)X_i^T$ 后面所乘的 $\mathbf{1}_p$ 换为其他 $p$ 维向量, 且该向量的分量不能都相等, 构造

$$\omega_i^k(\beta) = (Y_i - \beta^T X_i)X_i^T \alpha_k,$$

其中 $k = 1, 2, 3, 4, 5$, $\alpha_k$ 相互独立. 针对不同的维数, 有区别地在经验似然方法里加入了估计式的个数, 一方面很好的解决了假设检验问题, 另一方面也大大地节省了计算成本, 提高了计算效率.

除了以上工作, 我们还可以考虑将 HEL 方法, AHEL 方法, KEL 方法运用到高维删失数据当中, 结合 B-J 估计, K-M 估计等尝试解决高维删失数据下的假设检验问题, 这也将是今后我们努力和挑战的方向.

# 参考文献

[1] D.R. Thomas, G.L. Grunkemeier, 1975. Confidence Interval Estimation of Survival Probabilities for Censored Data. Journal of the American Statistical Association. 70: 865-871.

[2] A. Owen, 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika. 75: 237-249.

[3] A. Owen, 1991. Empirical likelihood for linear models. The Annals of Statistics. 19: 1725-1747.

[4] A. Owen, 1990. Empirical likelihood ratio confidence regions. The Annals of Statistics. 18: 90-120.

[5] M. Tsao, 2004. Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. The Annals of Statistics. 32: 1215-1221.

[6] W. Lu, Y. Liang, 2006. Empirical likelihood inference for linear transformation models. Journal of Multivariate Analysis. 97: 1586-1599.

[7] P. McCullagh, J.A. Nelder, 1983. Generalized linear models. London: Chapman and Hall.

[8] B. Zhang, 1997. Quantile processes in the presence of auxiliary information. Annals of the Institute of Statistical Mathematics. 49: 35-55.

[9] S.X. Chen, 1994. Empirical likelihood ratio confidence intervals for linear regression coefficients. Journal of Multivariate Analysis. 49: 24-40.

[10] J.H. Chen, J. Qin, 1993. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. Biometrika. 80: 107-116.

[11] S. Chen, P. Hall, 1993. Smoothed empirical likelihood confidence intervals for quantiles. The Annals of Statistics. 21: 1166-1181.

[12] W. Zhou, B.Y. Jing, 2003. Adjusted empirical likelihood method for quantiles. Annals of the Institute of Statistical Mathematics. 55: 689-703.

[13] L. Peng, 2004. Empirical-likelihood-based confidence interval for the mean with a heavy-tailed distribution. The Annals of Statistics. 32: 1192-1214.

[14] Q. Wang, B. Jing, 1999. Empirical likelihood for partial linear model with fixed designs. Statistics and Probability Letters. 41: 425-433.

[15] B. Zhong, J.N.K. Rao, 2000. Empirical likelihood inference under stratified random sampling using auxiliary population information. Biometrika. 87: 929-938.

[16] Y. Kitamura, 2001. Asymptotic optimality of empirical likelihood for testing moment restrictions. Econometrica. 69: 1661-1672.

[17] A. Owen, 2001. Empirical Likelihood. Chapman&Hall/CRC.

[18] G. Li, Q.H. Wang, 2003. Empirical likelihood regression analysis for right censored data. Statistica Sinica. 13: 51-68.

[19] M. Zhou, G. Li, 2008. Empirical likelihood analysis of the Buckley-James estimator. Journal of Multivariate Analysis. 99: 649-664.

[20] P.S. Zhong, H.J. Cui, 2010. Empirical likelihood for median regression model with designed censoring variables. Journal of Multivariate Analysis. 101: 240-251.

[21] J.B. Li, Z.S. Huang, H. Lian, 2014. Empirical likelihood inference for general transformation models with right censored data. Statistics and Computing. 24: 985-995.

[22] J. Chen, A.M. Variyath, B. Abraham, 2008. Adjusted empirical likelihood and its properties. Journal of Computational and Graphical Statistics. 17: 426-443.

[23] J. Chen, R.R. Sitter, C. Wu, 2002. Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. Biometrika. 89: 230-237.

