湖南師範大學

博士学位论文

维数发散的高维数据的经验似然推断

Empirical Likelihood Inference for High-dimensional Data with a Diverging Number of Parameters

学 科 专 业 统计学

学 位 类 型 ☑科学学位 □专业学位

研 究 生 姓 名 方江林

导师姓名、职称 刘万荣 教授

论 文 编 号

湖南师范大学学位评定委员会办公室

二零一七年三月

分 类 号________ 密 级________

学校代码 10542 学号 201301100098

维数发散的高维数据的经验似然推断

Empirical Likelihood Inference for High-dimensional Data with a Diverging Number of Parameters

博 士 生 姓 名: 方 江 林

指导教师姓名、职称: 刘 万 荣 教 授

学 科 专 业: 统 计 学

研 究 方 向: 数 理 统 计

湖南师范大学学位评定委员会办公室

二零一七年三月

摘　要

在生物信息、医学、金融分析等诸多应用领域，经常出现高维数据，删失数据等复杂数据。数据维数不断增大给数据分析带来了重大挑战。一方面：维数的增大会导致“维数灾难”问题；另一方面：经典大样本统计推断理论一般都是建立在维数固定且相对较小，而样本量趋于无穷的假设下，在数据维数p随着样本容量n一起趋向无穷时，特别是在“超高维”$(p > n)$数据情形下，经典统计理论的结论可能不再有效。因此，如何对这些高维数据进行统计推断是统计学研究的一个重要课题。

经验似然方法是由Owen(1988)提出的一种非参数统计推断方法，与传统的正态逼近方法相比较，具有许多优势。例如：由经验似然方法所构造的参数的经验似然置信域不需要估计渐近方差，其形状完全由数据决定，而且还具有域保持性和变换不变性。本文在样本维数p随容量n一起趋向无穷情形下应用经验似然方法研究复杂数据的统计推断。另外，变量选择也是高维数据分析的一个重要问题。本文也研究了样本维数p随容量n一起趋向无穷情形下基于惩罚经验似然方法的半参数模型和可加危险率模型的变量选择和参数估计问题。

本文主要包括了以下几个方面的内容。

第二章研究了高维情形下半参数模型的统计推断问题。首先，利用经验似然方法构造了参数的估计量及其置信域。证明了在一定条件下，当样本维数p和容量n都趋向无穷情形时，经验似然比渐近分布为正态分布，并证明了通过经验似然方法得到的参数估计量具有一致性；其次，将惩罚经验似然方法推广到高维稀疏情形下半参数模型的变量选择和参数估计问题。证明了在一定条件下，当样本维数p发散，即样本维数p和容量n一起趋向无穷情形时，惩罚经验似然比统计量具有渐近 χ_q^2 分布，同时证明了惩罚经验似然方法具有Oracle性质。

第三章研究了高维删失情形下可加危险率模型的统计推断问题。首先，利用经验似然方法构造了参数的估计量及参数置信域或参数分量的置信区间(置信域)。证明了在一定条件下，当样本维数p发散时，通过经验似然方法得到的参数估计量具有一致性，并证明了关于参数和参数分量的经验似然比渐近分布分别为正态分布和 χ_q^2 分布；其次，将惩罚经验似然方法推广到高维稀疏删失情形下可加危险率模型的变量选择和参数估计问题。获得了在一定条件下，当样本维数p和容量n一起趋向无

穷情形时，惩罚经验似然统计量的渐近分布 – χ_q^2 分布，并证明了惩罚经验似然方法具有Oracle性质。

第四章研究了高维情形下异方差部分线性单指标模型的统计推断问题。利用经验似然方法构造了参数的估计量及参数置信域或参数分量的置信区间(置信域)。证明了在一定条件下，当样本维数p发散时，关于参数和参数分量的经验似然比渐近分布分别为正态分布和 χ_q^2 分布。

第五章研究了高维情形下两样本的统计推断问题。利用经验似然方法构造了两样本均值之差和两线性模型系数之差的估计量及其置信域。证明了在一定条件下，当样本维数p发散时，经验似然比渐近分布为正态分布，并证明了通过经验似然方法得到的参数估计量具有一致性。

我们通过模拟和实例分析验证了本文所提出的基于经验似然方法的高维数据分析理论结果及其优良性。

关键词: 经验似然；惩罚经验似然；高维数据；删失数据；半参数模型；可加危险率模型；部分线性单指标模型；变量选择。

ABSTRACT

In the biological information, health studies, financial analysis and so on, complicated data such as high dimensional data, censored data are often encountered. With the increase of dimension, data analysis becomes more and more difficult. On the one hand, the increase of the dimension will result in "Curse of Dimensionality"; On the other hand, the classical theory statistical inferences for large sample are generally based on the assumptions that the dimension is fixed and and relatively small and the sample size tends to infinite. When the dimension p tends to infinite with the sample size n, particularly in the "super high dimension" $(p > n)$ situation, the outcomes of the classical statistical theory may be no longer valid. Therefore, how to deal with these complicated data to derive statistical inference has become a hot research issue in statistics.

Empirical likelihood method proposed by Owen (1988) is a nonparametric statistical inference method. Compared with the traditional asymptotic normality method, empirical likelihood method has many advantages. For example, constructing confidence regions of parameters by using empirical likelihood method does not care about estimating asymptotic variance of parameters. The shape and orientation of confidence regions constructed by using empirical likelihood method are determined entirely by data, and also these confidence regions are range preserving and transformation respecting. In this thesis, we are mainly interesting in the statistical inference of complicated data, where dimensionality $p \to \infty$, as $n \to \infty$. In addition, variable selection is one of the hot issues of the high dimensional data analysis in statistics. In this thesis, we also study variable selection and parameter estimation of the semiparametric model and the additive hazards model by using penalized empirical likelihood method, where dimensionality $p \to \infty$, as $n \to \infty$.

The main contents in this thesis include the following several chapters.

The second chapter investigates the question of statistical inference for the high dimensional semiparametric model. Firstly, we construct estimators and confidence regions of the unknown parameters by using empirical likelihood method. With the diverging dimensionality, i.e., $p \to \infty$ as $n \to \infty$, we prove that, under some mild conditions, the asymptotic distribution of the empirical log-likelihood

ratio statistics for unknown parameters is an asymptotically normal distribution, and prove that the empirical likelihood estimator has the asymptotic consistent property. Secondly, with the situation of diverging dimensionality, a penalized empirical likelihood method for estimating parameters and variable selection for the semiparametric model is proposed. We prove that, under some regularity conditions, the penalized empirical log-likelihood ratio for the high dimensional sparse semiparametric model has an asymptotically Chi-square distribution, and prove that the penalized empirical likelihood estimator has the Oracle property.

The third chapter investigates the question of statistical inference for the high dimensional additive hazards model with censored data. We construct estimators of the unknown parameters, which has the asymptotic consistent property. An empirical log-likelihood ratio statistics for unknown parameters and an empirical log-likelihood ratio statistics for the component of the unknown parameters is proposed. It is proved that, with the situation of diverging dimensionality, the proposed statistics have the asymptotic normal distribution and the asymptotic Chi-square distribution under some mild conditions, respectively, which can be used to construct the confidence regions for unknown parameters or the confidence regions (intervals) for the component of the parameters. In addition, we propose a penalized empirical likelihood method for estimating parameters and variable selection for the high dimensional sparse additive hazards model with censored data. The proposed penalized empirical log-likelihood ratio for unknown parameters has the asymptotic Chi-square distribution under some mild conditions, and the penalized empirical likelihood estimator has the Oracle property.

The fourth chapter investigates the question of statistical inference for the high dimensional heteroscedastic partially linear single-index model. An empirical log-likelihood ratio statistics for unknown parameters and an empirical log-likelihood ratio statistics for the component of the unknown parameters is proposed. It is proved that, with the situation of diverging dimensionality, the proposed statistics have the asymptotic normal distribution and the asymptotic Chi-square distribution under some mild conditions, respectively, which can be used to construct the confidence regions for unknown parameters or the confidence regions (intervals) for the component of the parameters.

The fifth chapter investigates the question of statistical inference for grow-

ing dimensional two sample problems. We construct confidence regions for the difference of the means of two samples and the difference in value between coefficients of two sample linear model by using empirical likelihood method. Under some mild conditions, the proposed empirical log-likelihood ratio statistics has an asymptotically normal distribution, and the empirical likelihood estimator for the difference in value between coefficients of two sample linear model has the asymptotic consistent property.

The proposed methods in this thesis are illustrated with simulation studies and real data examples.

Key Words: Empirical likelihood; Penalized Empirical likelihood; High dimensional data; Censored data; Semiparametric models; Additive hazards models; Partially linear single-index models; Variable selection.

目　录

第一章 绪 论

本章阐述了论文研究背景、文献综述、相关知识以及主要研究内容与结构。

§1.1 研究背景

随着社会的发展和技术的进步，经济、金融、生物信息、医学、信息技术等各个领域每时每刻都在产生大量的数据，并且数据的复杂程度越来越大。如果用变量表示信息，那么数据信息量越来越大意味着其维数越来越大，例如：基因组数据、微阵数据、卫星提供的大量图像数据等等，其维数可能达到成千上万，甚至维数增加的速度比样本容量增加的速度快很多。这些用多个变量描述某一现象的数据，称为高维数据。随着数据维数的不断增加，数据提供关于某一现象的信息更加丰富和完整，但同时也会给数据分析和处理带来许多困难。在经典的大样本统计推断理论中，通常都假设数据维数相对较小且固定，而样本容量可趋于无穷。在高维数据情形下，特别是数据维数大于样本容量时，经典推断统计理论已经显出局限性，不能满足现实分析的需求。在不同的领域内，科学家们研究高维数据的目的可能各不相同，但他们会面临同样的问题：即怎么分析和处理这些高维数据？在“大数据”时代已经降临的今天，高维数据的出现给统计学者的研究带来了新的挑战，这迫切需要探索和发展新的理论和方法来处理高维数据问题。因此，高维数据分析理论—样本维数p随容量n同时趋向无穷假设下的统计推断理论的研究具有重要理论和实际意义。

在1937年，Wilks证明了参数模型的似然比统计量的渐近分布为标准的χ^2分布。Owen于1988年将Wilks的方法推广到非参数模型，同样证明了经验似然比统计量的渐近分布为标准的χ^2分布，该方法被称为经验似然方法。经验似然方法不需要对样本服从某种分布的假设，但仍然具有参数似然的优点(如：Wilks现象和Bartlett可修正性)。在构造置信域时比传统方法更具有优势，例如：在复杂数据(删失数据、缺失数据等)分析中，参数估计量的方差计算非常复杂，但经验似然方法不需要用到估计量的方差，这使得统计推断更加简捷而且结果更加精确。另外，经验似然方法还具有不需要对置信域的形状施加约束、不需要构造枢轴统计量等优

点。因此，自从该方法被提出以来，被许多统计学者关注和推广。本文研究维数发散的高维数据的经验似然推断。

§1.2 文献综述

经验似然的思想最早可追溯到1975年，Thomas和Grunkemeier在文献 [1] 中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置信区间。Owen (1988,1990) [2, 3] 在独立同分布情形下构造了经验似然比统计量，并证明了其渐近分布为标准的χ_k^2分布。这是一种有效的非参数统计推断方法，被称之为经验似然，Owen (2001) 在文献 [4] 中对这种方法作了详细阐述。由于经验似然方法具有如上节所介绍的优点，该方法引起了许多统计学者的兴趣，并被成功地应用到各种模型和各种领域。Owen (1991) [5] 将经验似然应用于线性回归模型，Chen (1993, 1994) [6]、 [7] 研究了线性回归模型的经验似然推断。Diciccio，Hall 和 Romano (1991) [8] 将经验似然应用到均值的光滑函数，并证明了经验似然是可以Bartlett校正的，Chen，Mulayath 和 Abraham (2008) [9] 提出了一种校正的经验似然方法。Kolaczyk (1994) 在文献 [10] 中将经验似然推广到广义线性模型，Chen和Cui (2003) [11] 提出了一种拓展的经验似然方法改进广义线性模型参数估计的效率。Qin 和 Lawless (1994) [12] 将经验似然应用于估计方程，证明了参数的经验似然估计渐近具有正态性以及经验似然比统计量具有渐近χ_k^2分布。Wang 和 Jing (1999) [13]、 [14] 中分别研究了固定设计和完全随机设计实验时部分线性模型的经验似然推断，Shi 和 Lau (2000) [15] 也讨论了部分线性模型的经验似然推断，Lu (2009) [16] 进一步将经验似然方法推广到异方差部分线性模型。Xue 和 Zhu (2006) [17] 讨论了单指标模型的经验似然推断，Zhu 和 Xue (2006) [18] 又研究部分线性单指标模型的经验似然方法。Chen 和 Qin (2000) [19] 将经验似然方法应用到非参数回归模型，Fan 和 zhang (2004) [20] ，Qin 和 Tsao (2005) [21] 也分别在不同条件下讨论了非参数回归模型的经验似然推断。Chen 和 Hall (1993) [22] ，Chen (1996) [23] 各自将经验似然应用于其指定的半参数模型；Bertail在文献 [24] 中讨论了更一般的半参数模型的经验似然推断；Wang，Cui 和 Li (2013) [25] 将经验似然方法应用于半参数估计方程。You 和 Zhou (2006) [26] 研究了部分线性变系数模型参数部分的经

验似然推断；Huang 和 Zhang (2009) [27] 将经验似然应用于部分线性模型非参数部分的推断；此后，Huang 和 Zhang (2010) [28]，Xue 和 Wang (2012) [29] 分别研究了变系数单指标模型和单指标变系数模型的经验似然推断。Chen 和 Qin (1993) [30]，Zhong 和 Rao (2000) [31] 利用经验似然方法研究了抽样调查问题；Zhang (1997a, b) [32, 33] 对分位数回归的经验似然推断进行了研究；Chan 和 Ling (2006) [34] 将经验似然推广到GARCH模型；Kitamura (2001) [35]，Bravo (2004) [36] 将经验似然推断方法应用到经济模型的研究。

删失数据是生存分析中一种最常见的数据类型，近年来，越来越多的文献将经验似然方法推广到删失数据情形下各种统计模型中。Qin 和 Jing (2001a, b) [37]、[38] 研究了删失情形下线性模型和部分线性模型的经验似然推断，随后，Qin 和 Jing (2001c) [39] 又将经验似然方法推广到随机删失情形下的Cox回归模型。Wang 和 Jing (2001) [40] 证明了在右删失数据情形下，经验似然比统计量渐近分布为加权χ^2。Wang 和 Li (2002) [41] 研究了随机删失情形下部分线性模型参数的经验似然推断；Pan 和 Zhou (2002) [42] 研究了删失情形下累积危险率函数的经验似然推断；Qin 和 Tsao (2003) [43] 讨论了基于删失数据的中位数回归模型的经验似然推断；Li 和 Wang (2003) [44] 将经验似然应用到由删失情形下的回归分析；Lu 和 Liang (2006) [45] 将经验似然推广到线性变换模型；Qin 和 Zhao (2007) [46] 应用经验似然推断随机删失情形下平均剩余生命。

对于两样本问题，Jing (1995) [47] 将经验似然方法应用到一元两样本均值推断问题，随后，Liu，Zou 和 Zhang (2008) [48] 将其结论推广到多元情形。Qin Zhao (2000) [49] 研究了完全数据下两样本期望经验似然推断，Zi，Zou 和 Liu (2012) [50] 讨论了两样本线性模型系数之差的经验似然推断。Chen 和 Qin (1993) [51] ，Zhou 和 Liang (2005) [52] 分别在不同条件下研究了删失数据情形下两样本期望的经验似然，Shen 和 He (2006) [53] 给出了右删失情形下两生存函数之差的经验似然推断方法，Ren (2008) [54] 就各种删失情形下两样本半参数模型进行加权经验似然推断。

随着“大数据”时代的来临，高维数据中的统计推断问题成为当前统计学的一个研究热点，许多统计学者已成功地将经验似然方法拓展到许多高维数据场合。当样本维数p随样本容量$n(p = o_p(n^{1/3}))$一起趋向无穷时，Hjort，Mckeague 和 Keilegom (2009) [55] 在一定条件下得出了经验似然比统计量渐近分布为正态分布的结论。Chen，Peng 和 Qin [56] 进一步

研究了样本维数对经验似然方法的影响，证明了当$p = o_p(n^{1/2})$时，经验似然方法依然适用，改进了 [55] 的结果。Li，Lin 和 Zhu (2012) [57] 将经验似然应用到高维情形下的变系数部分线性模型，并提出了改进的经验似然方法以提高其统计推断的效率；Tang，Li 和 Lian (2013) [58] 讨论了高维删失情形下部分线性比率危险率模型的经验似然推断。

变量选择是高维数据统计推断中最重要的研究课题之一。经典的变量选择方法包括逐步回归、基于各种准则的最优子集选择法等。Akaike (1973) [59] 提出了基于信息论的赤池信息准则(AIC)选择最优变量子集，Schwartz (1978) [60] 研究了基于贝叶斯方法的贝叶斯信息准则(BIC)，Stone (1979) [61] 等讨论和比较了这两种准则，两者在不同情形下各有其优势。Mallows (1973) [62] 提出了基于模型误差的C_p准则的最优子集选择方法，类似的还有Akaika (1969) [63] 的FPE准则、Allen (1974) [64] 的PRESS准则、Tibshirani 和 Knight (1999) [65] 的协方差膨胀准则(CIC)等。经典的变量选择方法简单易懂，但也存在一些缺点，例如：由于选择过程中忽略了随机误差项导致无法给出估计的理论性质；由于选择过程的离散性导致选择结果不具有稳定性，且产生误差累积；计算复杂，随着变量个数增加计算量急剧增加。因此，这些经典的变量选择方法不再适用于做高维数据的变量选择。

对于高维数据的变量选择问题，一类基于惩罚函数的变量选择方法在近些年受到广泛关注。这类方法的思想是在进行参数估计的同时，利用惩罚函数将较小的系数估计值压缩为零而将系数估计值较大的保留，在估计出系数的同时选择出重要变量。基于惩罚函数的变量选择方法既能克服经典变量选择方法的局限性，又能同时实现变量选择和系数估计两个目标，因此被统计学者们广泛关注和推广。Hoerl 和 Kennard (1970) [66] 给出了最早的惩罚函数–岭(Ridge)回归惩罚函数(L_2惩罚)。随后，惩罚函数文献包括：桥(Bridge)回归惩罚函数(L_q惩罚)(Frank 和 Friedman (1993) [67])；Lasso惩罚函数(L_1惩罚)(Tibshirani (1996) [68])；光滑切片绝对偏差惩罚函数(SCAD)(Fan 和 Li (2001) [69])；弹性网(Elastic-net)惩罚函数(Zou Hastie (2005) [70])；Adaptive Lasso惩罚(Zou (2006) [71])；Dantzig selector惩罚(Candes 和 Tao (2007) [72])；非凹惩罚函数(MCP)(Zhang (2010) [73])等。当样本维数p随样本容量n一起趋向无穷时，许多统计学者也对惩罚函数方法进行了研究，文献包括：Fan 和 Peng (2004) [74] ，Zhao 和 Yu (2006) [75])，Huang，Horowitz 和 Ma (2008) [76] ，Huang 和 Ma (2008)

[77]，Zou 和 Zhang (2009) [78]，Lin 和 Lv (2013) [79] 等。

Bartolucci (2007) [80] 提出了一种惩罚经验似然(PEL)方法，并证明了惩罚经验似然比也具有Wilks现象；Otsu (2007) [81] 研究了半参数模型的惩罚经验似然方法；Hou，Song 和 Wang (2014) [82] 将惩罚经验似然方法应用到Cox比率危险率模型。当$n \to \infty$，$p \to \infty$时，Tang 和 Leng (2010) [83] 给出了惩罚高维经验似然方法，随后，Leng Tang (2012) [84] 又将这种方法推广到高维估计方程。文献 [83] 和 [84] 都要求$p < n$，当$p > n$时，Lahiri 和 Mukhopadhyay [85] 研究了总体均值的惩罚经验似然推断，并给出了惩罚经验似然比统计量的渐近性质。

§1.3 经验似然

Owen [2-5] 提出了经验似然思想和方法。

设$X_1, \cdots, X_n$为p维随机样本，其经验累积分布函数定义为：$F_n = \frac{1}{n}\sum\limits_{i=1}^{n}\delta_{X_i}$，其中$\delta_x(B) = I(x \in B)$，$I(\cdot)$为示性函数。如果$X_1, \cdots, X_n$相互独立且具有共同累积分布$F_0$，那么

$$L(F) = \prod_{i=1}^{n} F(\{X_i\})$$

为分布函数F的非参数似然函数，其中

$$F(\{X_i\}) = F(X_i) - F(X_i-) = p_i$$

为F在X_i的概率质量。下面定理表明经验分布函数 F_n 使 F 达到最大，即 F_n 是 F 的非参数极大似然估计。

定理 1.3.1[4] 假设$X_1, \cdots, X_n \in \mathbb{R}^p$相互独立且具有共同累积分布 F_0，F_n 为其经验累积分布函数，F 为其任意分布函数。如果 $F \neq F_n$，那么 $L(F) < L(F_n)$。

定义非参数似然比：

$$R(F) = \frac{L(F)}{L(F_n)}.$$

如果我们感兴趣的参数$\theta = T(F)$是关于分布F的函数，那么可以定义截面似然比(profile likelihood ratio)函数：

$$R(\theta) = \sup\{R(F) | T(F) = \theta,\ F \in \mathfrak{F}\}, \tag{1.3.1}$$

其中$\mathfrak{F}$表示某一分布函数类。当$R(\theta_0) < r_0$时，经验似然假设检验拒绝原假设H_0： $T(F_0) = \theta_0$。另外，经验似然置信域为$\{\theta|R(\theta) > r_0\}$，这里$r_0$为临界值，可以通过经验似然比统计量的渐近分布(类似参数似然的Wilks定理)来确定。由于上述截面经验似然比函数计算复杂，在实际应用中很难实现，于是Owen定义了另一种等价的经验似然比函数。

假设分布 F_p 取值 X_i 的概率为 p_i，$i = 1, \cdots, n$，利用$\sum\limits_{i=1}^{n} p_i = 1$，Owen定义下面的经验似然函数

$$L(\theta) = \sup\{\prod_{i=1}^{n} p_i | T(F_p) = \theta,\ \sum_{i=1}^{n} p_i = 1,\ p_i \geq 0\}. \tag{1.3.2}$$

当$\sum\limits_{i=1}^{n} p_i = 1$时，$p_i = \frac{1}{n}$使$L(\theta)$取到极大值$n^{-n}$，因此，经验似然比函数又可以定义为

$$R(\theta) = \sup\{\prod_{i=1}^{n} (np_i) | T(F_p) = \theta,\ \sum_{i=1}^{n} p_i = 1,\ p_i \geq 0\}. \tag{1.3.3}$$

下面以总体均值为例说明经验似然方法具体推断过程。

令$\int xdF(x) = \mu$，分布 F 取值 X_i 的概率为 p_i，利用 (1.3.3) 得总体均值μ的经验似然比为

$$R(\mu) = \sup\{\prod_{i=1}^{n} (np_i) | \sum_{i=1}^{n} X_i p_i = \mu,\ \sum_{i=1}^{n} p_i = 1,\ p_i \geq 0\}. \tag{1.3.4}$$

利用Lagrange条件极值方法得

$$p_i = \frac{1}{n} \frac{1}{\lambda^{\top}(X_i - \mu)},\ i = 1, \cdots, n, \tag{1.3.5}$$

其中λ为lagrange乘数，由下面方程所确定，

$$\sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda^{\top}(X_i - \mu)}. \tag{1.3.6}$$

联合 (1.3.5) 和 (1.3.6) 得对数经验似然比函数为

$$l(\mu) = -2\log(R(\mu)) = 2\sum_{i=1}^{n} \log\{1 + \lambda^{\top}(X_i - \mu)\}. \tag{1.3.7}$$

定理 1.3.2[4] 假设 $X_1,\cdots,X_n\in\mathbb{R}^p$ 相互独立且具有共同分布 F_0，$E(X_i)=\mu$，$\mathrm{Var}(X_i)=V_0$，$\mathrm{Rank}(V_0)=q>0$，如果总体均值参数真实值为 μ_0，那么

$$-2\log(R(\mu_0))\xrightarrow{d}\chi_q^2. \tag{1.3.8}$$

利用定理 (1.3.8)，我们可以构造关于 μ 的置信水平为 $(1-\alpha)$ 的经验似然置信域，

$$I_\alpha=\{\mu|-2\log(R(\mu))\le\chi_{q,\alpha}^2\},$$

其中 $\chi_{q,\alpha}^2$ 满足 $P(\chi_q^2>\chi_{q,\alpha}^2)=\alpha$。

惩罚变量选择普遍采用“损失函数+惩罚函数”的变量选择方法，类似地，惩罚经验似然通常也使用“经验似然比函数+惩罚函数”方式，例如：Tang 和 leng [83] 定义的总体均值 μ 的惩罚经验似然为

$$l_p(\mu)=2\sum_{i=1}^n\log\{1+\lambda^\top(X_i-\mu)\}+n\sum_{i=1}^n p_\tau(|\mu_i|), \tag{1.3.9}$$

其中 $p_\tau(\cdot)$ 为惩罚函数，τ 为规范化参数，一般采用交叉验证 (CV) 或广义交叉验证 (GCV) 来选择 τ 。通过最小化 (1.3.9) 可以完成变量选择。同时得到了参数估计值。不同的罚函数 $p_\tau(\cdot)$ 对应不同的惩罚经验似然方法。常用的惩罚函数有 Lasso 回归惩罚(L_1惩罚) [68]：

$$p_\tau(|\theta|)=\tau|\theta|. \tag{1.3.10}$$

Ridge 回归惩罚(L_2惩罚) [66]：

$$p_\tau(|\theta|)=\tau|\theta|^2. \tag{1.3.11}$$

Bridge 惩罚(L_q惩罚) [67]：

$$p_\tau(|\theta|)=\tau|\theta|^q. \tag{1.3.12}$$

SCAD 惩罚 [69] (通常令$a=3.7$)：

$$p_\tau(|\theta|)=\begin{cases}\tau|\theta|, & \text{当 } |\theta|\le\tau,\\ -\frac{(\theta^2-2a\tau|\theta|)}{2(a-1)}, & \text{当 } \tau<|\theta|<a\tau,\\ \frac{(a+1)\tau^2}{2}, & \text{当 } |\theta|\ge a\tau.\end{cases} \tag{1.3.13}$$

弹性网惩罚 (Elastic net) [70]：

$$p_{\tau,\ \gamma}(|\theta|)=\gamma\beta^2+\tau|\beta|. \tag{1.3.14}$$

Adaptive Lasso 惩罚 [71] ($\omega = |\hat{\beta}^{*}|^{-\gamma}$是对不同参数分配的权重)：

$$p_\tau(|\theta|) = \tau\omega|\beta|. \tag{1.3.15}$$

非凹惩罚 (MCP) [72] ：

$$p_\tau(|\theta|) = a\tau^2 - \frac{(|\beta| - a\tau)^2}{a} I(|\beta| < a\tau),\ a > 1. \tag{1.3.16}$$

§1.4 本文主要研究工作及内容结构

本文主要研究了样本维数 p 随样本容量 n 趋向无穷情形的经验似然推断问题，共分六章，各章内容安排如下：

第一章介绍了本论文的研究背景、文献综述、相关知识以及本文内容结构。

第二章研究了高维情形下半参数模型的经验似然推断。首先，通过半参数估计方程构造关于未知参数的经验似然比统计量，证明了在一定条件下，当 $n \to \infty$，$p \to \infty$ 时，经验似然比渐近分布为正态分布，并证明了通过经验似然方法得到的参数估计量具有一致性；其次，利用惩罚经验似然方法研究了高维稀疏半参数模型的变量选择和参数估计问题，证明了惩罚经验似然比统计量具有渐近 χ^2_q 分布，而且惩罚经验似然方法也具有变量选择的Oracle 性质。数值模拟和实证分析结果进一步验证了经验似然和惩罚经验似然方法的优良性质。

第三章将经验似然方法推广到高维删失情形下的可加危险率模型。定义了未知参数(部分未知参数)的经验似然比统计量，并证明了该统计量渐近于正态分布(χ^2_q 分布)，所得结果可以用来构造未知参数(部分未知参数)的置信域(置信区间或置信域)。对于高维稀疏删失情形，利用惩罚经验似然方法研究了可加危险率模型的变量选择和参数估计问题。在一定条件下给出了惩罚经验似然统计量的渐近分布 – χ^2_q 分布，并讨论了惩罚经验似然的Oracle性质。通过数值模拟，对经验似然(惩罚经验似然)方法和其他已经存在的一些方法构造的置信域(置信区间)的覆盖概率(区间长度)以及估计的均方误差等方面进行了比较，结果显示经验似然(惩罚经验似然)方法具有其优越性。

第四章研究了高维情形下异方差部分线性单指标模型的经验似然推断。在双重稳健估计方法基础上建立估计方程，构造异方差部分线性单指标模型的经验似然比统计量，证明了在一定条件下该统计量渐近分布为正态分布，并构造了参数的经验似然置信域。在实际应用中，特别是高维情形，往往部分参数的置信区间或置信域更值得关注，因此，我们进一步构造了关于部分参数的经验似然比统计量，证明了该统计量具有渐近 χ_q^2 分布，并利用这个结论构造了部分参数的经验似然置信区间或置信域。数值模拟和实例分析比较了双重稳健方法和经验似然方法的覆盖概率(或置信区间长度)，结果表明经验似然方法更好。

第五章利用经验似然方法研究了高维情形下的两样本问题的统计推断。在高维情形下，两样本均值的经验似然比的渐近分布为正态分布，其结果可以用来检验两样本均值是否相等假设以及构造两样本均值之差的置信域。类似地，利用经验似然方法对两样本线性模型系数进行了统计推断，即检验两样本线性模型系数是否相等假设以及构造两样本线性模型系数之差的置信域。数值模拟和实证分析比较了经验似然和已有的其他方法。

第六章对本文进行总结，并对进一步的研究工作做了展望。

第二章 高维半参数模型的经验似然推断

本章，我们讨论半参数模型的经验似然推断问题。当 $n \to \infty$，$p \to \infty$ 时，首先，我们构造关于未知参数的经验似然比统计量，证明了在一定条件下，经验似然比渐近分布为正态分布，并证明了通过经验似然方法得到的参数估计量具有一致性；其次，我们利用惩罚经验似然方法研究了高维稀疏半参数模型的变量选择和参数估计问题，其惩罚经验似然比统计量具有渐近 χ_q^2 分布，而且惩罚经验似然方法也具有变量选择的Oracle性质。

本章内容结构如下：第一节主要介绍高维半参数模型经验似然方法及其主要结论；第二节，利用惩罚经验似然研究高维稀疏半参数模型的变量选择、参数估计和假设检验问题以及相应的一些结果；第三节为定理的证明；第四和第五节分别为有限样本情形下的数值模拟结果和实例分析。

§2.1 高维半参数模型的经验似然

半参数模型是上世纪八十年代发展起来的一种重要统计模型，它结合了参数模型和非参数模型的优点，因而受到广大统计学者广泛关注，如文献 [22]–[27] 等。我们通过利用包含未知函数$h(\cdot)$的估计方程来考虑半参数模型，其形式如下：

$$E\{g(\mathbf{X}, H(T), \theta)\} = 0, \tag{2.1.1}$$

其中$\mathbf{X} = (X^\top, T^\top)^\top$ 为随机向量，$\theta = (\theta_1, \theta_2, \cdots, \theta_p)^\top \in \Theta_\theta$ 为未知参数，$\Theta_\theta \in \mathbb{R}^p$， g 是由已知函数组成的 r- 维向量，$H(T) \in \mathbb{R}^k$ 是 T 的未知光滑函数，满足

$$H(t) = (H_1(t), H_2(t), \cdots, H_k(t))^\top = E(\varphi(\cdot) \mid T = t),$$

$\varphi(\cdot) = (\varphi_1(\cdot), \cdots, \varphi_k(\cdot))^\top$ 是已知的可测函数。

模型 (2.1.1) 包括了多种已知的半参数模型，例如：部分线性模型 (partially linear model)：

$$g(\mathbf{X}, H(T), \theta) = Y - X^\top\theta - h(T); \tag{2.1.2}$$

变系数部分线性模型(partially linear varying-coefficient model)：

$$g(\mathbf{X}, H(T), \theta) = Y - X_1^\top \theta - h(T)^\top X_2; (\mathbf{X} = (X_1^\top, X_2^\top)^\top) \tag{2.1.3}$$

单指标模型 (single index model)：

$$g(\mathbf{X}, H(T), \theta) = Y - h(X^\top \theta). \tag{2.1.4}$$

另外还有，部分线性EV模型 (partially linear errors-in-variables model) 、半参数变系数线性模型 (semiparametric varying-coefficient linear model) 等。

假设 $\{(\mathbf{X}_i^\top, T_i^\top)^\top\}_{i=1}^n$ 相互独立且具有共同分布 F， $F \in \mathfrak{F}$， $\mathfrak{F} = \mathfrak{F}_x \times \mathfrak{F}_t \subset \mathbb{R}^{p+1+k_1} \times [a, b]$，其中，$k_1$为正整数，$a$和$b$为常数。估计方程中的函数

$$g(\mathbf{X}, H(T), \theta) = (g_1(\mathbf{X}, H(T), \theta), g_2(\mathbf{X}, H(T), \theta), \cdots, g_r(\mathbf{X}, H(T), \theta))^\top$$

满足

$$E\{g(\mathbf{X}, H(T), \theta)\} = 0, \quad (r \geq p).$$

当 k 和 p 为固定常数时，Wang et al. (2013) [25] 给出了模型 (2.1.1) 中参数 θ 的经验似然比统计量，并证明了该统计量的渐近分布为 χ_p^2 分布。我们考虑，当 $n \to \infty$，$k, p, r \to \infty$情形下模型 (2.1.1) 的经验似然推断。首先定义关于参数 θ 的经验似然函数：

$$L(\beta) = \sup\{\prod_{i=1}^n (nq_i) : \sum_{i=1}^n q_i = 1, q_i \geq 0, \sum_{i=1}^n q_i g(\mathbf{X}_i, H(T_i), \theta) = 0\}. \tag{2.1.5}$$

由于 (2.1.5) 中包含未知函数 $H(t)$，因此上式定义的经验似然函数不能直接用来做关于参数 θ 的经验似然推断。为了解决这个问题，我们用核估计 $\hat{h}(t) = \sum_{j=1}^n \mathcal{W}_{nj}(t)\varphi(X_j)$ 代替 (2.1.5) 中的未知函数 $h(t)$，其中：$\mathcal{W}_{nj}(t) = K(\frac{t-T_j}{h}) / \sum_{j=1}^n K(\frac{t-T_j}{h})$，$K(\cdot)$ 是核函数，h为窗宽。被估计的经验似然函数可以定义为

$$\tilde{L}(\beta) = \sup\{\prod_{i=1}^n (nq_i) : \sum_{i=1}^n q_i = 1, q_i \geq 0, \sum_{i=1}^n q_i g(\mathbf{X}_i, \hat{H}(T_i), \theta) = 0\}. \tag{2.1.6}$$

其被估计的对数经验似然比函数为

$$\tilde{l}(\theta) = -2[\log\{\tilde{L}(\theta)\} - n\log(n)]. \tag{2.1.7}$$

由lagrange乘数法，得到 (2.1.6) 中的 q_i, $i = 1, \cdots, n$,

$$q_i = \frac{1}{n} \frac{1}{1 + \lambda^\top g(\mathbf{X}_i, \hat{H}(T_i), \theta)}, \tag{2.1.8}$$

其中lagrange乘数λ满足：

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g(\mathbf{X}_i,\hat{H}(T_i),\theta)}{1+\lambda^\top g(\mathbf{X}_i,\hat{H}(T_i),\theta)}=0. \tag{2.1.9}$$

联合 (2.1.7)–(2.1.9) ，得被估计的对数经验似然比函数

$$\tilde{l}(\theta)=-2[\log\{\tilde{L}(\theta)\}-n\log(n)]=2\sum_{i=1}^{n}\log\{1+\lambda^\top g(\mathbf{X}_i,\hat{H}(T_i),\theta)\}. \tag{2.1.10}$$

求解使 (2.1.7) 达到最大的θ等价于求解使 (2.1.10) 达到最小的 θ，因此我们通过求解 (2.1.10) 的最小值获得 θ 的估计值 $\hat{\theta}$。$\hat{\theta}$ 可以表示为:

$$\hat{\theta}=\arg\min_{\theta\in\Theta_\theta}\max_{\lambda\in\hat{\Lambda}_n(\theta)}2\sum_{i=1}^{n}\log\{1+\lambda^\top g(\mathbf{X}_i,\hat{H}(T_i),\theta)\}, \tag{2.1.11}$$

其中：$\hat{\Lambda}_n(\theta)=\{\lambda\in\mathbb{R}^r:\lambda^\top g(\mathbf{X}_i,\hat{H}(T_i),\theta)\in\mathcal{V},i=1,2,\cdots,n\}$，$\mathcal{V}$ 为包括零点的开区间。

当k和p为固定常数时，Wang et al. (2013) [25] 证明了在一定条件下：$\|\lambda\|=O_p(n^{-1/2})$。这个结论在$k$和$p$随样本容量$n$一起趋向无穷时不再成立。为了得到被估计的对数经验似然比统计量的渐近分布，我们给出如下假设条件：

假设1. $\{h_j(t)\}_{j=1}^{k}$ 满足一阶 Lipschitz 条件；

假设2. $K(t)$是定义在有界紧支撑上的对称密度函数，窗宽$h=O_p(\frac{1}{\sqrt[3]{n}})$；

假设3. T 的密度函数 $r(t)$ 满足：

$$0<\inf_{a\le t\le b}r(t)\le\sup_{a\le t\le b}r(t)<\infty;$$

假设4. $\sup_t E\{\|\varphi_l(X)\|^2|T=t\}<\infty,\quad l=1,2,\cdots,k$；

假设5. 存在满足下面要求的 $u(\mathbf{X},T)$ ：

$$\frac{\partial^2 g(\mathbf{X},H(T),\theta)}{\partial H_j\partial H_l}\le u(\mathbf{X},T),E\{u^2(\mathbf{X},T)\}<\infty,\ (j,l=1,2,\cdots,k),$$

$$\sup_t E\{\|\frac{\partial g(\mathbf{X},H(T),\theta)}{\partial H_l}\|^2|T=t\}<\infty,\ l=1,2,\cdots,k;$$

假设6. 存在满足下列要求的 $v_1(\mathbf{X},T)$ 和 $v_2(\mathbf{X},T)$ ：

$$\frac{\partial g(\mathbf{X},H(T),\theta)}{\partial\theta_l}<v_1(\mathbf{X},T),E\{v_1^2(\mathbf{X},T)\}<\infty,\ (l=1,2,\cdots,p),$$

$$\frac{\partial^2 g(\mathbf{X},H(T),\theta)}{\partial\theta_i\partial\theta_j} \le v_2(\mathbf{X},T), E\{v_2^2(\mathbf{X},T)\} < \infty,\ (i,j=1,2,\cdots,p);$$

假设7. Θ_θ 是包含于 $\mathbb{R}^p$ 中的紧集，而且参数真实值 $\theta_0 \in \Theta_\theta$ 是方程 $E\{g(\mathbf{X},\hat{H}(T),\theta)\}=0$的唯一解；

假设8. 记 $g(\mathbf{X},H(T),\theta)_t$ 为 $g(\mathbf{X},H(T),\theta)$ 得第 t 个分量，存在$\alpha(\alpha \ge 12)$使得

$$E\{\sup_{\theta\in\Theta_\theta}\|g(\mathbf{X},H(T),\theta)_t\|\} = o_p(n^{1/\alpha});$$

假设9. 记 $\Sigma = E\{g(\mathbf{X},H(T),\theta)g(\mathbf{X},H(T),\theta)^\top\}$，$\Sigma$ 的特征值都是有限的正常数；

假设10. 当$n\to\infty$时，$k\to\infty$，$p\to\infty$，$pn^{-(1/6)}\to 0$，$kn^{-(1/6)}\to 0$，$p/r\to c_0$ ($0<c_0<1$)。

假设 1-5 是为了保证非参数估计不影响估计的经验似然比统计量的渐近结果，假设 7-8 能使得 (2.1.10) 的极小值存在以及其展开式中余项趋向于零，假设 6，9-10 确保 θ 的经验似然估计的理论性质。由于没有关于$g(\mathbf{X},H(T),\theta)$的附加信息，推导$\tilde{l}(\theta)$的渐近分布比较困难，因此，我们给出了较强的假设条件6-10确保经验似然比统计量的渐近分布。对于具体的半参数模型，假设条件6-10可以适当减弱，本章第4节给出了两个具体的例子。

使得 (2.1.10) 达到最小的θ，称为参数 θ 的经验似然估计，记为 $\hat{\theta}$。下面，我们给出高维情形下经验似然估计$\hat{\theta}$的理论性质。

定理 2.1.1 设参数 θ 的真实值为 θ_0，在假设 1-10 成立下，有

$$\lim_{n\to\infty} P(\|\hat{\theta}-\theta_0\| = O_p\{(p/n)^{1/2}\}) = 1.$$

定理 2.1.1 说明经验似然估计 $\hat{\theta}$ 具有渐近一致性，收敛于真实参数的速度为 $O_p\{(p/n)^{1/2}\}$。

定理 2.1.2 设参数 θ 的真实值为 θ_0，在假设 1-10 成立下，当 $n\to\infty$ 时，有

$$\sqrt{n}B_nV^{-1/2}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,G),$$

其中$\xrightarrow{d}$表示依分布收敛，$B_n\in\mathbb{R}^{q\times p}$，$B_nB_n^\top\to G$，$G$为$q\times q$矩阵($q$为正整数),

$$V = [E\{\frac{\partial g(\mathbf{X},H(T),\theta)}{\partial\theta}\}^\top E\{g(\mathbf{X},H(T),\theta)g(\mathbf{X},H(T),\theta)^\top\}^{-1}E\{\frac{\partial g(\mathbf{X},H(T),\theta)}{\partial\theta}\}]^{-1}.$$

定理 2.1.2 中向量左乘 B_n 表示将 p 维向量 $(p \to \infty)$ 投影到 q 维空间(q为正整数)上，$\hat{\theta} - \theta_0$ 的投影向量 (q 维) 具有渐近正态性。定理 2.1.2 表明高维半参数模型中参数θ的经验似然估计是渐近有效的。

下面定理 2.1.3 给出了估计的经验似然比统计量的渐近分布。

定理 2.1.3 在假设 1-10 成立下，如果参数 θ 的真实值为 θ_0，而且假设 8 中 $\alpha \geq 36$，那么

$$(2p)^{-\frac{1}{2}}(\tilde{l}(\theta_0) - p) \xrightarrow{d} N(0,1). \tag{2.1.12}$$

定理 2.1.3 的结论既能用来检验假设：

$$H_0 : \theta = \theta_0 \quad , \quad H_1 : \theta \neq \theta_0,$$

也能用来构造参数θ的置信域。令

$$I_\alpha(\theta) = \{\theta : \tilde{l}(\theta) \leq p + z_\alpha \sqrt{2p}\},$$

其中 z_α 是标准正态分布的上侧 $\alpha-$分位数。参数 θ 的渐近置信水平为 $1 - \alpha$ 的置信域为$I_\alpha(\theta)$，即

$$P(\theta \in I_\alpha(\theta)) = 1 - \alpha + o_p(1).$$

§2.2 高维稀疏半参数模型的惩罚经验似然

在高维数据分析中，往往数据维数会很大，提供的信息也特别多，但是其中很多的信息可能是噪声而非有价值的信息，因此在建立统计模型进行统计推断的时候，假设所有变量中只有一部分变量的系数不为零是合理的，即高维稀疏假设。对于高维稀疏问题，如何准确选择出这些非零系数的变量称为统计推断中的变量选择问题。解决高维稀疏情形下的变量选择问题最见的有效方法是惩罚方法。在高维稀疏假设下，我们利用惩罚经验似然方法对模型 (2.1.1) 进行统计推断。正如一般的惩罚变量选择方法一样，我们采用“损失函数+惩罚函数”的模式，即“经验似然比函数+惩罚函数”构造惩罚经验似然比函数$\tilde{l}_p(\theta)$，其形式如下：

$$\tilde{l}_p(\theta) = \sum_{i=1}^{n} \log\{1 + \lambda^\top g(\mathbf{X}_i, \hat{H}(T_i), \theta)\} + n \sum_{j=1}^{p} p_\nu(|\theta_j|), \tag{2.2.1}$$

其中：$p_\nu(|\beta_j|)$ 为惩罚函数，ν 为规范化参数。常用惩罚函数见 §1.3 中所述，本文采用 Fan 和 Li (2001) [2009] 中给出的SCAD惩罚函数，其一阶导数如下：

$$p_\nu^{'}(t) = \nu\{I(t \le \nu) + \frac{(a\nu - t)_+}{(a-1)\nu} I(t > \nu)\},$$