[24] S.X. Chen, L. Peng, Y.L. Qin, 2009. Effects of data dimension on empirical likelihood. Biometrika. 96: 1-12.

[25] N.L. Hjort, I.W. Mckeague, I.V. Keilegom, 2009. Extending the scop of empirical likelihood. The Annals of Statistics. 37: 1079-1111.

[26] C.Y. Tang, C. Leng, 2010. Penalized high-dimensional empirical likelihood. Biometrika. 97: 905-920.

[27] 黎子良, 郑祖康, 1993. 生存分析. 浙江科学技术出版社.

[28] S. Portnoy, 1984. Asymptotic behavior of $M$-estimators of p regression parameters when $p^2/n$ is large; $I$, consistency. The Annals of Statistics. 12: 1298-1309.

[29] S. Portnoy, 1985. Asymptotic behavior of $M$-estimators of p regression parameters when $p^2/n$ is large; $II$, normal approximation. The Annals of Statistics. 13: 1403-1417.

[30] J. Bradic, J. Fan, W. Wang, 2011. Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. Journal of the Royal Statistical Society Series B. 73: 325-349.

[31] R. Tibshirani, 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B. 58: 267-288.

[32] H. Zou, 2006. The adaptive Lasso and its oracle properties. Journal of the American Statistical Association. 101: 1418-1429.

[33] N. Meinshausen, 2007. Relaxed lasso. Computational Statistics and Data Analysis. 52: 374-393.

[34] N. Meinshausen, L. Meier, P. Buhlmann, 2009. P-values for high-dimensional regression. Journal of the American Statistical Association. 104: 1671-1681.

[35] L. Peng, Y.C. Qi, R.D. Wang, 2014. Empirical likelihood test for high dimensional linear models. Statistics and Probability Letters. 86: 85-90.

[36] 王大荣, 2009. 分散度量模型中的变量选择. 北京工业大学博士学位论文.

[37] 石坚, 2007. 高维线性模型中的经验似然. 系统科学与数学. 27:124-133.

[38] Y. Liu, C. Zou, Z. Wang, 2013. Calibration of the empirical likelihood for high-dimensional data. Annals of the Institute of Statistical Mathematics. 69: 529-550.

[39] H. Guo, C.L. Zou, Z.J. Wang, 2014. Empirical likelihood for high-dimensional linear regression models. Metrika. 77: 921-945.

[40] M. Zheng, W. Yu, 2012. Empirical likelihood method for multivariate cox regression. Computational Statistics. 3: 1241-1267.

[41] W. Yu, H.E. Barmi, Z.L. Ying, 2011. Restricted one way analysis of variance using the empirical likelihood ratio test. Journal of Multivariate Analysis. 102: 629-640.

[42] A.M. Variyath, J. Chen, B. Abraham, 2010. Empirical likelihood based variable selection. Journal of Statistical Planning and Inference. 140: 971-981.

# 致 谢

研究生生活转瞬即逝,回首往昔,想起那些在我求知道路上给予我极大帮助的老师、同学和亲人,感慨万千,心中满是感恩.

本文是在我的导师周秀轻老师的悉心指导和亲切关怀下完成的,她严谨的治学精神,严肃的科学态度,精益求精的工作作风深深地感染和激励着我,周秀轻老师不仅在学习上给我悉心的指导,在生活上亦给予我无微不至的关怀,在此谨向周秀轻老师表示诚挚的谢意和崇高的敬意.

本文最终得以顺利完成,也是与数学科学学院其他老师的帮助分不开的.在此衷心感谢刘国祥老师、解锋昌老师、朱全新老师、梁志彬老师、杜秀丽老师、王晓谦老师以及高启兵老师三年来对我学习上的悉心指导,他们的谆谆教诲使我终生难忘.

我还要感谢我的同学刑雅琴、康大林、陈高燕、叶鹏、任盈盈等,他们在学习和生活上给予我很多关心和帮助.

此外,我还要感谢学校及院系的领导、老师们为我们三年的研究生学习生活所提供的各种保障和便利.

最后,我要特别感谢日夜辛劳,默默奉献的父母,你们的关心和支持为我的学习和生活提供了最坚强的后盾!

曾力立

江苏南京

2016年3月