其中：$I(\cdot)$为示性函数，$a = 3.7$。

令 $\theta_0 = (\theta_{01}, \theta_{02}, \cdots, \theta_{0p})^\top \in \mathbb{R}^p$，$\mathcal{A} = \{j : \theta_{0j} \neq 0\}$，$s = |\mathcal{A}|$ 未知。不失一般性，设 $\theta = (\theta^{(1)\top}, \theta^{(2)\top})^\top$，其中 $\theta^{(1)} \in \mathbb{R}^s$ 和 $\theta^{(2)} \in \mathbb{R}^{p-s}$ 分别表示非零和零参数分量，即：$\theta_0 = (\theta_0^{(1)\top}, 0^\top)^\top$。针对惩罚函数$p_\nu(\cdot)$，我们给出如下假设：

假设*11.* 当 $n \to \infty$时，$\nu(p/n)^{1/2} \to \infty$，$\min\limits_{j \in \mathcal{A}} \theta_{0j}/\nu \to \infty$；

假设*12.* $\max\limits_{j \in \mathcal{A}} P_\nu^{'}(|\theta_{0j}|) = o\{(np)^{-1/2}\}$，$\max\limits_{j \in \mathcal{A}} P_\nu^{''}(|\theta_{0j}|) = o\{(p)^{-1/2}\}$。

令 $\mathcal{I}_p = (D_1^\top, D_2^\top)$，其中 $\mathcal{I}_p$ 为 p–阶单位矩阵，$D_1 \in \mathbb{R}^{s \times p}$，$D_2 \in \mathbb{R}^{(p-s) \times p}$，$\|D\|$ 表示矩阵 D 的 Frobenius 范数：$\|D\| = \{tr(D^\top D)\}^{1/2}$。使 (2.2.1) 达到最小的$\theta$称为惩罚经验似然估计，记为：$\hat{\theta}_P$，

$$\hat{\theta}_P = \arg\min_{\theta \in \Theta_\theta} \max_{\lambda \in \hat{\Lambda}_n(\theta)} \{\sum_{i=1}^n \log\{1 + \lambda^\top g(\mathbf{X}_i, \hat{H}(T_i), \theta)\} + n \sum_{j=1}^p p_\nu(|\theta_j|)\}.$$

设$\hat{\theta}_p = (\hat{\theta}_p^{(1)\top}, \hat{\theta}_p^{(2)\top})^\top$，下面定理给出了惩罚经验似然估计的理论性质。

定理 2.2.1 在假设1-12成立下，当 $n \to \infty$ 时，惩罚经验似然估计 $\hat{\theta}_P$ 满足：

(1). $\lim\limits_{n \to \infty} p(\hat{\theta}_p^{(2)} = 0) = 1$;

(2). $\sqrt{n} B_n V_p^{-1/2} (\hat{\theta}_p^{(1)} - \theta_0^{(1)}) \xrightarrow{d} N(0, G)$，其中：

$$V_p = D_1 V - D_1 V D_2^\top (D_2 V D_2^\top)^{-1} D_2 V,$$

其中V由定理2.1.2给定，$B_n \in \mathbb{R}^{q \times p}$，$B_n B_n^\top \to G$，$G \in \mathbb{R}^{q \times q}$，$q$为固定正整数。

定理2.2.1表示惩罚经验似然估计 $\hat{\theta}_p$ 具有一致性和Oracle性质，且非零参数分量 $\theta^{(1)}$ 的估计 $\hat{\theta}_p^{(1)}$ 具有渐近正态性。

考虑如下假设检验：

$$H_0: \quad L_n\theta_0 = 0\ , \qquad H_1 \quad L_n\theta_0 \neq 0,$$

其中：$L_n \in \mathbb{R}^{q\times p}$ 满足 $L_nL_n^{\top} = I_q$，q 为固定正整数，I_q 为 q–阶单位矩阵。由 (2.2.1)，惩罚经验似然比检验统计量定义为：

$$\tilde{l}_p(L_n) = -2\{\tilde{l}_p(\hat{\theta}_p) - \min_{\theta, L_n\theta_0=0} \tilde{l}_p(\theta)\}. \tag{2.2.2}$$

下面定理给出了惩罚经验似然比检验统计量的渐近分布。

定理 2.2.2 在假设1-12成立下，当 $n \to \infty$ 以及原假设H_0成立时，

$$\tilde{l}_p(L_n) \xrightarrow{d} \chi_q^2.$$

根据定理 2.2.2，我们可以构造 $L_n\theta_0$ 的置信水平为$(1-\alpha)$的渐近置信域：

$$\mathcal{I}_\alpha(L_n\theta) = \{\theta : -2\{\tilde{l}_p(\hat{\theta}_p) - \min_{\theta, L_n\theta=0} \tilde{l}_p(\theta)\} \le \chi^2_{q,(1-\alpha)}\}, \tag{2.2.3}$$

其中：$\chi^2_{q,(1-\alpha)}$ 是 χ^2_q 分布的 $(1-\alpha)$ 分位数。即：当$n\to\infty$时，

$$P(L_n\theta_0 \in \mathcal{I}_\alpha(L_n\theta)) \to 1-\alpha.$$

评论 本章研究了高维情形下半参数模型的经验似然推断和高维稀疏情形下利用惩罚经验似然对半参数模型进行变量选择和参数估计，这些方法都要求 $p < n$。当 $p \ge n$ 时，可以先用 Fan 和 Lv (2008) [86] 的SIS方法降维，即将样本的维数 p 降到小于 n 之后，再用本章方法进行统计推断。

§2.3 定理的证明

本节将给出本章几个定理的证明。整个证明过程中，不同地方出现的 C 的具体的值可能不同，但都表示一个正的常数。令：$\gamma_n = O_p((p/n)^{1/2})$，$D_n = \{\theta \mid \|\theta-\theta_0\| \le C\gamma_n\}$，$g(\theta) = g(\mathbf{X}, H(T), \theta)$，$g_i(\theta) = g(\mathbf{X}_i, H(T_i), \theta)$，$\hat{g}_i(\theta) = g(\mathbf{X}_i, \hat{H}(T_i), \theta)$。证明定理之前，我们先给出下面的引理。

引理 2.3.1[87] 如果假设2-3成立，那么：

(1). $E\{\mathcal{W}_{ni}(T_l)\}^2 \le C(n^2h),\ i \ne l;$

(2). $E\{\mathcal{W}_{ni}(T_l)\}^4 \le C(n^4h),\ i \ne l;$

(3). $E\{\mathcal{W}_{ni}(t)\}^2 \le C(n^2h),\ i = 1, \cdots, n.$

引理 2.3.2 如果假设1-4成立，那么：

$$\|\hat{H}_l(T_i) - H_l(T_i)\| = O_p(n^{-1/3}), \quad i = 1, 2, \cdots, n; \quad l = 1, \cdots, k.$$

证明： 因为$\hat{H}(t) = \sum_{j=1}^{n} \mathcal{W}_{nj}(T_i)\varphi(X_j),\ i = 1, \cdots, n$，所以

$$\begin{aligned}
\hat{H}_l(T_i) - H_l(T_i) &= \sum_{j=1}^{n} \mathcal{W}_{nj}(T_i)\varphi_l(X_j) - H_l(T_i) \\
&= \sum_{j=1}^{n} \mathcal{W}_{nj}(T_i)\varphi_l(X_j) - H_l(T_j) + h_l(T_j) - H_l(T_i) \\
&= R_{1i} + R_{2i},
\end{aligned} \tag{2.3.1}$$

其中：$R_{1i} = \sum\limits_{j=1}^{n} \mathcal{W}_{nj}(T_i)(\varphi_l(X_j) - H_l(T_j))$，$R_{2i} = \sum\limits_{j=1}^{n} \mathcal{W}_{nj}(T_i)(H_l(T_j) - H_l(T_i))$。由核函数的性质、假设4和引理 2.3.1 可得

$$\begin{aligned}
E\|R_{1i}\|^2 &= E\|\sum_{j=1}^{n} \mathcal{W}_{nj}(T_i)(\varphi_l(X_j) - H_l(T_j))\|^2 \\
&\le C\sum_{j=1}^{n} \mathcal{W}_{nj}^2(T_i)E\{(\varphi_l(X_j) - E(\varphi_l(X_j)|T_j))^2|T_i, X_j\} \\
&= O_p((nh)^{-1}).
\end{aligned} \tag{2.3.2}$$

同理，由假设1、假设3和引理2.3.1 (1) 可得

$$\begin{aligned}
E\|R_{2i}\|^2 &= E\|\sum_{j=1}^{n} \mathcal{W}_{nj}(T_i)(H_l(T_j) - H_l(T_i))\|^2 \\
&\le E\{\sum_{j=1}^{n} \mathcal{W}_{nj}(T_i)|T_j - T_i|\}^2 \le h^2\sum_{j=1}^{n} E\{\mathcal{W}_{nj}^2(T_i)|\frac{T_j - T_i}{h}|^2\} \\
&= h^2\sum_{j=1}^{n} E\{\mathcal{W}_{nj}^2(T_i)|\frac{T_j - T_i}{h}|^2 I(|\frac{T_j - T_i}{h}| \le \rho)\} \\
&+ h^2\sum_{j=1}^{n} E\{\mathcal{W}_{nj}^2(T_i)|\frac{T_j - T_i}{h}|^2 I(|\frac{T_j - T_i}{h}| > \rho)\} \\
&\le nh^2\rho^2 E\{\mathcal{W}_{nj}^2(T_i)\} = O_p(n^{-4/3}).
\end{aligned} \tag{2.3.3}$$

联合 (2.3.1)–(2.3.3)，我们有

$$\|\hat{H}_l(T_i) - H_l(T_i)\|^2 = O_p((nh)^{-1}) + O_p(n^{-4/3}) = O_p(n^{-2/3}).$$

综上所述，引理 2.3.2 证明完毕。 □

引理 2.3.3 假设 $E\{\frac{\partial g_i(\theta)}{\partial h}|T=t\}=0$ 成立以及假设1-5、9成立下, 我们有

(1) $\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}\hat{g}_i(\theta)=\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}g_i(\theta)+o_p(1)$;

(2) $\frac{1}{n}\sum\limits_{i=1}^{n}\hat{g}_i(\theta)\hat{g}_i^\top(\theta)=\frac{1}{n}\sum\limits_{i=1}^{n}g_i(\theta)g_i^\top(\theta)+o_p(1)$.

证明： 由Taylor展开公式得

$$
\begin{aligned}
\hat{g}_i(\theta) &= g_i(\theta)+\sum_{l=1}^{k}\frac{\partial g_i(\theta)}{\partial H_l}(\hat{H}_l(T_i)-H_l(T_i))+\xi_i\\
&= g_i(\theta)+\sum_{l=1}^{k}\frac{\partial g_i(\theta)}{\partial H_l}\{\sum_{j=1}^{n}\mathcal{W}_{nj}(T_i)\varphi_l(X_j)-H_l(T_i)\}+\xi_i\\
&= g_i(\theta)+\sum_{l=1}^{k}\frac{\partial g_i(\theta)}{\partial H_l}\{\sum_{j=1}^{n}\mathcal{W}_{nj}(T_i)(\varphi_l(X_j)-H_l(T_j))\}\\
&\quad+\sum_{l=1}^{k}\frac{\partial g_i(\theta)}{\partial H_l}\{\sum_{j=1}^{n}\mathcal{W}_{nj}(T_i)(H_l(T_j)-H_l(T_i))\}+\xi_i,
\end{aligned}
\tag{2.3.4}
$$

其中：$\xi_i=O_p(\|\hat{H}(T_i)-H(T_i)\|^2)$。由 (2.3.4) 可得

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{g}_i(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_i(\theta)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i+A_{n1}+A_{n2}, \tag{2.3.5}
$$

其中：

$$
\begin{aligned}
A_{n1} &= \sum_{l=1}^{k}A_{n1,l}=\sum_{l=1}^{k}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial g_i(\theta)}{\partial H_l}\sum_{j=1}^{n}\mathcal{W}_{nj}(T_i)(H_l(T_j)-H_l(T_i)),\\
A_{n2} &= \sum_{l=1}^{k}A_{n2,l}=\sum_{l=1}^{k}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial g_i(\theta)}{\partial H_l}\sum_{j=1}^{n}\mathcal{W}_{nj}(T_i)(\varphi_l(X_j)-H_l(T_j)).
\end{aligned}
$$

根据引理 2.3.2 ，当 $n\to\infty$时，我们有

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i=\sqrt{n}\times(O_p(n^{-1/3}))^2=O_p(n^{-1/6})=o_p(1). \tag{2.3.6}
$$

联合假设1，2，5和引理 2.3.1(1) 得

$$\begin{aligned}
E\|A_{n1}\|^2 &\leq \sum_{l=1}^{k} E\|A_{n1,l}\|^2 \\
&= \frac{1}{n}\sum_{l=1}^{k} E\|\sum_{i=1}^{n}\frac{\partial g_i(\theta)}{\partial H_l}\sum_{j=1}^{n}\mathcal{W}_{nj}(T_i)(H_l(T_j)-H_l(T_i))\|^2 \\
&= \frac{1}{n}\sum_{l=1}^{k}\sum_{i=1}^{n} E\{E\{\|\frac{\partial g_i(\theta)}{\partial H_l}\|^2|T_i\}\{\sum_{j=1}^{n}\mathcal{W}_{nj}(T_i)(H_l(T_j)-H_l(T_i))\}^2\} \\
&\leq Ch^2r\sum_{l=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{n} E\{\mathcal{W}_{nj}^2(T_i)|\frac{T_j-T_i}{h}|^2\} \\
&= O_p(krh). \qquad (2.3.7)
\end{aligned}$$

类似 Wang et al.[25] 中引理 3 的证明过程，容易证明

$$\begin{aligned}
E(A_{n2,l}A_{n2,l}^{\tau}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\{(\frac{\partial g_i(\theta)}{\partial H_l})(\frac{\partial g_i(\theta)}{\partial H_l})^{\top}\mathcal{W}_{nj}^2(\varphi_l(X_j)-H_l(T_j))^2\} \\
&= O_p(\frac{r^2}{nh}). \qquad (2.3.8)
\end{aligned}$$

由 (2.3.8) 式可得

$$E\|A_{n2}\|^2 = O_p(\frac{kr^2}{nh}). \qquad (2.3.9)$$

联合假设 2，9 以及 (2.3.7)–(2.3.9) 有 $E\|A_{n1}\|^2 = o_p(1)$ 和 $E\|A_{n2}\|^2 = o_p(1)$ 成立。因此，

$$A_{n1} = o_p(1); \quad A_{n2} = o_p(1). \qquad (2.3.10)$$

结合 (2.3.5)–(2.3.10) 可得

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{g}_i(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i(\theta) + o_p(1).$$

前面完成了引理 2.3.3 (1) 的证明，下面开始证明引理 2.3.3 (2)。

根据 (2.3.5) 可得

$$\begin{aligned}\sum_{i=1}^{n}\hat{g}_i(\theta)\hat{g}_i^\top(\theta)=&\sum_{i=1}^{n}r_{i1}r_{i1}^\top+2\sum_{i=1}^{n}r_{i1}r_{i2}^\top+2\sum_{i=1}^{n}r_{i1}r_{i3}^\top+\sum_{i=1}^{n}r_{i2}r_{i2}^\top\\&+2\sum_{i=1}^{n}r_{i2}r_{i3}^\top+\sum_{i=1}^{n}r_{i3}r_{i3}^\top+2\sum_{i=1}^{n}r_{i1}\xi_i^\top\\&+2\sum_{i=1}^{n}r_{i2}\xi_i^\top+2\sum_{i=1}^{n}r_{i3}\xi_i^\top+\sum_{i=1}^{n}\xi_i\xi_i^\top\\=&R_1+2R_2+2R_3+R_4+2R_5\\&+R_6+2R_7+2R_8+2R_9+R_{10},\end{aligned}\tag{2.3.11}$$

其中：$\xi_i=O_p(\|\hat{H}(T_i)-H(T_i)\|^2)$，$r_{i1}=g_i$，

$$r_{i2}=\sum_{l=1}^{k}\frac{\partial g_i(\theta)}{\partial H_l}\sum_{j=1}^{n}\mathcal{W}_{nj}(T_i)(\varphi_l(X_j)-H_l(T_j)),$$

$$r_{i3}=\sum_{l=1}^{k}\frac{\partial g_i(\theta)}{\partial H_l}\sum_{j=1}^{n}\mathcal{W}_{nj}(T_i)(H_l(T_j)-H_l(T_i)).$$

$R_{2,st}$ 表示 R_2中第 s 行第 t 列位置的元素，也表示向量 r_{ij} 的第$s-$个分量，$j=1,2$。利用柯西-施瓦兹不等式得到下面不等式

$$|R_{2,st}|\le(\sum_{i=1}^{n}r_{i1,s}^2)^{1/2}(\sum_{i=1}^{n}r_{i2,t}^2)^{1/2}.$$

利用假设 2和5，类似于引理 2.3.2 的证明，我们有 $r_{i1,s}=o_p(n^{1/\alpha})$ 和

$$r_{i2,t}^2\le\sum_{l=1}^{k}\|\frac{\partial g_i(\theta)_t}{\partial H_l}\|^2\|\sum_{j=1}^{n}\mathcal{W}_{nj}(T_i)(\varphi_l(X_j)-H_l(T_j))\|^2=O_p(\frac{k}{nh}),$$

其中：$g_i(\theta)_t$ 表示向量 $g_i(\theta)$ 的第$t-$个分量，$t=1,2\cdots,r$。因此，我们有

$$|R_{2,st}|\le n\times o_p(n^{1/\alpha})\times O_p(\sqrt{\frac{k}{nh}})=o_p(n^{5/6}),$$

和

$$\|\frac{1}{n}R_2\|=\frac{r}{n}\times o_p(n^{5/6})=o_p(1).\tag{2.3.12}$$

同理可证：

$$\|\frac{1}{n}R_i\|=o_p(1),\ i=3,\cdots,10.\tag{2.3.13}$$

由 (2.3.11)–(2.3.13) 得

$$\frac{1}{n}\sum_{i=1}^{n}\hat{g}_i(\theta)\hat{g}_i^{\top}(\theta)=\frac{1}{n}\sum_{i=1}^{n}g_i(\theta)g_i^{\top}(\theta)+o_p(1).$$

综上所述，引理 2.3.3 (2) 证明完毕。 □

引理 2.3.4 在假设1-12成立下，$\lambda_{\theta_0}=\arg\max_{\lambda\in\hat{\Lambda}_n(\theta_0)}\tilde{l}(\lambda,\theta_0)$ 存在的概率趋向于1，而且$\|\lambda_{\theta_0}\|=O_p(\gamma_n)$，以及$\|\lambda_{\hat{\theta}}\|=O_p(\gamma_n)$。

证明： 本引理证明过程类似于 Leng 和 Tang (2012)[84] 中引理2和引理4的证明，因此省略。 □

下面开始证明本章给出的几个定理。

定理2.1.1和定理2.1.2的证明： 由引理 2.3.1–2.3.3 可知，被估计的经验似然函数中用核估计代替其中非参数部分不影响经验似然函数的渐近结果，因此定理2.1.1和定理2.1.2的剩下的证明完全类似于 Leng 和 Tang (2012)[84] 中定理1和定理2的证明，因此省略。 □

定理2.1.3的证明： 由假设8可以得出

$$\max_{1\le i\le n}\sup_{\theta\in\Theta_\theta}\|g_i(\theta)\|=o_p(n^{1/\alpha}r^{1/2}). \tag{2.3.14}$$

结合引理 2.3.3 和引理 2.3.4 以及 (2.3.14),我们有

$$\begin{aligned}\max_{1\le i\le n}|\lambda^{\top}\hat{g}_i(\theta_0)|&=\max_{1\le i\le n}|\lambda^{\top}g_i(\theta)|+o_p(1)\le\|\lambda\|\max_{1\le i\le n}\|g_i(\theta)\|+o_p(1)\\&=o_p(\gamma_n n^{1/\alpha}r^{1/2})=o_p(1).\end{aligned}$$

将等式 (2.1.9) 左边 Taylor 展开得

$$\begin{aligned}0&=\frac{1}{n}\sum_{i=1}^{n}\hat{g}_i(\theta_0)-\frac{1}{n}\sum_{i=1}^{n}\hat{g}_i(\theta_0)\hat{g}_i^{\top}(\theta_0)\lambda+R_{n1}\\&=\frac{1}{n}\sum_{i=1}^{n}g_i(\theta_0)-\frac{1}{n}\sum_{i=1}^{n}g_i(\theta_0)g_i^{\top}(\theta_0)\lambda+R_{n2},\end{aligned} \tag{2.3.15}$$

其中：$|\xi_i|\le\max_{1\le i\le n}|\lambda^{\top}\hat{g}_i(\theta_0)|$ 以及

$$R_{n1}=\frac{1}{n}\sum_{i=1}^{n}\hat{g}_i(\theta_0)\frac{(\lambda^{\top}\hat{g}_i(\theta_0))^2}{(1+\xi_i)^3},$$

$$R_{n2}=\frac{1}{n}\sum_{i=1}^{n}g_i(\theta_0)\frac{(\lambda^\top g_i(\theta_0))^2}{(1+\xi_i)^3}+o_p(1).$$

因为：$\max_{1\le i\le n}|\lambda^\top\hat{g}_i(\theta_0)|=o_p(1)$，所以：$\max_{1\le i\le n}|\xi_i|=o_p(1)$。联合 (2.3.14) 和 (2.3.15) 式，由引理 2.3.4 可以得到

$$\begin{aligned}\|R_{nl}\|&\le\|\lambda\|^2\max_{1\le i\le n}\|g_i(\theta)\|^3+o_p(1)\\&=o_p(\gamma_n^2 n^{3/\alpha}r^{3/2})=o_p(1),\quad l=1,2.\end{aligned}\tag{2.3.16}$$

令 $\bar{g}_n=\frac{1}{n}\sum\limits_{i=1}^{n}g_i(\theta_0)$和$S_n=\frac{1}{n}\sum\limits_{i=1}^{n}g_i(\theta_0)g_i^\top(\theta_0)$，由 (2.3.15)和(2.3.16) 式，我们有

$$\lambda=S_n^{-1}\bar{g}_n+S_n^{-1}R_{n2}.\tag{2.3.17}$$

首先将 $\tilde{l}(\theta_0)$ Taylor 展开，然后根据引理 2.3.3 得出，

$$\begin{aligned}\tilde{l}(\theta_0)=&2\sum_{i=1}^{n}\lambda^\top g_i(\theta_0)-\sum_{i=1}^{n}(\lambda^\top g_i(\theta_0))^2+\frac{2}{3}\sum_{i=1}^{n}(\lambda^\top g_i(\theta_0))^3(1+\eta_i)^{-4}+o_p(1)\\=&n\bar{g}_n^\top S_n^{-1}\bar{g}_n-nR_{n2}^\top S_n^{-1}R_{n2}+\frac{2}{3}\sum_{i=1}^{n}\frac{(\lambda^\top g_i(\theta_0))^3}{(1+\eta_i)^4}+o_p(1),\end{aligned}\tag{2.3.18}$$

其中：$|\eta_i|\le\max_{1\le i\le n}|\lambda^\top g_i(\theta_0)|$。由假设 8-10 可以推导出：当 $n\to\infty$ 时，S_n 依概率收敛于 $\Sigma(\theta_0)$。因此，联合 (2.3.17) 和 (2.3.18) 式，我们得到

$$\|nR_{n2}^\top S_n^{-1}R_{n2}\|=o_p(\sqrt{p})$$

和

$$\begin{aligned}|\frac{2}{3}\sum_{i=1}^{n}\frac{(\lambda^\top g_i(\theta_0))^3}{(1+\eta_i)^4}|&\le\frac{2}{3}\|\lambda\|^3\sum_{i=1}^{n}\|g_i(\theta_0)\|^3\{1+o_p(1)\}\\&=O_p(p^3n^{3/\alpha-1/2})=o_p(\sqrt{p}).\end{aligned}$$

类似于 Chen et al.(2009)[56] 中引理 5 和引理 6 的证明，容易得出：

$$(2p)^{-\frac{1}{2}}\{n\bar{g}_n^\top\Sigma(\theta_0)^{-1}\bar{g}_n-p\}\xrightarrow{d}N(0,1)$$

和

$$n\bar{g}_n^\top(\Sigma(\theta_0)^{-1}-S_n^{-1})\bar{g}_n=o_p(\sqrt{p}).$$

因此 (2.1.12) 式成立。综上所述，定理 2.1.3 证明完毕。 □

定理2.2.1的证明：根据 Newey 和 Smith (2004)[88]，(2.2.1) 给出的惩罚经验似然比函数在区域 D_n 中得到最小值存在。首先，对于任意的$\theta \in D_n$，由 (2.1.14) 和引理 2.3.2 得出

$$\max_{1\le i\le n} \|\lambda^{\top} g(\theta)\| = o_p(1).$$

然后，利用Taylor定理和引理 2.3.3 可得

$$\begin{aligned}
\frac{1}{n}\frac{\partial \tilde{l}_p(\theta)}{\partial \theta_j} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda^{\top}\partial \hat{g}_i(\theta)/\partial\theta_j}{1+\lambda^{\top}\hat{g}_i(\theta)} + P_{\nu}^{'}(|\theta_j|)\mathrm{sign}(\theta_j)\\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda^{\top}\partial g_i(\theta)/\partial\theta_j}{1+\lambda^{\top}g_i(\theta)} + o_p(1) + P_{\nu}^{'}(|\theta_j|)\mathrm{sign}(\theta_j)\\
&= \frac{1}{n}\sum_{i=1}^{n}\lambda^{\top}\{\frac{\partial g_i(\theta_0)}{\partial\theta_j} + \frac{\partial^2 g_i(\theta_0)}{\partial\theta_j\partial\theta^{\top}}(\theta-\theta_0)\} + o_p(1) + P_{\nu}^{'}(|\theta_j|)\mathrm{sign}(\theta_j)\\
&= A_1 + A_2 + P_{\nu}^{'}(|\theta_j|)\mathrm{sign}(\theta_j) + o_p(1).
\end{aligned}$$

再根据假设 6，我们有

$$\begin{aligned}
\max_{j\notin\mathcal{A}}(|A_1|) &= \max_{j\notin\mathcal{A}}\frac{1}{n}|\sum_{i=1}^{n}\lambda^{\top}\{E(\frac{\partial g(\theta_0)}{\partial\theta_j}) + (\frac{\partial g_i(\theta_0)}{\partial\theta_j} - E\frac{\partial g(\theta_0)}{\partial\theta_j})\}|\\
&\le \max_{j\notin\mathcal{A}}|\lambda^{\top}E(\frac{\partial g(\theta_0)}{\partial\theta_j})| + \frac{1}{n}\|\sum_{i=1}^{n}\lambda^{\top}(\frac{\partial g_i(\theta_0)}{\partial\theta_j} - E\frac{\partial g(\theta_0)}{\partial\theta_j})\|\\
&= o_p(1),
\end{aligned}$$

和

$$\begin{aligned}
\max_{j\notin\mathcal{A}}(|A_2|) &= \max_{j\notin\mathcal{A}}\frac{1}{n}|\sum_{i=1}^{n}\lambda^{\top}\{[E(\frac{\partial^2 g(\theta_0)}{\partial\theta_j\partial\theta^{\top}}) + (\frac{\partial^2 g_i(\theta_0)}{\partial\theta_j\partial\theta^{\top}} - E\frac{\partial^2 g(\theta_0)}{\partial\theta_j\partial\theta^{\top}})](\theta-\theta_0)\}|\\
&= o_p(1).
\end{aligned}$$

由于：$|\theta_j|_{\{j\notin\mathcal{A}\}} \le \gamma_n$，再利用假设11可以得到 $P_{\nu}^{'}(|\theta_j|)_{\{j\notin\mathcal{A}\}} = \nu$ 和

$$P_{\nu}^{'}(|\theta_j|)\mathrm{sign}(\theta_j)_{\{j\notin\mathcal{A}\}} = \nu\mathrm{sign}(\theta_j)_{\{j\notin\mathcal{A}\}}.$$

因此，当 $n\to\infty$ 时，对于任意的 $j\notin\mathcal{A}$，θ_j 的符号由$\frac{\partial\tilde{l}_p(\theta)}{\partial\theta_j}$确定的概率趋向于1，即：

$$\lim_{n\to\infty} p(\hat{\theta}_p^{(2)} = 0) = 1.$$

定理 2.2.1 (1) 的结论得证，下面开始证明定理 2.2.1 (2) 的结论。我们考虑在约束条件 $D_2\theta = 0$ 成立求下 (2.2.1) 式的最小值。根据Lagrange条件极值方法，约束条件 $D_2\theta = 0$ 成立求下 (2.2.1) 式的最小值等价于求如下不带约束条件的目标函数的最小值，

$$l_p(\theta, \lambda, \mu) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \lambda^\top \hat{g}_i(\theta)) + \sum_{j=1}^{p} p_\nu(|\theta_j|) + \mu^\top D_2\theta, \tag{2.3.19}$$

其中：$\lambda \in \mathbb{R}^p$ 和 $\mu \in \mathbb{R}^{(p-s)}$ 都是Lagrange乘数。由引理 2.3.3 可知，求 (2.3.19) 式得最小值等价于求下面目标函数的最小值，

$$\tilde{l}_p(\theta, \lambda, \mu) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \lambda^\top g_i(\theta)) + \sum_{j=1}^{p} p_\nu(|\theta_j|) + \mu^\top D_2\theta. \tag{2.3.20}$$

接下来的定理 2.2.1 (2) 的证明过程类似于 Leng 和 Tang (2012)[84] 中定理3的证明，因此省略。综上所述，定理 2.2.1 证明完毕。 □

定理2.2.2的证明：首先，我们给出 $\tilde{l}_p(\hat{\theta}_p)$ 的渐近展开式，其中 $\hat{\theta}_p$ 是使得 (2.2.1) 式达到最小的参数θ。令 $\hat{y}_i = \hat{\lambda}^\top \hat{g}_i(\hat{\theta}_p)$ 和 $y_i = \hat{\lambda}^\top g_i(\hat{\theta}_p)$，$i = 1, 2, \cdots, n$。由引理 2.3.4 和 (2.3.14) 式可得，$\max_{1\le i\le n}|y_i| = o_p(1)$。利用Taylor公式将 $\tilde{l}_p(\hat{\theta}_p)$ 展开，

$$\begin{aligned}
\tilde{l}_p(\hat{\theta}_p) &= \sum_{i=1}^{n} \hat{y}_i - \sum_{i=1}^{n} \frac{\hat{y}_i^2}{2} + \sum_{i=1}^{n} \frac{\hat{y}_i^3}{3(1+\zeta_i)^4 + o_p(1)} \\
&= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{y_i^2}{2} + \sum_{i=1}^{n} \frac{y_i^3}{3(1+\zeta_i)^4 + o_p(1)} + o_p(1),
\end{aligned} \tag{2.3.21}$$

其中：$|\zeta_i| \le |y_i|$。令：$Q_{1n}(\theta, \lambda) = n^{-1}\sum_{i=1}^{n} \frac{g_i(\theta)}{1+\lambda^\tau g_i(\theta)}$ 和 $\Sigma = \Sigma(\theta_0)$，由 (2.3.17)-(2.3.19) 可得 $\hat{\lambda}$ 的渐近表达式，

$$\hat{\lambda} = \{\Sigma^{-1} + \Sigma^{-1}G(V - VD_2^\top(D_2VD_2^\top)^{-1}D_2V)G^\top\Sigma^{-1}\}(Q_{1n}(\theta_0, 0) + o_p(1)).$$

令：$\bar{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\theta)$，由 (2.3.21) 可得惩罚经验似然比统计量的渐近表达式，

$$2\tilde{l}_p(\hat{\theta}_p) = n\bar{g}_n(\theta_0)^\top D_2^\top(D_2\Sigma D_2^\top)^{-1}D_2\bar{g}_n(\theta_0) + o_p(1). \tag{2.3.22}$$

因为 $L_nL_n^\top = I_q$，所以在原假设 $H_0 : L_n\theta_0 = 0$ 成立下，存在 $\widetilde{D}_2$，满足：$\widetilde{D}_2\theta = 0$ 和 $\widetilde{D}_2\widetilde{D}_2^\top = I_{p-d+q}$。由引理 2.3.3 知，使下式达到最小的 $\breve{\theta}$ 就

是 θ 的估计量，

$$\breve{l}_p(\theta,\lambda,\mu)=\frac{1}{n}\sum_{i=1}^{n}\log(1+\lambda^\top g_i(\theta))+\sum_{j=1}^{p}p_\nu(|\theta_j|)+\mu^\top\tilde{D}_2\theta. \tag{2.3.23}$$

用 $(\breve{\theta},\breve{\lambda},\breve{\mu})$ 表示使得 (2.3.23) 达到最小的参数 (θ,λ,μ)，类似于定理 2.1.1(1) 的证明可得

$$\lim_{n\to\infty}p(\breve{\theta}^{(2)}=0)=1.$$

因此，由假设12，我们有

$$\lim_{n\to\infty}p(n\{\sum_{j=1}^{p}p_\nu(|\hat{\theta}_{p,j}|)-\sum_{j=1}^{p}p_\nu(|\breve{\theta}_j|)\}=0)=1.$$

用 $\widetilde{D}_2$ 代替 (2.3.22) 式中的 D_2 得，

$$2\tilde{l}_p(\hat{\theta}_p)_{L_n\theta=0}=2\tilde{l}_p(\breve{\theta})=n\bar{g}_n(\theta_0)^\top\tilde{D}_2^\top(\tilde{D}_2\Sigma\tilde{D}_2^\top)^{-1}\tilde{D}_2\bar{g}_n(\theta_0)+o_p(1). \tag{2.3.24}$$

联合 (2.3.22)–(2.3.24)，我们有

$$\widetilde{l}_p(L_n)=n\bar{g}_n(\theta_0)^\tau\Sigma^{-1/2}(A_1-A_2)\Sigma^{-1/2}\bar{g}_n(\theta_0)+o_p(1),$$

其中：

$$P_1=\Sigma^{-1/2}GV\widetilde{D}_2^\tau(\widetilde{D}_2V\widetilde{D}_2^\tau)^{-1}\widetilde{D}_2VG^\tau\Sigma^{-1/2},$$
$$P_2=\Sigma^{-1/2}GVD_2^\tau(D_2VD_2^\tau)^{-1}D_2VG^\tau\Sigma^{-1/2}.$$

因为 A_1-A_2 是 q 阶单位矩阵，所以存在 $q\times p$ 阶矩阵 A_n，使得$A_nA_n^\top=I_q$，$A_1-A_2=A_n^\top A_n$。由Lindeberg 和 Feller 中心极限定理得

$$\sqrt{n}A_n\Sigma^{-1/2}\bar{g}_n(\theta_0)\xrightarrow{d}N(0,I_q).$$

所以：

$$n\bar{g}_n(\theta_0)^\tau\Sigma^{-1/2}(A_1-A_2)\Sigma^{-1/2}\bar{g}_n(\theta_0)\xrightarrow{d}\chi_q^2.$$

综述所述，定理 2.3.2 证明完毕。 □

§2.4 数值模拟

在这一节中，我们通过数值模拟验证本章所提出的方法的优良性质。

首先，我们给出求经验似然和惩罚经验似然估计的算法。由于高维经验似然函数的支撑是非凸的，这使得经验似然估计和经验似然比统计量的计算是非常复杂的，尤其是惩罚经验似然函数中的罚函数通常不可微的，因此惩罚经验似然估计和惩罚经验似然比统计量的计算更加复杂。本节模拟过程中，采用 Owen (2001)[4] 中的迭代最小二乘算法计算经验似然比统计量。对于惩罚经验似然比统计量的计算，采用 Fan 和 Li (2001)[69] 中提出的局部二次逼近算法 (LQA)。在计算经验似然和惩罚经验似然估计过程中，选择用 Owen (2001)[4] 中解决非线性优化问题的嵌套算法 (nested algorithm)分别最小化 (2.1.10) 和 (2.2.1)。在最小化 (2.2.1) 时，需要选择一个较好地初始值 θ^0 ，因为我们使用的算法的收敛性和收敛速度都和初始值有关。在求惩罚经验似然估计过程中，令 $\theta_j^{(k)}$ 为第$k-$次迭代结果，如果 $\theta_j^{(k)}$ 比较接近零，那么我们令 $\widehat{\theta}_j^{(k)}=0$。当 $\theta_j^{(k)}\neq 0$ 时，为使计算过程更简单，用下式 ($p_\nu(|\theta_j|)$ 的渐近表达式) 代替$p_\nu(|\theta_j|)$，

$$p_\nu(|\theta_j^{(k)}|)+\frac{1}{2}\{p'_\nu(|\theta_j^{(k)}|)/|\theta_j^{(k)}|\}\{\theta_j^2-(\theta_j^{(k)})^2\}.$$

在惩罚经验似然估计和惩罚经验似然比的计算过程中，惩罚函数 $p_\nu(|\theta_j|)$ 中的惩罚参数 ν 利用 Variyath et al.(2010)[95] 中给出的基于经验似然的贝叶斯信息准则(BIC)确定

$$BIC(\nu)=-2\tilde{l}_p(\theta_\nu)+df_\nu.\log(n),$$

其中：df_ν 为非零参数个数。

下面两个例子模拟说明分别在高维和高维稀疏情形下由 (2.1.1) 所确定的半参数模型的经验似然似然和惩罚经验似然推断方法的理论性质。

例 1： 考虑变系数部分线性模型：

$$Y_i=X_i^\top\theta+Z_i^\top u(T_i)+\varepsilon_i,\quad i=1,\cdots,n, \tag{2.4.1}$$

其中：Y_i 为响应变量，$\mathbf{X}_i=(X_i,Z_i,T_i)$ 为相应的协变量，$u(\cdot)=(u_1(\cdot),\cdots,u_{k_1}(\cdot))^\top$ 为未知光滑的q元回归函数，$\theta=(\theta_1,\theta_2,\cdots,\theta_p)^\top$ 为 p维回归系数，ε_i 为零均值有限方差且独立同分布的随机误差，满足：

$$E(\epsilon_i|X_i,Z_i,T_i)=0,\ a.s.$$

对于模型 (2.4.1) ，估计函数 $g(\cdot)$ 可表示为

$$g(\mathbf{X}, H(T), \theta) = (X - (M_3^{-1}M_1^\top)^\top Z)\{(Y - M_2M_3^{-1}Z) - (X - (M_3^{-1}M_1^\top)^\top Z)^\top\theta\},$$

其中：$\mathbf{X} = (X, Y, Z)$，$H(t) = (E(XZ^\top|T = t), E(YZ^\top|T = t), E(ZZ^\top|T = t))$，$M_1 = E(XZ^\top|T)$，$M_2 = E(YZ^\top|T)$ 以及 $M_3 = E(ZZ^\top|T)$。

在第二节中给出的假设条件6-10是相对保守的，针对具体变系数部分线性模型，有比假设条件6-10更弱的相应假设条件如下：

假设6'. $E(\varepsilon|X, Z, T) = 0$，a.s.。存在整数$d \geq 4$，使得$E(\|Z\varepsilon\|^d) < \infty$，$E(\|Z\|^d) < \infty$，$E(\|\varepsilon\|^d) < \infty$；

假设7'. 令$\Upsilon(\theta) = \varepsilon(X - M_3^{-1}M_1^\top Z)$，以及$\Upsilon(\theta)_t$是$\Upsilon(\theta)$的第$t$个分量，$t = 1, \cdots r$。存在整数$d \geq 4$以及正的常数$C_0$，当$n \to \infty$时，有$E(\|\Upsilon(\theta)/\sqrt{p}\|^d) < C_0$，$E(\|XZ^\top/\sqrt{p}\|^d) < C_0$, $E(\|M_3^{-1}M_1^\top ZZ^\top/\sqrt{p}\|^d) < C_0$ 和

$$\frac{1}{r}\sum_{t=1}^{r} E(|\Upsilon(\theta)_t|(\|XZ^\top/\sqrt{p}\|^4 + \|M_3^{-1}M_1^\top ZZ^\top/\sqrt{p}\|^4)) < C_0;$$

假设8'. 存在正的常数C_0，当$n \to \infty$时，有

$$\max_{1\leq t_1,t_2,t_3\leq r} E(\Upsilon(\theta)_{t_1}\Upsilon(\theta)_{t_2}\Upsilon(\theta)_{t_3})^2 < C_0;$$

假设9'. 令$\Sigma_1 = E(\varepsilon(X - M_3^{-1}M_1^\top Z)(X - M_3^{-1}M_1^\top Z)^\top)$。$\Sigma_1$的特征值非零且有限；

假设10'. $E(\varepsilon^3|X, Z, T) = 0$，a.s.。存在正整数 $d \geq 8$，使得 $p \to \infty$，$p^{3+2/(d-2)}/n \to 0$和$p/r \to c_0$ ($0 < c_0 < 1$) as $n \to \infty$。

在例1的模拟中，设协变量 T 服从均匀分布 $U(0, 1)$，协变量 X 服从多元正态分布 $N_p(0, \Sigma_p)$ ，其中，$(\Sigma_p)_{ij} = 0.2^{|i-j|}$，$1 \leq i, j \leq p$，协变量 $Z = (Z_1, Z_2)^\top$ 满足 $Z_1 = 1$ 和 $Z_2 \sim N(0, 1)$。非参数部分 $u(t) = (u_1(t), u_2(t))^\top$ 满足 $u_1(t) = sin\pi t$，$u_1(t) = 2t(1 - t)$。在核估计中，我们用Epanechnikov核函数，

$$K(t) = \frac{3}{4}(1 - t^2)_+.$$

核函数的窗宽 h 在满足假设2的前提下利用交叉验证方法确定。

令 $\theta = (1, 2, -1.5, 0, 2, 0, \cdots, 0)^\top$，考虑三种情形：$p = 10$，$p = 20$ 和 $p = 30$，相应的样本容量 $n = 200$，$n = 300$，$n = 400$ 和 $n = 500$。所有模拟重复1000次，其模拟结果见表 2.4.1−2.4.3。

表 2.4.1: 置信水平为0.9和0.95时EL和GLS(广义最小二乘)的置信域的覆盖概率

p	n	1-α=0.90		1-α=0.95	
		GLS	EL	GLS	EL
10	200	0.864	0.879	0.921	0.927
	300	0.872	0.884	0.926	0.934
	400	0.883	0.896	0.931	0.945
20	300	0.859	0.868	0.917	0.922
	400	0.867	0.873	0.924	0.931
	500	0.879	0.882	0.928	0.937
30	300	0.842	0.865	0.891	0.913
	400	0.856	0.871	0.895	0.923
	500	0.861	0.889	0.903	0.932

在显著性水平分别为 0.90 和 0.95 下，表 2.4.1 给出了分别利用经验似然和广义最小二乘(GLS)方法构造的参数 θ 的置信域的覆盖概率。广义最小二乘估计为：

$$\hat{\theta}_{GLS} = (\sum_{i=1}^{n} \hat{X}_i \hat{X}_i^\top)^{-1} (\sum_{i=1}^{n} \hat{X}_i \hat{Y}_i),$$

其中：$\hat{X}_i = X_i - (M_{3i}^{-1} M_{1i}^\top)^\tau Z_i$，$\hat{Y}_i = Y_i - M_{2i} M_{3i}^{-1} Z_i$，$M_{1i} = \hat{E}(X_i Z_i^\top | T_i)$，$M_{2i} = \hat{E}(Y_i Z_i^\top | T_i)$ 以及 $M_{3i} = \hat{E}(Z_i Z_i^\top | T_i)$。类似于 Lam 和 Fan (2008)[89]，可以证明GLS估计具有如下渐近分布，

$$\sqrt{n} B_n B \Sigma^{-1/2} (\hat{\theta}_{GLS} - \theta) \xrightarrow{d} N(0, G).$$

如果利用广义最小二乘法构造参数 θ 置信域，那么需要先给出 B 和 Σ的估计，$\hat{B} = \frac{1}{n} \sum\limits_{i=1}^{n} \hat{X}_i \hat{X}_i^\top$ 和 $\hat{\Sigma} = \frac{1}{n} \sum\limits_{i=1}^{n} \varepsilon_i^2 \hat{X}_i \hat{X}_i^\top$，其中，$\varepsilon_i = \hat{Y}_i - X_i^\top \hat{\theta}$。综上可得利用广义最小二乘方法构造的参数 θ 的置信域 $\mathcal{I}_{\alpha,GLS}(\theta)$ 为，

$$\mathcal{I}_{\alpha,GLS}(\theta) = \{\theta : n(\theta - \hat{\theta}_{GLS})^\top \hat{\Sigma}^{-1/2} \hat{B}^\top B_n^\top G^\top B_n \hat{B} \hat{\Sigma}^{-1/2} (\hat{\theta}_{GLS} - \theta) \le \chi^2_{p,(1-\alpha)}\}.$$

从表 2.4.1 可知，针对不同的维数 p，经验似然置信域的覆盖概率随样本容量 n 的增加而增加，当样本容量达到一定程度时，覆盖概率接近

表 2.4.2: EL和PEL估计的均值(标准差)的比较

p	n	Method	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_5$
10	200	EL	1.09(0.153)	1.92(0.261)	-1.42(0.248)	2.08(0.156)
		Bridge	0.93(0.186)	1.78(0.246)	-1.35(0.203)	1.87(0.195)
		Lasso	0.97(0.174)	1.80(0.248)	-1.47(0.182)	2.06(0.162)
		SCAD	0.99(0.128)	1.82(0.213)	-1.53(0.176)	2.03(0.173)
20	300	EL	1.08(0.210)	1.86(0.234)	-1.38(0.225)	2.11(0.231)
		Bridge	1.12(0.247)	1.72(0.289)	-1.42(0.254)	1.92(0.186)
		Lasso	1.10(0.236)	1.66(0.315)	-1.39(0.247)	1.93(0.193)
		SCAD	0.96(0.179)	1.79(0.232)	-1.43(0.218)	2.02(0.074)
30	400	EL	0.89(0.234)	1.83(0.257)	-1.33(0.273)	2.13(0.241)
		Bridge	0.87(0.287)	1.70(0.248)	-1.26(0.267)	1.75(0.258)
		Lasso	0.83(0.269)	1.74(0.251)	-1.31(0.271)	1.81(0.212)
		SCAD	0.86(0.245)	1.95(0.089)	-1.31(0.263)	1.87(0.187)

于给定的置信水平。另外，经验似然置信域的覆盖概率一致优于广义最小二乘法得到的置信域。

表 2.4.2 给出了非零参数的经验似然估计和三种惩罚函数 (Lasso，Bridge，SCAD) 下的惩罚经验似然估计的均值和标准差。从 2.4.2 可以看出，所有估计均值都比较接近真实参数，但是 SCAD 惩罚函数下的惩罚经验似然估计的标准差比其他几种估计的标准差更小。

各种惩罚函数下惩罚经验似然的变量选择结果总结在表 2.4.3。其结果显示：SCAD 惩罚函数下的惩罚经验似然方法得到的零参数个数比 Lasso 和 Bridge 惩罚函数下惩罚经验似然方法得到的零参数个数更接近真实的零参数的个数($p-4$)。这个结果进一步证明了我们提出的用 SCAD 惩罚函数下的惩罚经验似然方法对高维稀疏变系数部分线性模型进行变量选择具有优越性。

例 2: 考虑部分线性EV模型：

$$\begin{cases} Y_i = X_i^{\top}\theta + u(T_i) + \varepsilon_i, \\ Z_i = X_i + \vartheta_i, \quad i = 1, 2, \cdots, n, \end{cases} \tag{2.4.2}$$

表 2.4.3: 三种不同惩罚函数下PEL的变量选择结果

p	n	Method	Average number of zeros coefficients	
			Correct	Incorrect
10	200	Bridge	4.13[69%]	1.86
		Lasso	4.08[68%]	1.91
		SCAD	4.97[83%]	1.12
20	300	Bridge	12.89[81%]	3.24
		Lasso	13.93[87%]	2.08
		SCAD	14.87[93%]	1.35
30	400	Bridge	22.75[88%]	3.37
		Lasso	22.81[88%]	3.28
		SCAD	24.65[95%]	1.13

其中：Y_i 为响应变量，(T_i, Z_i) 为$(p \times \mathbb{R}^p)$上的可观测的随机向量，$X_i \in \mathbb{R}^p$ 是不可观测的协变量，θ 为 p维未知参数，$u(\cdot)$ 为未知函数，ϑ_i 为 p维测量误差变量， ε_i 为模型误差变量，满足：$E\{(\varepsilon_i, \vartheta_i^\top)^\top\} = 0$，$\mathrm{Cov}(\varepsilon_i, \vartheta_i) = \mathrm{diag}(\sigma^2, \Sigma_p)$，其中 σ^2 未知。考虑到模型的可识别性，假定 $\Sigma_p > 0$ 已知，ϑ_i 与 (X_i, Y_i, T_i) 相互独立。对于模型 (2.4.1)，估计函数 $g(\cdot)$ 可表示为

$$g(\mathbf{X}, H(T), \theta) = (Z - E(Z|T))\{(Y - E(Y|T)) - (Z - E(Z|T))^\top \theta\} + \Sigma_p \theta,$$

其中：$\mathbf{X} = (X, Y, Z)$，$H(t) = (E(Z|T = t), E(Y|T = t))$。当$\theta$为真实参数时有

$$E\{g(\mathbf{X}, H(T), \theta)\} = 0.$$

针对部分线性EV模型，本章第二节中的假设条件6-8可以放宽为：

假设6'. 存在$v_1(\mathbf{X}, T)$和常数C_2，使得

$$\frac{\partial g(\mathbf{X}, H(T), \theta)}{\partial \theta_l} < v_1(X, T), E\{v_1^2(\mathbf{X}, T)\} \le C_2 < \infty \quad (l = 1, 2, \cdots, p);$$

假设7'. $E(|\varepsilon|^4) < \infty$，$E(\|Z\|^4) < \infty$，$E(\|\vartheta\|^4) < \infty$；

假设8'. 令ϑ_{ij}为ϑ_i, $i = 1, \cdots, n$, $j = 1, \cdots, p$的第j个分量。存在正整数d，使得$|j_1 - j_2| > d$，$\mathrm{cov}(\vartheta_{ij_1}, \vartheta_{ij_2}) = 0$，$1 \le j_1, j_2 \le p$。

表 2.4.4: 置信水平为0.9和0.95时EL和ANS(统计量渐近正态性)的置信域的覆盖概率

p	n	1-α=0.90		1-α=0.95	
		ANS	EL	ANS	EL
20	500	0.848	0.854	0.895	0.903
	700	0.875	0.872	0.904	0.917
	900	0.881	0.885	0.932	0.936
30	800	0.845	0.849	0.871	0.873
	900	0.862	0.867	0.896	0.908
	1000	0.879	0.882	0.918	0.929

表 2.4.5: EL和SCAD-PEL估计的均值(标准差)的比较

p	n	Method	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_4$	$\hat{\theta}_5$
20	500	EL	2.15(0.272)	-1.67(0.241)	1.34(0.256)	2.06(0.132)
		SCAD	1.76(0.236)	-1.58(0.164)	0.93(0.225)	1.98(0.123)
30	1000	EL	2.18(0.253)	-1.75(0.279)	1.22(0.293)	2.15(0.298)
		SCAD	1.87(0.214)	-1.72(0.262)	1.26(0.281)	1.83(0.274)

在例2的模拟中，核函数和窗宽的选择类似于例1。协变量 T 服从均匀分布 $U(0,1)$，协变量 X 服从多元正态分布 $N_p(0,\Sigma_p)$ ，其中，$(\Sigma_p)_{ij} = 0.5^{|i-j|}$，$1 \le i,j \le p$，测量误差变量 ϑ 服从多元正态分布 $N_p(0,\Sigma_p)$ ，其中，$(\Sigma_p)_{ij} = 0.3^{|i-j|}$，$u(t) = \sin(2\pi t)$。

令 $\theta = (2,-1.5,0,1,2,0,\cdots,0)^\top$，考虑两种情形：$p = 20$ 和 $p = 30$，相应的样本容量 $n = 500$，$n = 700$ 和 $n = 1000$。所有模拟重复1000次，其结果总结见表 2.4.4–2.4.6。

为了比较经验似然置信域的覆盖概率，我们也给出了基于渐近正态分布的置信域。类似于 Chen，Zhong 和 Cui (2009)[90]，可以得到 θ 的矩估计，

$$\hat{\theta}_M = [\sum_{i=1}^n (Z_i-\hat{E}(Z_i|T_i))(Z_i-\hat{E}(Z_i|T_i))^\top - n\Sigma_p]^{-1}\sum_{i=1}^n (Z_i-\hat{E}(Z_i|T_i))(Y_i-\hat{E}(Y_i|T_i)),$$

表 2.4.6: 三种不同惩罚函数下PEL的变量选择结果

p	n	Method	Average number of zeros coefficients	
			Correct	Incorrect
20	500	Bridge	13.72[85%]	2.34
		Lasso	12.41[78%]	3.65
		SCAD	14.16[89%]	1.27
30	800	Bridge	23.24[89%]	2.58
		Lasso	21.69[83%]	4.31
		SCAD	24.73[95%]	1.46

以及 $\hat{\theta}_M$ 的渐近分布

$$\sqrt{n}^{-1/2}(\theta - \hat{\theta}_M) \xrightarrow{d} N(0, \Sigma_X^{-1}\Omega\Sigma_X^{-1}).$$

Ω 和 Σ_X 的一致估计分别为，

$$\hat{\Omega} = n^{-1}\sum_{i=1}^{n}\{(Z_i - \hat{E}(Z_i|T_i))((Y_i - \hat{E}(Y_i|T_i)) - (Z_i - \hat{E}(Z_i|T_i))^\top\hat{\theta}_M) + \Sigma_p\hat{\theta}_M\}^{\otimes 2},$$

和

$$\hat{\Sigma}_X = \frac{1}{n}\sum_{i=1}^{n}\{(Z_i - \hat{E}(Z_i|T_i))(Z_i - \hat{E}(Z_i|T_i))^\top - n\Sigma_p\},$$

其中：$A^{\otimes 2} = AA^\top$。综上可得基于渐近正态性的参数θ的置信域如下：

$$\mathcal{I}_{\alpha,ASN}(\theta) = \{\theta : \{n(\theta - \hat{\theta}_M)^\top\hat{\Sigma}_X\hat{\Omega}^{-1}\hat{\Sigma}_X(\theta - \hat{\theta}_M)\} \le \chi^2_{p,(1-\alpha)}\}.$$

表 2.4.4 给出了两种方法得到的置信域的覆盖概率模拟结果。其结果显示：经验似然置信域有更高的覆盖概率，而且，随着样本容量的增加置信域的覆盖概率越来越接近置信水平。非零参数 $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_4, \hat{\theta}_5)$ 的经验似然估计和基于SCAD惩罚函数的惩罚经验似然估计结果总结在表 2.4.5 中，可以看出，在高维稀疏下，基于SCAD惩罚函数的惩罚经验似然估计优于经验似然估计。最后，表 2.4.6 总结了各种惩罚函数下惩罚经验似然的变量选择结果，SCAD 惩罚函数下的惩罚经验似然方法的变量选择结果优于Lasso 和 Bridge 惩罚函数下惩罚经验似然方法。

表 2.5.1: SCAD惩罚函数下的惩罚经验似然方法变量选择结果及其系数估计

基因的probe IDs	SCAD-PEL估计
31536_at	-0.0935
36131_at	-0.0506
37761_at	0.1104
39837_s_at	-0.1191
40718_at	-0.0669
754_s_at	0.0916

§2.5 实例分析

下面将我们提出的方法应用到实际数据中去。急性淋巴细胞性白血病 (ALL) 数据 (http://www.bioconductor.org/)，包含128个白血病病人的T−细胞或B−细胞基因表达数据。Dudoit，Keles 和 Vanderlaan (2008)，Chen 和 Qin (2010) 分别在 [91] 和 [92] 中分析过这些数据。一个比较有实际意义的问题是如何利用这些基因表达数据对样本进行分类。我们考虑ALL数据中79个急性淋巴细胞性白血病病人的 B−细胞基因表达数据 (37 个 BCR/ABL 分子类型和 42 个 NEG 分子类型的 B−细胞基因表达数据)，利用部分线性逻辑斯蒂回归模型和本章提出的方法对其进行分析。由于样本的基因表达数据维数 p 高达12625，远远超过样本容量 n，我们先对其进行过滤：(1) 至少 75% 的样本被测得的强度达到 100 以上；(2) 样本间的强度变异系数处于 0.7 至 10 之间。79 个样本的基因表达数据过滤之后的维数 $p = 2396$。为了考虑年龄因素的影响，我们将其作为非参数部分的协变量引入到部分线性逻辑斯蒂回归回归模型。在79个样本中，有3个样本没有年龄信息，因此，我们最终考虑具有完整年龄信息的 76 个样本。通过基因过滤后的样本数据维数 $p = 2396$，$p > n$，因此，我们还需要利用 Fan 和 Lv (2008) [86] 的 SIS 方法降维，使得 $p < n$ 成立。通过 SIS，我们得到边际影响最大的 15 个基因表达数据。

利用本章提出的惩罚经验似然方法，对基因表达数据进行变量选择，得到了 6 个对急性淋巴细胞性白血病病人分类影响显著的 6 个基因变量以及变量的估计值，其结果如表 2.5.1 所示。

第三章 高维可加危险率模型的经验似然推断

生存分析是统计学重要研究领域，而可加危险率模型又是生存分析中最常见的模型之一，该模型由Aalen (1989)[93] 提出，其形式如下：

$$h(t|Z_j(t)) = h_0(t) + \sum_{k=1}^{p} \beta_k Z_{jk}(t), \quad j = 1, \cdots, n, \tag{3.1.1}$$

其中$Z(t) = (Z_1(t), Z_2(t), \cdots, Z_p(t))^\top$ 为协变量，$h(t|Z(t))$ 为危险率函数，$h_0(t)$ 为基准危险率函数，$\beta(t) = (\beta_1(t), \beta_2(t), \cdots, \beta_p(t))^\top$ 为回归系数。回归系数$\beta(t)$ 是关于时间t的函数，在实际应用中比较复杂，Lin 和 Yang (1994) [94] 提出了一种更简单的可加危险率模型：

$$h(t \mid Z_j(t)) = h_0(t) + \sum_{k=1}^{p} \beta_{jk} Z_k(t), \quad j = 1, \cdots, n, \tag{3.1.2}$$

其中回归系数 β_k 是常数，实际应用中更加方便。

本章，我们将考虑高维删失情形下模型 (3.1.2) 的经验似然推断问题。当 $n \to \infty$，$p \to \infty$ 时，首先，证明了在一定条件下，经验似然比渐近分布为正态分布以及通过经验似然方法得到的参数估计量具有一致性；其次，我们利用惩罚经验似然方法研究了高维稀疏情形下可加危险率模型 (3.1.2) 的变量选择和参数估计问题，其惩罚经验似然比统计量具有渐近 χ_q^2 分布，且惩罚经验似然方法具有变量选择的Oracle性质。

本章内容结构如下：第一节介绍高维情形下模型 (3.1.2) 的经验似然推断方法及其主要结论；第二节，利用惩罚经验似然研究删失情形下高维稀疏模型 (3.1.2) 的变量选择、参数估计和假设检验问题以及相应的一些结果；第三节为定理的证明；第四和第五节分别为有限样本情形下的数值模拟结果和实例分析。

§3.1 高维可加危险率模型的经验似然

设 $Z(t) = (Z_1(t), Z_2(t), \cdots, Z_p(t))^\top$ 为协变量， T 和 C 分别为事件发生时间和删失时间，关于 T 的危险率函数由模型 (3.1.2) 给定。定义$X = \min\{T, C\}$，$\triangle = I(T \leq C)$，其中$I(\cdot)$ 为示性函数。假设在给定 $Z(\cdot)$ 条件

下 T 和 C 相互独立。分别定义左连续计数过程 $Y_i(t)$、右连续计数过程 $N_i(t)$ 和 计数过程鞅 $M_i(t)$

$$Y_i(t) = I(X_i \geq t), \quad i = 1, \cdots, n,$$

$$N_i(t) = I(X_i \leq t, \triangle_i = 1),$$

$$M_i(t) = N_i(t) - \int_0^{\top} Y_i(s)\{h_0(s) + \beta^{\top} Z_i(s)\}ds.$$

Lin 和 Yang (1994) [94] 给出了模型 (3.1.2) 的拟得分函数

$$U(\beta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t) - \bar{Z}(t)\}\{dN_i(t) - Y_i(t)\beta^{\top}Z_i(t)dt\}, \tag{3.1.3}$$

其中 $\beta \in \mathbb{R}^p$，$\bar{Z}(t) = \sum_{i=1}^n Y_i(t)Z_i(t)/\sum_{i=1}^n Y_i(t)$，$\tau$ 为最大观测时间且满足 $P(X_i > \tau) > 0$。由 (3.1.3) 得到部分得分方程

$$\begin{aligned} S_n(\beta) &= \sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t) - \bar{Z}(t)\}\{dN_i(t) - Y_i(t)\beta^{\top}Z_i(t)dt\} \\ &= \sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t) - \bar{Z}(t)\}dM_i(t) \\ &= 0. \end{aligned} \tag{3.1.4}$$

根据 (3.1.4)，我们可以合理假设

$$\sum_{i=1}^{n} w_i \int_0^{\tau}\{Z_i(t) - \bar{Z}(t)\}d\hat{M}_i(t) = 0, \tag{3.1.5}$$

其中$\sum\limits_{i=1}^{n} w_i = 1,\ w_i \geq 0$，$\hat{M}_i(t)$ 为 $\Lambda_0(t) = \int_0^t h_0(s)ds$ 的 Aalen−Breslow 估计，

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s)\{\beta^{\top}Z_i(s)ds + d\hat{\Lambda}_0(s)\},$$

$$\hat{\Lambda}_0(t) = \int_0^{\top}\frac{\sum_{i=1}^n\{dN_i(s) - Y_i(s)\beta^{\top}Z_i(s)ds\}}{\sum_{i=1}^n Y_i(s)}.$$

定义辅助随机向量 η_i，$i = 1, \cdots, n$，

$$\begin{aligned} \eta_i \triangleq \eta_i(\beta) &= \int_0^{\tau}\{Z_i(t) - \bar{Z}(t)\}d\hat{M}_i(t) \\ &= \int_0^{\tau}\{Z_i(t) - \bar{Z}(t)\}dM_i(t) \\ &= \int_0^{\tau}\{Z_i(t) - \bar{Z}(t)\}\{dN_i(t) - Y_i(t)\beta^{\top}Z_i(t)dt\}. \end{aligned}$$

容易证明，$\eta_i(\beta)$ 是均值为零的鞅。对于$1 \le i \le n$，令

$$b_i = \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dN_i(t),$$

$$V_i = \int_0^\tau Y_i(t)\{Z_i(t) - \bar{Z}(t)\}\{Z_i(t) - \bar{Z}(t)\}^\top dt.$$

因此，$\eta_i(\beta)$ 可以表示为

$$\eta_i(\beta) = b_i - V_i\beta.$$

为了方便下面的定理及其证明过程的叙述，我们引入下面的记号

$$D = E[\int_0^\tau Y(t)\{Z(t) - e(t)\}^{\otimes 2} dt],$$

$$\Sigma = E[\int_0^\tau Y(t)\{Z(t) - e(t)\}^{\otimes 2}\{h_0(t) + \beta^\top Z(t)\} dt],$$

$$b = \frac{1}{n}\sum_{i=1}^n b_i = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dN_i(t),$$

$$V = \frac{1}{n}\sum_{i=1}^n V_i = \frac{1}{n}\sum_{i=1}^n \int_0^\tau Y_i(t)\{Z_i(t) - \bar{Z}(t)\}^{\otimes 2} dt,$$

其中 $w^{\otimes 0} = 1, w^{\otimes 1} = w, w^{\otimes 2} = ww^\top, s^{(k)}(t) = E\{Y(t)Z(t)^{\otimes k}\}, 0 \le k \le 2, e(t) = s^{(1)}(t)/s^{(0)}(t)$。由 D 和 Σ 的定义，有

$$D = E(V), \qquad \Sigma = E(\frac{1}{n}\sum_{i=1}^n \eta_i(\beta)\eta_i(\beta)^\top).$$

关于参数 β 的经验似然比函数为

$$R(\beta) = \sup\{\prod_{i=1}^n nw_i | w_i \ge 0, \sum_{i=1}^n w_i = 1, \sum_{i=1}^n w_i\eta_i(\beta) = 0\}. \tag{3.1.6}$$

由 Lagrange 乘数法，其对数经验似然比函数为

$$l(\beta) = 2\sum_{i=1}^n \log(1 + \lambda^\top \eta_i(\beta)), \tag{3.1.7}$$

其中 $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_p)^\top$ 为 Lagrange 乘数，满足如下方程

$$\frac{1}{n}\sum_{i=1}^n \frac{\eta_i(\beta)}{1 + \lambda^\top \eta_i(\beta)} = 0. \tag{3.1.8}$$

当 $n \to \infty$，$p \to \infty$ 时，$\|\lambda\| = O_p(n^{-\frac{1}{2}})$ 不再成立。为了得到样本维数发散条件下对数经验似然比统计量的渐近分布，需要假设条件如下：

假设1. 协变量 $Z(t) = (Z_1(t), Z_2(t), \cdots, Z_p(t))^\top$ 的支撑有界，对于 $1 \le k \le p$，$Z_k(t)$ 的边际密度函数 $f(t)$ 连续而且满足：

$$0 < \inf f(t) \le \sup f(t) < \infty;$$

假设2. 对于 $1 \le i \le n$，$P(C > \tau | Z(\cdot)) > 0$ 和 $P(\Delta_i = 1 | Z(\cdot)) > 0$ 成立。

假设3. $V_1, V_2, \cdots, V_n, \Sigma$ 和 D 的所有特征值都为有限正数。

假设4. 当 $n \to \infty, p \to \infty$ 时，$pn^{-(1/5)} \to 0.$

假设5. $\int_0^\tau \lambda_0(t)dt < \infty.$

使得 (3.1.7) 达到最小的 β，称为参数 β 的经验似然估计，记为 $\hat{\beta}$。如下定理给出了维数发散条件下经验似然估计 $\hat{\beta}$ 的理论性质。

定理 3.1.1 在假设 1-5 成立下，如果参数 β 的真实值为 β_0，那么

$$\lim_{n \to \infty} p(\|\hat{\beta} - \beta_0\| = O_p\{(p/n)^{1/2}\}) = 1.$$

定理 3.1.1 表明参数 β 的经验似然估计 $\hat{\beta}$ 具有渐近一致性。下面我们给出对数经验似然比统计量的渐近分布。

定理 3.1.2 令参数 β 的真实值为 β_0，在假设 1-5 成立下，当 $n \to \infty$ 时，有

$$(2p)^{-\frac{1}{2}}(l(\beta_0) - p) \xrightarrow{d} N(0, 1). \tag{3.1.9}$$

定理 3.1.2 的结论可以用来检验假设：

$$H_0 : \beta = \beta_0\ , \quad H_1 : \beta \ne \beta_0.$$

另外，定理 3.1.2 的结论还可以用来构造参数β的置信域。令

$$I_\alpha(\beta) = \{\beta : \tilde{l}(\beta) \le p + z_\alpha \sqrt{2p}\},$$

其中 z_α 是标准正态分布的上 $\alpha-$分位数。参数 β 的渐近置信水平为 $1-\alpha$ 的置信域为$I_\alpha(\beta)$，即

$$P(\beta \in I_\alpha(\beta)) = 1 - \alpha + o_p(1).$$

在实际应用中，特别是在高维情形下，很少考虑整个参数向量的置信域，因为只要 $p > 3$，很难想象或表示参数向量的置信域的形状。一

般情况下，我们只考虑参数向量的一个(或两个)分量的置信区间(置信域)。假设 $\beta=(\beta^{(1)\top},\beta^{(2)\top})^\top$，其中 $\beta^{(1)}$ 是 β 的 $k-$维分量(k 固定)。我们考虑构造 $\beta^{(1)}$ 的置信区间(置信域)。真实参数向量作相应划分：$\beta_0=(\beta_0^{(1)\top},\beta_0^{(2)\top})^\top$，类似地，$Z_i=(Z_i^{(1)\top},Z_i^{(2)\top})^\top$，$\bar{Z}=(\bar{Z}^{(1)\top},\ \bar{Z}^{(2)\top})^\top$，$b_i=(b_i^{(1)\top},b_i^{(2)\top})^\top$，$V_{1i}=(V_i^{(1)},V_i^{(2)})$，其中 V_{1i} 由矩阵 V_i 的前面 k 组成。

对于固定的 $\beta^{(1)}$，通过最小化 (3.1.7) 得到 $\beta^{(2)}$ 的估计量 $\hat{\beta}^{(2)}$ 。定义辅助随机向量

$$\tilde{\eta}_i(\beta)=b_i^{(1)}-V_i^{(1)}\beta^{(1)}-V_i^{(2)}\hat{\beta}^{(2)}.$$

关于参数分量 $\beta^{(1)}$ 的对数经验似然比表示为

$$\tilde{l}(\beta^{(1)})=2\sum_{i=1}^{n}log\{1+\lambda^{(1)\top}\tilde{\eta}_i(\beta)\},\tag{3.1.10}$$

其中 $\lambda^{(1)}\in\mathbb{R}^k$ 是 Lagrange 乘数。下面定理给出了关于参数分量 $\beta^{(1)}$ 的对数经验似然比统计量的渐近分布。

定理 3.1.3 在定理 3.1.2 的条件下，当 $n\to\infty$ 时，有

$$\tilde{l}(\beta_0^{(1)})\xrightarrow{d}\chi_k^2.$$

根据定理 3.1.3 ，参数分量 $\beta^{(1)}$ 的渐近置信区间(置信域)为$I_\alpha(\beta^{(1)})$

$$I_\alpha(\beta^{(1)})=\{\beta^{(1)}\in\mathbb{R}^k|\tilde{l}(\beta^{(1)})\le c_\alpha\},$$

其中 c_α 为临界值，对于任意的 $0<\alpha<1$ 都有 $p(\chi_k^2>c_\alpha)$成立。

另外，定理 3.1.3 还可以用来检验假设：

$$H_0:L_n\beta_0^{(1)}=0\ ,\quad H_1:L_n\beta_0^{(1)}\neq 0,$$

其中$L_n\in\mathbb{R}^{k\times p}$ 满足 $L_nL_n^\top=I_k$，k 为固定常数。

§3.2 高维稀疏可加危险率模型的惩罚经验似然

在高维数据分析中，变量选择有非常重要的意义。本节我们提出利用惩罚经验似然方法实现高维稀疏可加危险率模型的变量选择。定义惩罚经验似然比为：

$$l_p(\beta)=\sum_{i=1}^{n}\log(1+\lambda_\beta^\top\eta_i(\beta))+n\sum_{j=1}^{p}p_\nu(|\beta_j|),\tag{3.2.1}$$

其中：$\lambda_\beta = (\lambda_{\beta 1}, \lambda_{\beta 2}, \cdots, \lambda_{\beta p})^\top$为Lagrange乘数，$p_\nu(|\beta_j|)$为惩罚函数，$\nu$是平衡模型复杂程度与偏误的规范化参数。

常见的惩罚函数有：Frank和Friedman(1993)在[67]中提出的桥(Bridge)惩罚函数(L_q惩罚)；Tibshirani (1996) 在 [68] 中提出的Lasso惩罚函数；Fan 和 Li (2001) 在 [69] 中提出的光滑切片绝对偏差惩罚函数(SCAD)等。我们利用带有SCAD惩罚函数的惩罚经验似然方法实现高维稀疏可加危险率模型的变量选择。

记$\mathcal{A} = \{j : \beta_{oj} \neq 0\}$，$\mathcal{A}$是由真实参数向量$\beta_0$的非零分量组成的集合，$\mathcal{A}$的基数($|\mathcal{A}| = d$)未知，$d$有限或随$n$一起趋向无穷。不失一般性，设$\beta = (\beta^{(1)\top}, \beta^{(2)\top})^\top$，其中$\beta^{(1)} \in \mathbb{R}^d$ 和$\beta^{(2)} \in \mathbb{R}^{p-d}$分别表示真实参数向量$\beta_0$的非零分量和零分量，例如：$\beta_0 = (\beta_0^{(1)\top}, 0)^\top$。为了方便表示，定义：$I_p = (H_1^\top, H_2^\top)$，其中$H_1 \in \mathbb{R}^{d\times p}$，$H_2 \in \mathbb{R}^{(p-d)\times p}$。$\|A\|$表示矩阵$A$的Frobenius范数，$\|A\| = \{tr(A^\top A)\}^{\frac{1}{2}}$。

为了得到惩罚经验似然估计的Oracle性质，需要如下假设：

假设6. 当$n \to \infty$时, 规范化参数ν满足$\nu(p/n)^{\frac{1}{2}} \to \infty$ 和$\min_{j\in\mathcal{A}} \beta_{0j}/\nu \to \infty$；

假设7. $\max_{j\in\mathcal{A}} P'_\nu(|\beta_{0j}|) = o\{(np)^{-\frac{1}{2}}\}$，$\max_{j\in\mathcal{A}} P''_\nu(|\beta_{0j}|) = o\{(p)^{-\frac{1}{2}}\}$。

使(3.2.1)达到最小的β称为惩罚经验似然估计，记为$\hat\beta$。令$\hat\beta = (\hat\beta_1^\top, \hat\beta_2^\top)^\top$，下面定理给出了带有SCAD惩罚函数的惩罚经验似然估计$\hat\beta$的Oracle性质。

定理 3.2.1 在假设$1-7$成立下, 当$n \to \infty$时, $\hat\beta$满足：

(1). (稀疏性) 以概率1, 有$\hat\beta^{(2)} = 0$;

(2). (渐近正态性)

$$\sqrt{n}A_n I_{\mathcal{A}}^{-1/2}(\hat\beta^{(1)} - \beta_0^{(1)}) \xrightarrow{d} N(0, G),$$

其中：$A_n \in \mathbb{R}^{q\times p}$满足$A_n A_n^\top \to G$，$G \in \mathbb{R}^{q\times q}$，$q$固定，$A = \mathbf{D}^\top\Sigma^{-1}\mathbf{D}$，$I_{\mathcal{A}} = H_1 A^{-1}\Sigma A^{-1} H_1^\top - \Gamma_{12}\Gamma_{22}^{-1} H_2 A^{-1}\Sigma A^{-1} H_2^\top \Gamma_{22}^{-1}\Gamma_{21}$，$\Gamma_{ij} = H_i \Sigma^{-1} H_j^\top$，$i, j = 1, 2$.

定理3.2.1(1)显示，在估计参数零分量时，惩罚经验似然估计方法比经验似然估计方法更有效。$A_n(\hat\beta^{(1)} - \beta_0^{(1)})$表示维数($d-$维)发散的向量($\hat\beta^{(1)} - \beta_0^{(1)}$)的投影，投影向量维数($q-$维)有限，定理3.2.1(2)显示其极限分布为多元正态分布，即惩罚经验似然估计的任意有限维投影的具有渐近有效性。

下面考虑关于参数β的假设检验。考虑假设：

$$H_0: \quad L_n\beta_0 = 0\ , \qquad H_1 \quad L_n\beta_0 \neq 0,$$

其中：$L_n \in \mathbb{R}^{q\times d}$，$L_n L_n^\top = I_q$，$q$固定，$I_q$为$q-$阶单位矩阵。根据(3.2.1)式，构造如下惩罚经验似然比检验统计量：

$$\tilde{l}_p(L_n) = -2\{l_p(\hat{\beta}) - \min_{\beta, L_n\beta_0=0} l_p(\beta)\}.$$

下面定理给出了检验统计量$\tilde{l}_p(L_n)$的极限分布。

定理 3.2.2 在原假设H_0和假设条件$1-7$成立下，当$n\to\infty$时，

$$K_{B_n}^{-\frac{1}{2}} \widetilde{l}_p(L_n) K_{B_n}^{-\frac{1}{2}} \xrightarrow{d} \chi_q^2,$$

其中：$K_{B_n} = B_n A^{-\frac{1}{2}} \Sigma A^{-\frac{1}{2}} B_n^\top$，$B_n$为$q\times p$矩阵，满足$B_n B_n^\top = I_q$。

根据定理3.2.2，参数β的置信度为$(1-\alpha)100\%$的渐近置信域为：

$$V_\alpha(L_n\beta) = \{\beta : -2K_{B_n}^{-\frac{1}{2}}\{l_p(\hat{\beta}) - \min_{\beta, L_n\beta_{10}=0} l_p(\beta)\}K_{B_n}^{-\frac{1}{2}} \le \chi^2_{q,(1-\alpha)}\}.$$

即：当$n\to\infty$时，$p(L_n\beta_0 \in V_\alpha(L_n\beta)) \to 1-\alpha$。

§3.3 定理的证明

在这一节中，我们将给出本章几个定理的证明。整个证明过程中，不同地方出现的C的具体的值可能不同，但都表示一个正的常数。令：$a_n = c(p/n)^{\frac{1}{2}}$，$D_n = \{\beta \mid \|\beta - \beta_0\| \le a_n\}$。证明定理之前，我们先给出下面引理。

引理 3.3.1 如果假设1-5成立，那么：

$$(1)\ \ \max_{1\le i\le n} \|\eta_i(\beta)\| = O_p(\sqrt{p});\quad (2)\ \ \|\lambda\| = O_p(a_n).$$

证明： 类似于Lin和Lv(2013)[79]中引理A.3，在假设1，2和5成立下，存在常数$C>0$，有

$$P(|\eta_{ij}(\beta)| > C) \to 0,$$

其中：$\eta_{ij}(\beta)$是$\eta_i(\beta)$的第j个分量，$i=1,\cdots,n$，$j=1,\cdots,p$。因此，容易得出：

$$\max_{1\le i\le n} \|\eta_i(\beta)\| = O_p(\sqrt{p}).$$

引理3.3.1 (1)的结论得证，下面开始证明引理3.3.1 (2)。根据(3.1.8)式，存在$\lambda \in \mathbb{R}^p$，满足：

$$0 = \frac{1}{n}\sum_{i=1}^n \frac{\eta_i(\beta)}{1+\lambda^\top \eta_i(\beta)} =: \phi(\lambda).$$

令：$\lambda = \alpha\theta$，其中：$\alpha \geq 0$，$\theta \in \mathbb{R}^p$，$\|\theta\| = 1$。将$1/(1+\lambda^\top\eta_i(\beta)) = 1 - \lambda^\top\eta_i(\beta)/(1+\lambda^\top\eta_i(\beta))$代入$\theta^\top\phi(\lambda) = 0$可以得到：

$$|\theta^\top\eta_i(\beta)| \geq \frac{\alpha}{1+\alpha \max\limits_{1\leq i\leq n}\|\eta_i(\beta)\|}\theta^\top S_n\theta,$$

其中：$S_n = \frac{1}{n}\sum\limits_{i=1}^{n}\eta_i(\beta)\eta_i(\beta)^\top$。由于$\frac{1}{n}\frac{1}{1+\lambda^\top\eta_i(\beta)}$是概率测度，我们有

$$0 < 1+\lambda^\top\eta_i(\beta) \leq 1+\alpha \max_{1\leq i\leq n}\|\eta_i(\beta)\|.$$

因此：

$$\alpha[\theta^\top S_n\theta - \theta^\top\overline{\eta}_i(\beta)\max_{1\leq i\leq n}\|\eta_i(\beta)\|] \leq |\theta^\top\overline{\eta}_i(\beta)|, \tag{3.3.1}$$

其中：$\overline{\eta}_i(\beta) = \frac{1}{n}\sum\limits_{i=1}^{n}\eta_i(\beta)$。由引理3.3.1 (1)可得

$$|\theta^\top\overline{\eta}_i(\beta)| \leq \|\overline{\eta}_i(\beta)\| = O_p(\sqrt{p/n}). \tag{3.3.2}$$

根据(3.3.2)式和假设4，我们有

$$|\theta^\top\overline{\eta}_i(\beta)|\max_{1\leq i\leq n}\|\eta_i(\beta)\| = o_p(1). \tag{3.3.3}$$

由(3.3.1)式和(3.3.3)式，有

$$|\alpha[\theta^\top S_n\theta + o_p(1)] = O_p(\sqrt{p/n}). \tag{3.3.4}$$

类似于Li et al. (2012)[57] 中引理B.4，由假设条件1-5，可以证明

$$\|S_n - \Sigma\| = O_p(\sqrt{p/n}). \tag{3.3.5}$$

联合(3.3.4)和(3.3.5)得：存在常数$C_0 > 0$，当$n \to \infty$时，$P(\theta^\top S_n\theta \geq \frac{1}{2}C_0) \to 1$。因此，$\alpha = O_p(\sqrt{p/n})$，即$\|\lambda\| = \alpha = O_p(\sqrt{p/n})$。

综上所述，引理3.3.1得证。 □

定理3.1.1的证明：令：$\gamma \in \mathbb{R}^p$，$\|\gamma\| = 1$。我们首先证明：对于任意给定的$\varepsilon > 0$，存在正的常数c，使得下面等式成立

$$\lim_{n\to\infty} Pr\{|\min_{\|\gamma\|=1} l(\beta_0 + a_n\gamma) - l(\beta_0)| < \varepsilon\} = 1. \tag{3.3.6}$$

将$l(\beta)$在β_0处Taylor展开，有

$$l(\beta_0+a_n\gamma)=l(\beta_0)-2a_n\sum_{i=1}^{n}\frac{\lambda^\top V_i}{1+\lambda^\top\eta_i(\beta_0)}\gamma-a_n^2\sum_{i=1}^{n}\gamma^\top\frac{\lambda^\top V_iV_i^\top\lambda}{(1+\lambda^\top\eta_i(\beta_0))^2}\gamma+r_n,$$

其中：$r_n=-\frac{2}{3}a_n^3\sum_{i=1}^{n}\gamma^\top\frac{\lambda^\top V_iV_i^\top\lambda}{(1+\lambda^\top\eta_i(\tilde\beta))^3}\gamma\lambda^\top V_i\gamma$，$\tilde\beta\in D_n$。由假设3和引理3.3.1，我们有$\|V_i\|=O_p(\sqrt{p})$ 和

$$\|\lambda^\top\eta_i(\beta_0)\|\le\|\lambda^\top\|\|\eta_i(\beta_0)\|=o_p(1).$$

因此，

$$\begin{aligned}\|2a_n\sum_{i=1}^{n}\frac{\lambda^\top V_i}{1+\lambda^\top\eta_i(\beta_0)}\gamma\| &\le 2na_n\|\lambda^\top\|\tfrac{1}{n}\sum_{i=1}^{n}\|V_i\||\tfrac{1}{1+\lambda^\top\eta_i(\beta_0)}|\\ &=2na_nO_p\{\sqrt{p^2/n}\}=O_p\{\sqrt{p^3}\},\end{aligned}$$

$$\begin{aligned}\|a_n^2\sum_{i=1}^{n}\gamma^\top\frac{\lambda^\top V_iV_i^\top\lambda}{(1+\lambda^\top\eta_i(\beta_0))^2}\gamma\| &\le a_n^2\|\lambda^\top\|^2\sum_{i=1}^{n}\|V_i\|^2|\frac{1}{(1+\lambda^\top\eta_i(\beta_0))^2}|\\ &= a_n^2nO_p\{p^2/n\}=O_p\{p^3/n\},\end{aligned}$$

以及

$$\|r_n\|\le a_n^3\|\lambda^\top\|^3\sum_{i=1}^{n}\|V_i\|^3|\frac{1}{(1+\lambda^\top\eta_i(\beta_0))^3}|=O_p\{\sqrt{p^9/n^4}\}.$$

由假设4，我们有

$$\frac{1}{n}(l(\beta_0+a_n\gamma)-l(\beta_0))=\frac{1}{n}(O_p\{\sqrt{p^3}\}+O_p\{p^3/n\}+O_p\{\sqrt{p^9/n^4}\})=o_p(1).$$

从而

$$\lim_{n\to\infty}Pr\{|\min_{\|\gamma\|=1}l(\beta_0+a_n\gamma)-l(\beta_0)|<\varepsilon\}=1.$$

综上所述，(3.3.6)式成立。定理3.1.1证明完毕。 □

引理 3.3.2 在假设1-5成立条件下，当$n\to\infty$时，以概率1，由(3.1.7)式给出的$l(\beta)$在D_n内的最小值存在。

证明：本引理证明过程类似于Leng 和 Tang (2012)[84] 中引理4的证明，因此省略。 □

引理 3.3.3 在假设1-5成立下，当$n\to\infty$时，有

$$\frac{1}{n}\sum_{i=1}^{n}\eta_i(\beta)\eta_i(\beta)^\top\xrightarrow{p}\Sigma,$$

其中：$\xrightarrow{p}$表示依概率收敛。

证明：本引理证明过程类似于Li et al. (2012)[57] 中引理B.4的证明，因此省略。 □

定理3.1.2的证明： 由引理3.3.1和假设3，有

$$\max_{i=1,2,\cdots,n} \|\lambda^\top \eta_i(\beta_0)\| \leq \max_{i=1,2,\cdots,n} \|\lambda^\top\| \|\eta_i(\beta_0)\| = o_p(1). \tag{3.3.7}$$

将(3.1.8)式左边展开得

$$0 = \frac{1}{n}\sum_{i=1}^{n} \eta_i(\beta) - S_n\lambda + R_1, \tag{3.3.8}$$

其中：$S_n = \frac{1}{n}\sum_{i=1}^{n} \eta_i(\beta)\eta_i(\beta)^\top$，$R_1 = \frac{1}{n}\sum_{i=1}^{n} \eta_i(\beta)\frac{(\lambda^\top\eta_i(\beta))^2}{(1+\xi_i)^3}$，$|\xi_i| \leq |\lambda^\top\eta_i(\beta))|$。因为$\max\limits_{i=1,\cdots,n} \|\lambda^\top\eta_i(\beta_0)\| = o_p(1)$，所以$\max\limits_{i=1,\cdots,n} \|\xi_i\| = o_p(1)$。利用引理3.3.1和假设4可得

$$\|\frac{1}{n}\sum_{i=1}^{n} \eta_i(\beta)^\top\eta_i(\beta)\| = \|\frac{1}{n}\sum_{i=1}^{n} O_p(p)\| = O_p(p), \tag{3.3.9}$$

和

$$R_1 \leq \|\frac{1}{n}\sum_{i=1}^{n} \eta_i(\beta)^\top\eta_i(\beta)\|\|\lambda^\top\|^2\{1 + o_p(1)\} = O_p\{p^2/n\} = o_p(1).$$

令：$\bar{\eta}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \eta_i(\beta)$, 又(3.3.8)有

$$\lambda = S_n^{-1}\bar{\eta}_n(\beta) + S_n^{-1}R_1. \tag{3.3.10}$$

因为：$\log(1 + X_i) = X_i - \frac{X_i^2}{2} + \frac{X_i^3}{3(1+\zeta_i)^3}$，$|\zeta_i| \leq |X_i|$，所以

$$\begin{aligned}
l(\beta) &= 2\sum_{i=1}^{n} \log(1 + \lambda^\top\eta_i(\beta)) \\
&= 2\sum_{i=1}^{n} \lambda^\top\eta_i(\beta) - \sum_{i=1}^{n}(\lambda^\top\eta_i(\beta))^2 + \frac{2}{3}\sum_{i=1}^{n}(\lambda^\top\eta_i(\beta))^3(1 + \zeta_i)^{-4} \\
&= n\bar{\eta}_n(\beta)^\top S_n^{-1}\bar{\eta}_n(\beta) - nR_1^\top S_n^{-1}R_1 + \frac{2}{3}\sum_{i=1}^{n}(\lambda^\top\eta_i(\beta))^3(1 + \zeta_i)^{-4}.
\end{aligned}$$

由(3.3.7)-(3.3.9)，假设3-4和引理3.3.3，有

$$|nR_1^\top S_n^{-1}R_1| \leq n\|R_1\|^2\|S_n^{-1}\| = O_p\{p^{\frac{9}{2}}/n\} = o_p(\sqrt{p}),$$

和

$$\begin{aligned} |\frac{2}{3}\sum_{i=1}^{n}\frac{(\lambda^\top \eta_i(\beta))^3}{(1+\zeta_i)^3}| &\le n\|\lambda\|^3\|\frac{1}{n}\sum_{i=1}^{n}\eta_i(\beta)^\top\eta_i(\beta)\eta_i(\beta)\|\{1+o_p(1)\} \\ &= O_p(\sqrt{p^6/n}) = o_p(\sqrt{p}). \end{aligned}$$

因此，类似于Chen et al. (2009)[56] 中引理5和6的证明可得

$$(2p)^{-\frac{1}{2}}(l(\beta_0)-p) \xrightarrow{d} N(0,1).$$

综述所述，定理3.1.2证明完毕。 □

定理3.1.3的证明： 我们首先证明：$\max_{1\le i\le n}\|\tilde{\eta}_i(\beta)\| = o_p(\sqrt{n})$。

$$\begin{aligned} \tilde{\eta}_i(\beta) &= b_i^{(1)} - V_i^{(1)}\beta^{(1)} - V_i^{(2)}\hat{\beta}^{(2)} \\ &= b_i^{(1)} - V_{1i}\beta - V_i^{(2)}(\hat{\beta}^{(2)} - \beta^{(2)}). \end{aligned}$$

$$\max_{1\le i\le n}\|\tilde{\eta}_i(\beta)\| \le \max_{1\le i\le n}\|b_i^{(1)} - V_{1i}\beta\| + \max_{1\le i\le n}\|V_i^{(2)}(\hat{\beta}^{(2)}-\beta^{(2)})\|.$$

由定理3.1.1和引理5.3.1，有

$$\max_{1\le i\le n}\|\tilde{\eta}_i(\beta)\| = o_p(\sqrt{n}). \tag{3.3.11}$$

类似于Li和wang(2003)[44]中定理3.1证明过程，我们可以证明

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\eta}_i(\beta) \xrightarrow{d} N(0,\Lambda(\beta)), \tag{3.3.12}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\eta}_i(\beta)\tilde{\eta}_i(\beta)^\top \xrightarrow{p} \Lambda(\beta), \tag{3.3.13}$$

其中：$\Lambda(\beta) = E(\tilde{\eta}(\beta)\tilde{\eta}(\beta)^\top)$是$k\times k$阶正定矩阵。由(3.1.10)式右边Taylor展开得

$$\tilde{l}(\beta^{(1)}) = 2\sum_{i=1}^{n}\lambda^{*\top}\tilde{\eta}_i(\beta) - \sum_{i=1}^{n}\{\lambda^{*\top}\tilde{\eta}_i(\beta)\}^2 + o_p(1). \tag{3.3.14}$$

类似于Owen (1990)[3] 中定理1的证明，容易得出

$$\sum_{i=1}^{n}\{\lambda^{*\top}\tilde{\eta}_i(\beta)\}^2 = \sum_{i=1}^{n}\lambda^{*\top}\tilde{\eta}_i(\beta) + o_p(1), \tag{3.3.15}$$

$$\lambda^{*\top} = \{\sum_{i=1}^{n}\tilde{\eta}_i(\beta)\tilde{\eta}_i(\beta)^\top\}^{-1}\sum_{i=1}^{n}\tilde{\eta}_i(\beta) + o_p(1/\sqrt{n}). \tag{3.3.16}$$

联合(3.3.14)-(3.3.16)，有

$$\tilde{l}(\beta^{(1)}) = \{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\eta}_i(\beta)\}^\top\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\eta}_i(\beta)\tilde{\eta}_i(\beta)^\top\}^{-1}\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\eta}_i(\beta)\} + o_p(1).$$

因此，由(3.3.12)-(3.3.13)以及正态分布的性质可知，$\tilde{l}(\beta^{(1)})$的渐近分布是自由度为k的卡方分布。

综述所述，定理3.1.3证明完毕。 □

引理 3.3.4 在假设1-7成立条件下，当$n \to \infty$时，有：

(1): $\|\lambda\|_\beta = O_p(a_n)$；

(2): 以概率1，由(3.2.1)式给出的$l_p(\beta)$在D_n内的最小值存在。

证明： 本引理证明过程类似于Tang和Leng(2010)[83] 中引理4的证明，因此省略。 □

引理 3.3.5 在假设1-7成立条件下，有：$\|\Sigma - S_{n0}\| = O_p\{(p/n)^{\frac{1}{2}}\}$；$\|\mathbf{D} - V\| = O_p\{(p/n)^{\frac{1}{2}}\}$。

证明：本引理证明过程类似于Lin和Lv(2013)[79] 中引理A.4的证明，因此省略。 □

定理3.2.1的证明： 由引理3.3.4知，$l_p(\beta)$在D_n内的最小值存在。任意给定$\beta \in D_n$，有

$$\frac{1}{n}\frac{\partial l_p(\beta)}{\partial \beta_j} = \frac{1}{n}\sum_{i=1}^{n}\frac{-\lambda_\beta^\top V_i^{(j)}}{1+\lambda_\beta^\top \eta_i(\beta)} + P_\nu^{'}(|\beta_j|)\mathrm{sign}(\beta_j) = M_1 + M_2,$$

其中：$V_i^{(j)}$是矩阵V_i的第j列。由引理3.3.1，引理3.3.4和假设3-4可得

$$\max_{i=1,2,\cdots,n}\|\lambda_\beta^\top \eta_i(\beta_0)\| \le \max_{i=1,2,\cdots,n}\|\lambda_\beta^\top\|\|\eta_i(\beta_0)\| = o_p(1),$$

和

$$|M_1| \le \|\lambda_\beta\|\|V_i^j\||\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\lambda_\beta^\top \eta_i(\beta)}| = O_p(\frac{p}{\sqrt{n}})(1+o_p(1)) = o_p(1).$$

从而，当$n \to \infty$时，有

$$Pr(|M_1| > C) \to 0.$$

因为$|\beta_j|_{\{j \notin \mathcal{A}\}} \le a_n$，由假设6，对于足够大的$n$ ，我们有

$$P_\nu^{'}(|\beta_j|)_{\{j \notin \mathcal{A}\}} = \nu,$$

$$M_2 = P_\nu'(|\beta_j|)\mathrm{sign}(\beta_j)_{\{j\notin\mathcal{A}\}} = \nu\mathrm{sign}(\beta_j)_{\{j\notin\mathcal{A}\}}.$$

因此，当$n\to\infty$时，以概率1，对于$j\notin\mathcal{A}$，以概率1，β_j的符号由$\frac{\partial l_p(\beta)}{\partial\beta_j}$决定，这意味着，以概率1，$\hat\beta^{(2)}=0$，从而定理3.2.1(1)得证。

接下来我们开始证明定理3.2.1(2)。考虑在约束条件$H_2\beta=0$下(3.2.1)式的最小值问题。由Lagrange乘数法，求约束条件$H_2\beta=0$下(3.2.1)式的最小值等价于求下面这个新的无约束条件目标函数

$$\tilde{l}(\beta,\lambda_\beta,\mu)=\frac{1}{n}\sum_{i=1}^{n}\log(1+\lambda_\beta^\top\eta_i(\beta))+\sum_{j=1}^{p}p_\nu(|\beta_j|)+\mu^\top H_2\beta,$$

的极小值，其中：$\mu\in\mathbb{R}^{(p-d)}$ 也是Lagrange乘数。令：

$$\tilde{Q}_{1n}(\beta,\lambda_\beta,\mu)=\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i(\beta)}{(1+\lambda_\beta^\top\eta_i(\beta))},\quad \tilde{Q}_{3n}(\beta,\lambda_\beta,\mu)=H_2\beta,$$

$$\tilde{Q}_{2n}(\beta,\lambda_\beta,\mu)=\frac{1}{n}\sum_{i=1}^{n}\frac{-V_i^\top\lambda_\beta}{(1+\lambda_\beta^\top\eta_i(\beta))}+b(\beta)+H_2^\top\mu,$$

其中：

$$b(\beta)=\{P_\nu'(\beta_1)\mathrm{sign}(\beta_1),P_\nu'(\beta_2)\mathrm{sign}(\beta_2),\cdots,P_\nu'(\beta_d)\mathrm{sign}(\beta_d),0^\top\}^\top.$$

使(3.2.1)式达到最小的参数$(\hat\beta,\hat\lambda_\beta,\hat\mu)$满足

$$\tilde{Q}_{in}(\hat\beta,\hat\lambda_\beta,\hat\mu)=0,\quad i=1,2,3.$$

由于 $\|\lambda_\beta\|=O_p(a_n)$ 和 $\tilde{Q}_{2n}(\hat\beta,\hat\lambda_\beta,\hat\mu)=0$，可以得到 $\|\mu\|=O_p(a_n)$。将 $\tilde{Q}_{in}(\beta,\lambda_\beta,\mu)$ 在 $(\beta_0,0,0)$ 处Taylor展开得

$$\begin{pmatrix}-\tilde{Q}_{1n}(\beta_0,0,0)\\0\\0\end{pmatrix}=\begin{pmatrix}-\Sigma & -\mathbf{D} & 0\\ -\mathbf{D}^\top & 0 & H_2^\top\\ 0 & H_2 & 0\end{pmatrix}\begin{pmatrix}\hat\lambda_\beta-0\\ \hat\beta-\beta_0\\ \hat\mu-0\end{pmatrix}+R_n, \tag{3.3.17}$$

其中：$R_n=\sum_{k=1}^5 R_n^k$, $R_n^{(1)}=(R_{1n}^{(1)\top},R_{2n}^{(1)\top},0)^\top$，$R_{jn}^{(1)}\in\mathbb{R}^p$，$j=1,2$，$R_{jn}^{(1)}$的第$j$个成分为

$$R_{jn,k}^{(1)}=\frac{1}{2}(\hat\theta-\theta_0)^\top\frac{\partial^2\tilde{Q}_{jn,k}(\theta^*)}{\partial\theta\partial\theta^\top}(\hat\theta-\theta_0),$$

这里$\theta=(\beta^\top,\lambda^\top)$，$\theta^*=(\beta^{*\top},\lambda^{*\top})$ 满足$\|\beta^*-\beta_0\|\le\|\hat\beta-\beta_0\|$ 和$\|\lambda^*\|\le\|\hat\lambda\|$。令$S_{n0}=\frac{1}{n}\sum_{i=1}^n\eta_i(\beta_0)\eta_i(\beta_0)^\top$，有

$$R_n^{(2)}=\{0,b^\top(\beta_0),0\}^\top,$$

$$R_n^{(3)} = \{0, \{b^{'}(\beta_0)(\hat{\beta} - \beta_0)\}^\top, 0\}^\top,$$

$$R_n^{(4)} = \{\{(\Sigma - S_{n0})\hat{\lambda}\}^\top + \{(\mathbf{D} - V)(\hat{\beta} - \beta_0)\}^\top, 0, 0\}^\top,$$

和

$$R_n^{(5)} = \{0, \{\mathbf{D}^\top - V^\top)\hat{\lambda}\}^\top, 0\}^\top.$$

由引理3.3.4，假设4以及Cauchy-Schwarz不等式得$\|\hat{\theta} - \theta_0\| = O_p(a_n)$，和

$$\|R_{jn}^{(1)}\|^2 \le \|\hat{\theta} - \theta_0\|^4 |\sum_{k,l=1}^{p} \frac{\partial^2 \tilde{Q}_{jn}(\theta^*)}{\partial\theta_k \partial\theta_l}| = O_p\{p^2 a_n^4\}, \qquad j = 1,\ 2.$$

从而，有

$$\|R_n^{(1)}\| = O_p\{(p^4/n^2)^{\frac{1}{2}}\} = o_p(n^{-\frac{1}{2}}).$$

由假设6-7，易得

$$\|R_n^{(2)}\| = o_p(n^{-\frac{1}{2}}), \qquad \|R_n^{(3)}\| = o_p(n^{-\frac{1}{2}}).$$

由引理3.3.5可知

$$\|R_n^{(4)}\| = O_p\{(p/n)^{\frac{1}{2}}\} O_p(a_n) = o_p(n^{-\frac{1}{2}}), \quad \|R_n^{(5)}\| = o_p(n^{-\frac{1}{2}}).$$

令

$$\vartheta = (\beta^\top \ \ \mu^\top)^\top, \ \ M_{11} = -\Sigma, \ \ M_{12} = (-\mathbf{D} \ \ 0), \ \ M_{21} = M_{12}^\top,$$

和

$$M_{22} = \begin{pmatrix} 0 & H_2^\top \\ H_2 & 0 \end{pmatrix}, \ \ M = \begin{pmatrix} M_{11} & M_{12}^\top \\ M_{21} & M_{22} \end{pmatrix}.$$

由(3.3.17)的逆运算可得

$$(\hat{\lambda} - 0 \ \ \hat{\vartheta} - \vartheta_0)^\top = M^{-1}\{(-Q_{1n}(\beta_0, 0, 0) \ \ 0 \ \ 0)^\top + R_n\},$$

其中：$\|R_n\| \le \sum_{k=1}^5 \|R_n^k\| = o_p(n^{-\frac{1}{2}})$。由分块矩阵的求逆方法，有

$$M^{-1} = \begin{pmatrix} M_{11}^{-1} + M_{11}^{-1} M_{12} B^{-1} M_{21} M_{11}^{-1} & -M_{11}^{-1} M_{12} B^{-1} \\ -B^{-1} M_{21} M_{11}^{-1} & B^{-1} \end{pmatrix},$$

其中

$$B = M_{22} - M_{21} M_{11}^{-1} M_{12} = \begin{pmatrix} A & H_2^\top \\ H_2 & 0 \end{pmatrix},$$

以及$A = \mathbf{D}^\top \Sigma^{-1} \mathbf{D}$。因此

$$\hat{\vartheta} - \vartheta_0 = B^{-1} M_{21} M_{11}^{-1} \widetilde{Q}_{1n}(\beta_0, 0, 0) + o_p(n^{-\frac{1}{2}}),$$

和

$$B^{-1} = \begin{pmatrix} A^{-1} - A^{-1} H_2^\top (H_2 A^{-1} H_2^\top)^{-1} H_2 A^{-1} & A^{-1} H_2^\top (H_2 A^{-1} H_2^\top)^{-1} \\ (H_2 A^{-1} H_2^\top)^{-1} H_2 A^{-1} & -(H_2 A^{-1} H_2^\top)^{-1} \end{pmatrix}.$$

从而

$$\hat{\beta} - \beta_0 = \{A^{-1} - A^{-1} H_2^\top (H_2 A^{-1} H_2^\top)^{-1} H_2 A^{-1}\}\{\overline{\eta}_n(\beta_0) + R_{1n}\}, \tag{3.3.18}$$

其中R_{1n}是R_n中相应的分量，且$\|R_{1n}\| = o_p(n^{-\frac{1}{2}})$。对于$\beta$的非零分量$\beta_1$，有

$$\hat{\beta}^{(1)} - \beta_0^{(1)\top} = \{H_1 A^{-1} - H_1 A^{-1} H_2^\top (H_2 A^{-1} H_2^\top)^{-1} H_2 A^{-1}\}\{\overline{\eta}_n(\beta_0) + R_{1n}\}.$$

$n^{\frac{1}{2}}\hat{\beta}_1$的渐近协方差矩阵为

$$I_{\mathcal{A}} = H_1 A^{-1} \Sigma A^{-1} H_1^\top - \Gamma_{12} \Gamma_{22}^{-1} H_2 A^{-1} \Sigma A^{-1} H_2^\top \Gamma_{22}^{-1} \Gamma_{21},$$

其中：$\Gamma_{ij} = H_i A^{-1} H_j^\top$，$i, j = 1, 2$。令$Y_{ni} = \frac{1}{\sqrt{n}} T_{ni}$，其中

$$T_{ni} = A_n I_{\mathcal{A}}^{-\frac{1}{2}} (H_1 A^{-1} - H_1 A^{-1} H_2^\top (H_2 A^{-1} H_2^\top)^{-1} H_2 A^{-1}) \eta_i(\beta_0).$$

容易证明

$$Pr(\|Y_{ni}\| > \epsilon) \le n^{-1} \epsilon^{-2} E\|T_{ni}\|^2 = O(\frac{1}{n})$$

和

$$E\|Y_{ni}\|^4 = n^{-2} E(T_{ni}^\top T_{ni})^2 = O(p^2/n^2).$$

因此

$$\sum_{i=1}^{n} E\|Y_{ni}\|^2 I(\|Y_{ni}\| > \epsilon) \le n\{E\|Y_{n1}\|^4\}^{\frac{1}{2}} \{Pr(\|Y_{ni}\| > \epsilon)\}^{\frac{1}{2}} \to 0.$$

当$A_n A_n^\top \to G$时，由Lindeberg-Feller中心极限定理可得

$$\sqrt{n} A_n I_{\mathcal{A}}^{-\frac{1}{2}} \{H_1 A^{-1} - H_1 A^{-1} H_2^\top (H_2 A^{-1} H_2^\top)^{-1} H_2 A^{-1}\} \overline{\eta}_n(\beta_0) \xrightarrow{d} N(0, G).$$

最后，由于$\|\sqrt{n} A_n I_{\mathcal{A}}^{-\frac{1}{2}} R_{1n}\| = o_p(1)$，从而，定理3.2.1证明完毕。 □

定理3.2.2的证明： 令$y_i = \hat{\lambda}^\top \eta_i(\hat{\beta})$。由引理3.3.4知，$|y_i| = o_p(1)$。利用Taylor公式，得到$l(\hat{\beta})$的渐近展开式

$$l(\hat{\beta}) = \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{y_i^2}{2} + \sum_{i=1}^n \frac{y_i^3}{3(1+\xi_i)^4 + o_p(1)},$$

其中：$\hat{\beta}$是使(3.2.1)式达到最小的β，$|\xi_i| \le |y_i|$。由(3.3.10)式，有

$$\lambda = S_n^{-1}\overline{\eta}_n(\beta) + S_n^{-1}R_1,$$

其中

$$R_1 = \frac{1}{n}\sum_{i=1}^n \eta_i(\beta)\frac{(\lambda^\top \eta_i(\beta))^2}{(1+\xi_i)^2}, \quad |\xi_i| \le |y_i|.$$

类似于Hou et al. (2014)[82]中定理2.2和Tang和Leng(2010)[83]中定理4的证明，将展开式中的$\hat{\beta}$ 和$\hat{\lambda}$代入y_i得

$$2l(\hat{\beta}) = n\overline{\eta}_n(\beta_0)^\top A^{-1}H_2^\top(H_2A^{-1}H_2^\top)^{-1}H_2A^{-1}\overline{\eta}_n(\beta_0) + o_p(1). \tag{3.3.19}$$

在原假设H_0成立下，因为$L_nL_n^\top = I_q$，所以存在$\widetilde{H}_2$使得$\widetilde{H}_2\beta = 0$，$\widetilde{H}_2\widetilde{H}_2^\top = I_{p-d+q}$。类似于定理3.2.1的推导过程，$\beta$的估计量$\tilde{\beta}$能够通过最小化

$$\tilde{l}(\beta,\lambda,\mu) = \sum_{i=1}^n \log(1+\lambda^\top \eta_i(\beta)) + \sum_{j=1}^p p_\nu(|\beta_j|) + \mu^\top \widetilde{H}_2\beta. \tag{3.3.20}$$

获得。设(3.3.20)式在$(\tilde{\beta},\tilde{\lambda},\tilde{\mu})$处达到最小。由定理3.2.1(1)，以概率1，有$\tilde{\beta}_2 = 0$。再由假设6-7，以概率1，我们有$n\{\sum_{j=1}^p p_\nu(|\hat{\beta}_j|) - \sum_{j=1}^p p_\nu(|\tilde{\beta}_j|)\} = 0$。通过在(3.3.18)式中用$\widetilde{H}_2$代替$H_2$可得

$$2l(\tilde{\beta}) = n\overline{\eta}_n(\beta_0)^\top A^{-1}\widetilde{H}_2^\top(\widetilde{H}_2A^{-1}\widetilde{H}_2^\top)^{-1}\widetilde{H}_2A^{-1}\overline{\eta}_n(\beta_0) + o_p(1). \tag{3.3.21}$$

联合(3.3.19)式和(3.3.21)式，有

$$\widetilde{l}_p(L_n) = n\overline{\eta}_n(\beta_0)^\top A^{-\frac{1}{2}}\{P_1 - P_2\}A^{-\frac{1}{2}}\overline{\eta}_n(\beta_0) + o_p(1),$$

其中

$$P_1 = A^{-\frac{1}{2}}\widetilde{H}_2^\top(\widetilde{H}_2A^{-1}\widetilde{H}_2^\top)^{-1}\widetilde{H}_2A^{-\frac{1}{2}}$$

和

$$P_2 = A^{-\frac{1}{2}}H_2^\top(H_2A^{-1}H_2^\top)^{-1}H_2A^{-\frac{1}{2}}.$$

因为P_1-P_2是秩为q的对称幂等矩阵，所以P_1-P_2可以表示为$B_n^\top B_n$，B_n是$q\times p$阶矩阵满足$B_nB_n^\top=I_q$。令$K_{B_n}=B_nA^{-\frac{1}{2}}\Sigma A^{-\frac{1}{2}}B_n^\top$，容易证明

$$K_{B_n}^{-\frac{1}{2}}\widetilde{l}_p(L_n)K_{B_n}^{-\frac{1}{2}}=nK_{B_n}^{-\frac{1}{2}}\overline{\eta}_n(\beta_0)^\top A^{-\frac{1}{2}}\{P_1-P_2\}A^{-\frac{1}{2}}\overline{\eta}_n(\beta_0)K_{B_n}^{-\frac{1}{2}}+o_p(1).$$

类似于定理3.2.1的证明，我们有

$$\sqrt{n}K_{B_n}^{-\frac{1}{2}}B_nA^{-\frac{1}{2}}\overline{\eta}_n(\beta_0)\xrightarrow{d}N(0,I_q).$$

从而

$$nK_{B_n}^{-\frac{1}{2}}\overline{\eta}_n(\beta_0)^\top A^{-\frac{1}{2}}\{P_1-P_2\}A^{-\frac{1}{2}}\overline{\eta}_n(\beta_0)K_{B_n}^{-\frac{1}{2}}\xrightarrow{d}\chi_q^2.$$

综上所述，定理3.2.2证明完毕。 □

§3.4 数值模拟

本节通过数值模拟进一步验证本章所给出的理论结果。

本节模拟中，我们利用Owen(2001)[4]中给出的迭代最小二乘算法求经验似然比的最小值，利用Fan和Li(2001)[69]中给出的局部二次逼近算法(LQA)求其惩罚经验似然比的最小值。在计算经验似然和惩罚经验似然估计过程中，选择用Owen(2001)[4]中解决非线性优化问题的嵌套算法(nested algorithm)分别最小化 (3.1.7) 和 (3.2.1)。下面给出嵌套算法求经验似然和惩罚经验似然估计的步骤。

步0. 给出β的初始值β_0；

步1. 对于给定的初始值β_0，使用牛顿算法求使得(3.2.1)式达到最小的$\lambda^{(k)}$；

步2. 对于步1得到的$\lambda^{(k)}$，使用局部二次逼近算法求使得(3.2.1)式达到最小的$\beta^{(k)}$；

步3. 令$\beta_0=\beta^{(k)}$，转步1，重复迭代直到$\beta^{(k)}$收敛。

在求惩罚经验似然估计过程中，假设β_0是一个较好地初始值，$\beta_j^{(k)}$是 β_j 的第k步的估计值。如果$\beta_j^{(k)}$非常接近零，我们令$\widehat{\beta}_j^{(k)}=0$。当$\beta_j^{(k)}\neq 0$时，$p_\nu(|\beta_j|)$可由

$$p_\nu(|\beta_j^{(k)}|)+\frac{1}{2}\{p'_\nu(|\beta_j^{(k)}|)/|\beta_j^{(k)}|\}\{\beta_j^2-(\beta_j^{(k)})^2\}$$

二次近似。上述步骤重复迭代，直到收敛。

关于惩罚函数的规范化参数ν的选择问题，我们使用Variyath et al.(2010) [95] 中给出的基于经验似然的贝叶斯信息准则

$$BIC(\nu) = -2l_p(\beta_\nu) + \mathrm{df}_\nu.\log(n)$$

选择ν，其中：df_ν 非零系数对于的变量的个数。

考虑模型

$$h(t|\mathbf{Z}) = 1 + \beta_o^\top \mathbf{Z}, \tag{3.4.1}$$

其中$\beta_o^\top \mathbf{Z} > -1$，$\mathbf{Z}$服从均值零协方差矩阵为$(0.5^{|i-j|})_{i,j=1}^p$的多元正态分布。

第一个模拟考虑利用经验似然方法构造的参数β的置信域的覆盖概率。令$\beta_o = (u, \cdots, u)$或$\beta_o = (u, \cdots, u, u_0)$，$u = (1,0,0)$，在参数$\beta_o$中重复$q$次，$\beta_o$的维数 p 不同其相应的 q 也不同。考虑：$p = 20(q = 6, u_0 = (1,0))$, $30\ (q = 10)$, $40(q = 13, u_0 = 1)$，$50(q = 16, u_0 = (1,0))$，$n = 100, \cdots, 400$。删失时间C由均匀分布$U(0,c)$产生，通过调整c可以确定删失比率。我们选择的删失比率大约为25%，其相应的$c = 3$。置信水平分别取为0.90和0.95。所有模拟重复1000次，其结果总结在表3.4.1。

从表3.4.1可以看出，对于不同的维数p，经验似然置信域的覆盖概率随样本容量 n 的增加而增加，当样本容量达到一定程度时，覆盖概率接近给定的置信水平。

在第二个模拟中，我们将比较Lin和Ying(1994)[94]的估计方法、Lin和Lv (2013)[79]的估计方法(L-SCAD、L-MCP 和L-SICA方法)、经验似然和惩罚经验似然估计方法的效率。在模型(3.4.1)中，令 $\beta = (3, 1.5, 0, 0, 2, 0, \cdots, 0)^\top$，$p = 20$，所有模拟重复1000次。删失比率大约15%时，表3.4.2给出了上述几种估计方法对非零参数分量β_1、β_2和β_5的估计结果(均值及其均方误差根)。结果显示，Lin 和 Lv 的估计方法、经验似然和惩罚经验似然估计方法的均方误差根都小于 Lin 和 Ying 的估计方法。特别地，惩罚经验似然估计方法的均方误差根是上述所有估计方法中最小的，而且估计均值更接近真实参数。

第三个模拟将对Lin和Lv的变量选择方法和惩罚经验似然变量选择方法进行比较。在模型(3.4.1)中，令$\beta = (3, 1.5, 0, 0, 2, 0, \cdots, 0)^\top$，$p = 20$。针对惩罚经验似然方法，我们考虑了三种不同惩罚函数：Lasso、Bridge和SCAD 惩罚函数。所有模拟重复1000次，其结果总结在表3.4.3。表3.4.3中列标题为“C”的列表示真实参数分量为零时估计结果也为零的分量个数

表 3.4.1: 置信水平分别为0.90和0.95时经验似然方法构造的置信域的覆盖概率(删失比率大约为25%)

Sample dimension(p)	Sample size (n)	1-α=0.90	1-α=0.95
20	100	0.873	0.922
	200	0.886	0.934
	300	0.902	0.951
30	100	0.867	0.913
	200	0.882	0.938
	300	0.895	0.945
40	200	0.846	0.876
	300	0.861	0.903
	400	0.879	0.924
50	200	0.834	0.865
	300	0.853	0.897
	400	0.872	0.915

表 3.4.2: Lin和Ying的估计、Lin和Lv的估计(L-SCAD，L-MCP和L-SICA)、经验似然估计和SCAD－惩罚经验似然估计结果的比较(删失比率大约15%，SCAD惩罚函数中$a = 3.7$)。表中显示的值为相应估计的均值(均方误差根)

n	Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_5$
100	LY	3.964 (0.781)	2.142 (0.571)	2.673 (0.685)
	EL	3.625 (0.536)	1.169 (0.346)	2.238 (0.279)
	PEL	3.412 (0.327)	1.285 (0.314)	2.214 (0.236)
	L-MCP	2.493 (0.518)	1.791 (0.427)	2.352 (0.361)
	L-SCAD	3.561 (0.469)	1.147 (0.359)	1.736 (0.354)
	L-SICA	2.478 (0.512)	1.203 (0.328)	2.341 (0.372)
200	LY	3.758 (0.643)	1.926 (0.464)	2.551 (0.492)
	EL	3.525 (0.485)	1.841 (0.297)	2.147 (0.241)
	PEL	3.297 (0.286)	1.732 (0.278)	2.135 (0.194)
	L-MCP	2.612 (0.374)	1.214 (0.362)	1.698 (0.327)
	L-SCAD	3.453 (0.351)	1.253 (0.294)	1.821 (0.286)
	L-SICA	2.694 (0.325)	1.817 (0.257)	1.793 (0.258)
300	LY	3.642 (0.591)	1.854 (0.379)	2.476 (0.438)
	EL	3.385 (0.427)	1.697 (0.241)	2.109 (0.206)
	PEL	3.231 (0.223)	1.645 (0.185)	2.084 (0.125)
	L-MCP	2.746 (0.336)	1.743 (0.307)	2.275 (0.282)
	L-SCAD	3.314 (0.329)	1.718 (0.245)	2.162 (0.214)
	L-SICA	2.761 (0.248)	1.262 (0.213)	1.845 (0.183)

表 **3.4.3**: 变量选择结果(删失比率大约15%)

n	Method	C(17)	I(0)	Method	C(17)	I(0)
100	SCAD	16.12	0.16	L-MCP	13.87	0.92
	Bridge	13.65	0.71	L-SCAD	14.52	0.67
	Lasso	8.41	1.08	L-SICA	15.39	0.45
200	SCAD	16.35	0.05	L-MCP	14.46	0.81
	Bridge	13.42	0.56	L-SCAD	15.14	0.53
	Lasso	8.32	0.94	L-SICA	15.75	0.37
300	SCAD	16.78	0.00	L-MCP	15.63	0.24
	Bridge	14.36	0.42	L-SCAD	16.21	0.19
	Lasso	8.54	0.87	L-SICA	16,56	0.08

表 **3.4.4**: β_1真实值为3，β_1的给定值没有落在EL和PEL置信域内的比率(删失比率大约15%)

p	n	Method	2.8	2.9	3	3.1	3.2
10	100	EL	31.2	12.9	7.6	13.4	30.6
		PEL	42.3	26.5	6.4	25.6	44.2
20	200	EL	58.6	21.7	8.2	22.5	57.8
		PEL	67.4	35.8	6.7	34.9	68.1
30	300	EL	71.2	32.9	7.9	31.4	70.6
		PEL	82.3	46.5	5.9	47.6	84.2
40	400	EL	78.6	41.7	7.2	38.5	75.8
		PEL	87.4	53.8	5.7	51.9	86.1

表 3.5.1: 利用SCAD-PEL、Bridge-PEL和Lasso-PEL方法对Sorlie数据集的变量选择的结果

Method of PEL	number of selected genes
SCAD-PEL	18
Bridge-PEL	21
Lasso-PEL	63

平均值，列标题为"I"的列表示真实参数分量为零但估计结果不为零的分量个数平均数。模拟结果显示，带有SCAD惩罚函数的惩罚经验似然方法选择变量结果比其他方法更好，估计结果为零的参数分量个数平均值更接近$p-3$，并且能准确地选择出三个重要变量(系数不为零的变量)。这个模拟结果进一步验证了本章提出的带有SCAD惩罚函数的惩罚经验似然方法对高维稀疏可加危险率模型进行变量选择方法的优点。

在最后一个模拟中验证了$\beta_1 \notin V_\alpha(L_n\beta)$的频率，$V_\alpha(L_n\beta)$是根据经验似然或惩罚经验似然方法构造的参数分量的置信域或置信区间，β_1是参数β 的第一个分量。在模型(3.4.1)中，令$\beta = (3, 1.5, 0, 0, 2, 0, \cdots, 0)^\top$，$L_n = (1, 0, \cdots, 0)^\top$。模拟重复1000次，其结果总结在表3.4.4中。从表3.4.4可以看出：(1) 当$\beta_1 = 3$为参数真实值时，$\beta_1 \notin V_\alpha(L_n\beta)$的频率接近置信水平0.05，这恰好验证了定理3.1.3和定理3.2.2的结论；(2) 根据经验似然或惩罚经验似然方法构造的假设检验的势随着$|\beta_1 - \beta_{10}|$的增加而增加；(3)在高维稀疏情形下，在构造参数置信域或对参数进行假设检验时，本章提出的惩罚经验似然方法比经验似然方法更好。

§3.5 实例分析

下面将利用本章提出的方法进行实例分析。Sorlie数据集包含115位被诊断为乳腺癌女性患者的生存时间和549个基因表达数据。 Sorlie et al(2000)，Martinussen 和 Scheike(2009)分别在[96]和[97]中分析过这些数据。

基因变量个数为549，样本容量为115，不考虑基因变量交互作用前提下，我们利用惩罚经验似然方法选择对样本个体生存时间有重要影响的基因变量建立可加危险率模型分析乳腺癌女性患者的危险率。其结果总结在表3.5.1和表3.5.2。

表 3.5.2: SCAD-PEL和Bridge-PEL选择变量的系数估计值

Gene name	SCAD	Bridge	Gene name	SCAD	Bridge
X8	-0.04067	-0.04858	X112	-0.03042	-0.03819
X9	-0.06397	-0.07354	X131	0.04558	0.04169
X22	-0.04187	-0.05120	X149	0.02674	
X28	-0.04299	-0.05077	X189	-0.03162	-0.03783
X41	-0.03473	-0.04206	X190	-0.05153	-0.06126
X42	-0.02678	-0.03390	X207		-0.03080
X62	-0.05861	-0.06754	X218	0.04543	0.04185
X64		-0.03319	X224	-0.03600	-0.04423
X68		-0.02695	X229	-0.04347	-0.05197
X76	-0.07078	-0.08045	X236	0.04469	0.04014
X89		-0.04158	X249	-0.05003	-0.06162

表3.5.1给出了三种惩罚函数下的惩罚经验似然方法的变量选择结果。因为Lasso惩罚函数不具有Oracle性质，Lasso-惩罚经验似然方法在变量选择时常将零系数错误地估计为非零系数，所以表3.5.1中Lasso-惩罚经验似然方法选择变量的数量明显大于比另外两种方法。SCAD-惩罚经验似然和Bridge-惩罚经验似然方法选择变量系数估计结果见表3.5.2。结果显示，三种不同惩罚函数下的惩罚经验似然方法都选择了17个相同的基因变量，这体现了惩罚经验似然方法在变量选择是具有一致性。另外，从表3.5.1和表3.5.2可以看出，针对高维稀疏可加危险率模型的变量选择问题，SCAD-惩罚经验似然比Bridge-惩罚经验似然和Lasso-惩罚经验似然方法更合理，这进一步验证了本章结论的合理性。

第四章 高维异方差部分线性单指标模型的经验似然推断

部分线性单指标模型自从被Carroll et al.[98]提出以来，受到统计学者的广泛关注。例如：Yu和Ruppert(2002)[99]给出了关于部分线性单指标模型的惩罚样条估计方法； Xia et al. (2002)[100]研究了降维后的参数估计方法；Xia 和 Härdle(2006)[101]讨论了部分线性单指标模型的半参数估计；Zhu和Xue(2006)[18]研究了部分线性单指标模型的经验似然推断；Ma和Zhu(2013)[102]给出了异方差部分线性单指标模型的双重稳健且有效的估计；Lai 和 Wang (2014)[103]讨论了响应变量随机确实情形下部分线性单指标模型的半参数有效估计，等等。

本章将考虑如下异方差部分线性单指标模型的经验似然推断：

$$Y_i = X_i^\top \beta + g(Z_i^\top \theta) + \varepsilon_i, \quad (i = 1, \cdots, n), \tag{4.1.1}$$

其中：$X_i \in \mathbb{R}^p$和$Z_i \in \mathbb{R}^r$为维数分别为p和r的随机变量，Y_i为响应变量，β和θ 为未知参数，$g(\cdot)$为未知函数，ε_i为相互独立的随机误差变量满足

$$E(\varepsilon_i|X_i, Z_i) = 0, \quad \mathrm{Var}(\varepsilon_i|X_i, Z_i) = v(X_i, Z_i) > 0,$$

$v(X, Z)$ 为关于 (X, Z) 的函数。为了保证模型的可识别性，假设θ的第一个分量等于1。令$Z_{-1} = (z_2, z_3, \cdots, z_r)^\top$。当 $n \to \infty$，$p \to \infty$ 时，首先，我们在双重稳健有效估计方法基础上建立估计方程，构造异方差部分线性单指标模型的经验似然比统计量，在一定条件下证明了该统计量渐近分布为正态分布，并构造了参数的经验似然置信域；其次，我们进一步构造了关于部分参数的经验似然比统计量，在一定条件下证明了该统计量具有渐近 χ_q^2 分布，并利用这个结论构造了部分参数的经验似然置信区间或置信域。

本章内容结构如下：第一节介绍高维情形下模型(4.1.1)的经验似然推断方法及其主要结论；第二节讨论模型(4.1.1)的两种特殊情形：部分线性模型和单指标模型的经验似然推断；第三节为定理的证明；第四和第五节分别为有限样本情形下的数值模拟结果和实例分析。

§4.1 方法与主要结果

首先，我们介绍 Ma和 Zhu(2013)[102] 提出的关于模型(4.1.1)中参数$(\beta^\top, \theta^\top)^\top$ 的估计方法。为了估计参数$(\beta^\top, \theta^\top)^\top$，Ma和Zhu给出了如下估计

方程：

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w(X_i,Z_i)\{Y_i - X_i^\top\beta - f(Z_i^\top\theta)\}[a(X_i,Z_i) - \tilde{E}\{a(X_i,Z_i)|Z_i^\top\theta)\}] = 0, \tag{4.1.2}$$

其中：$f(Z_i^\top\theta)$ 和 $\tilde{E}\{\cdot|Z_i^\top\theta)\}$都是关于$Z_i^\top\theta$的函数，$w(X_i,Z_i) = \mathrm{Var}(\varepsilon_i|X_i,Z_i)^{-1}$，$a(\cdot,\cdot) \in \mathbb{R}^{p+r-1}$为关于$X$和$Z$函数。Ma和Zhu指出：如果$w_i$和 $g(\cdot)$ 的一致估计都能求出，那么$(\beta^\top,\theta^\top)^\top$的估计$(\hat{\beta}^\top,\hat{\theta}^\top)^\top$能通过求解估计方程 (4.1.2) 得到。但是，这样通过估计方程(4.1.2)得到的估计$(\hat{\beta}^\top,\hat{\theta}^\top)^\top$ 不具有双重稳健性。为了得到模型(4.1.1)中参数的双重稳健有效估计，Ma 和 Zhu 给出了一种改进的加权估计方程：

$$\begin{cases} \frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}\tilde{\varepsilon}_i\hat{w}(X_i,Z_i)[X_i - \frac{\hat{E}\{\hat{w}(X,Z)X|Z_i^\top\theta\}}{\hat{E}\{\hat{w}(X,Z)|Z_i^\top\theta\}}] = 0, \\ \frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}\tilde{\varepsilon}_i\hat{w}(X_i,Z_i)\hat{g}'(Z_i^\top\theta)[Z_{-1,i} - \frac{\hat{E}\{\hat{w}(X,Z)Z_{-1}|Z_i^\top\theta\}}{\hat{E}\{\hat{w}(X,Z)|Z_i^\top\theta\}}] = 0, \end{cases} \tag{4.1.3}$$

其中：$\tilde{\varepsilon}_i = Y_i - X_i^\top\beta - \hat{g}(Z_i^\top\theta)$，$w(X_i,Z_i) = E(\varepsilon_i^2|X_i,Z_i)^{-1}$，$\hat{g}(Z_i^\top\theta)$，$\hat{g}'(Z_i^\top\theta)$，$\hat{E}\{\hat{w}(X,Z)|Z_i^\top\theta\}$，$\hat{E}\{\hat{w}(X,Z)X|Z_i^\top\theta\}$和$\hat{E}\{\hat{w}(X,Z)Z_{-1}|Z_i^\top\theta\}$是利用核估计方法得到的非参数估计。令$\eta_i = \eta(X_i,Z_i)$满足$\mathrm{Var}(\varepsilon_i|X_i,Z_i) = \mathrm{Var}(\varepsilon_i|\eta_i)$，$K_h(\cdot) = h^{-1}K(\cdot/h)$，其中$K_h(\cdot)$是核函数，其窗宽为$h \to 0$。对于窗宽$h_1$，$h_2$和$h_3$，有

$$\hat{g}(Z_i^\top\theta) = \sum_{i\neq j} K_{h_1}(Z_i^\top\theta - Z_j^\top\theta)(Y_i - X_i^\top\beta)/\sum_{i\neq j} K_{h_1}(Z_i^\top\theta - Z_j^\top\theta),$$

$$\begin{aligned}\hat{g}'(Z_i^\top\theta) =& h_1^{-1}\{\sum_{i\neq j} K'_{h_1}(Z_i^\top\theta - Z_j^\top\theta)(Y_i - X_i^\top\beta)\sum_{i\neq j} K_{h_1}(Z_i^\top\theta - Z_j^\top\theta)\} \\ & - \sum_{i\neq j} K_{h_1}(Z_i^\top\theta - Z_j^\top\theta)(Y_i - X_i^\top\beta) \\ & \times \sum_{i\neq j} K'_{h_1}(Z_i^\top\theta - Z_j^\top\theta)\}/\{\sum_{i\neq j} K_{h_1}(Z_i^\top\theta - Z_j^\top\theta)\}^2,\end{aligned}$$

$$\hat{w}(X_i,Z_i) = \sum_{i\neq j} K_{h_2}(\eta_i - \eta_j)/\sum_{i\neq j} K_{h_2}(\eta_i - \eta_j)e_i^2,$$

$$\begin{aligned}\hat{E}\{\hat{w}(X,Z)|Z_i^\top\theta\} =& \sum_{i\neq j} K_{h_3}(Z_i^\top\theta - Z_j^\top\theta)\hat{w}(X_i,Z_i)/\sum_{i\neq j} K_{h_3}(Z_i^\top\theta - Z_j^\top\theta), \\ \hat{E}\{\hat{w}(X,Z)X|Z_i^\top\theta\} =& \sum_{i\neq j} K_{h_3}(Z_i^\top\theta - Z_j^\top\theta)\hat{w}(X_i,Z_i)X_i/\sum_{i\neq j} K_{h_3}(Z_i^\top\theta - Z_j^\top\theta), \\ \hat{E}\{\hat{w}(X,Z)Z_{-1,i}|Z_i^\top\theta\} =& \sum_{i\neq j} K_{h_3}(Z_i^\top\theta - Z_j^\top\theta)\hat{w}(X_i,Z_i)Z_{-1,i}/\sum_{i\neq j} K_{h_3}(Z_i^\top\theta - Z_j^\top\theta).\end{aligned}$$

下面给出求估计量$(\beta^\top,\theta^\top)^\top$的算法：

步1. 通过求解估计方程(4.1.2)得到的参数$(\beta^\top,\theta^\top)^\top$的估计量的初始值，记为$(\tilde{\beta}^\top,\tilde{\theta}^\top)^\top$；

步2. 计算$\hat{g}(Z_i^\top\theta)$和$\hat{g}'(Z_i^\top\theta)$，令$\tilde{Y}=Y-\hat{g}(Z_i^\top\tilde{\theta})$；

步3. 计算残差e_i，$e_i=Y_i-X_i^\top\tilde{\beta}-\hat{g}(Z_i^\top\tilde{\theta})$。根据方差模型利用$\eta(X_i,Z_i)$，$e_i^2$和 $(\tilde{\beta}^\top,\tilde{\theta}^\top)^\top$ 求$\sigma^2(X,Z)$的估计。令$\hat{w}_i=\hat{\sigma}^{-2}(X_i,Z_i)$；

步4. 利用数据$(X_i\hat{w}_i,Z_i^\top\tilde{\theta},\hat{w}_i)$进行非参数回归得到 $\hat{E}(X\hat{w}(X,Z)|Z_i^\top\tilde{\theta})$和 $\hat{E}(\tilde{w}(X,Z)|Z_i^\top\tilde{\theta})$。令$\tilde{X}_i=X_i-\hat{E}(X\hat{w}(X,Z)|Z_i^\top\tilde{\theta})/\hat{E}(\tilde{w}(X,Z)|Z_i^\top\tilde{\theta})$；

步5. 利用数据$(Z_{-1,i}\hat{w}_i,Z_i^\top\tilde{\theta})$进行非参数回归得到 $\hat{E}(Z_{-1}\hat{w}(X,Z)|Z_i^\top\tilde{\theta})$。令$\tilde{Z}_{-1}=Z_{-1}-\hat{E}(Z_{-1}\hat{w}(X,Z)|Z_i^\top\tilde{\theta})/\hat{E}(\hat{w}(X,Z)|Z_i^\top\tilde{\theta})$；

步6. 利用$\hat{\beta}=\tilde{E}(\tilde{w}\tilde{X}X^\top)^{-1}\tilde{E}(\tilde{w}\tilde{X}\tilde{Y})$作为$\beta$新的估计，其中$\tilde{E}$是指样本平均。通过解如下方程得到$\theta$的估计量$\hat{\theta}$

$$\sum_{i=1}^n\{Y_i-X_i^\top\hat{\beta}-\hat{g}(Z_i^\top\tilde{\theta})\}\hat{w}_i\hat{g}'(Z_i^\top\tilde{\theta})\tilde{Z}_{-1,i}=0.$$

Ma和Zhu在一定条件下证明了：(1) 半参数有效得分为

$$S_{eff}=w\varepsilon\left(X^\top-\frac{E(wX^\top|Z^\top\theta)}{E(w|Z^\top\theta)},g'(Z^\top\theta)\left\{Z_{-1}^\top-\frac{E(wZ_{-1}^\top|Z^\top\theta)}{E(w|Z^\top\theta)}\right\}\right)^\top,\quad(4.1.4)$$

(2)通过估计方程(4.1.3)得到的估计量$(\hat{\beta}^\top,\hat{\theta}^\top)^\top$双重稳健且有效。

在文献[102]中，虽然作者允许协变量是高维的，但协变量的维数确实固定的，不能随着样本容量n一起趋向无穷。

我们先将文献[102]的结论推广到样本维数发散情形，即：当$n\to\infty$时，$p,r\to\infty$。为了得到维数发散情形下估计量$(\hat{\beta}^\top,\hat{\theta}^\top)^\top$的渐近性质，需要如下假设：

假设1. 令$\mathrm{Var}(X_i)=\Sigma_{xi}$，$\mathrm{Var}(Z_i)=\Sigma_{zi}$，对于给定的常数$0<C_1<C_2$，$\Sigma_{xi}$和$\Sigma_{zi}$ 的特征值满足$C_1\le\gamma_1(\Sigma_{xi})\le\cdots\le\gamma_p(\Sigma_{xi})\le C_2$和$C_1\le\gamma_1(\Sigma_{zi})\le\cdots\le\gamma_r(\Sigma_{zi})\le C_2$，$i=1\cdots n$。存在常数$\delta>0$满足$E(\varepsilon^{4+\delta}|X,Z)<\infty$；

假设2. 存在$v(\cdot)$和$\eta=\eta(X,Z)$满足$E(\varepsilon^2|X,Z)=v(\eta)$。存在给定常数$0<C_1<C_2$，使得$0<C_1<v(\cdot)<C_2<\infty$。$\mathrm{Var}(X_i|\eta(X_i,Z_i))$的特征值为有限的正数；

假设3. 存在$v_1(X,Z)$满足

$$\frac{\partial^2E(X|Z^\top\theta)}{\partial\theta_i\partial\theta_j},\frac{\partial^2E(Z|Z^\top\theta)}{\partial\theta_i\partial\theta_j},\frac{\partial^2E(w|Z^\top\theta)}{\partial\theta_i\partial\theta_j},\frac{\partial^2E(wZ|Z^\top\theta)}{\partial\theta_i\partial\theta_j},$$

$$\frac{\partial^2 E(wX|Z^\top\theta)}{\partial\theta_i\partial\theta_j} < v_1(X,Z), Ev_1^2 < \infty, (i,j = 1\cdots p).$$

存在$v_2(X,Z)$满足

$$\frac{\partial^3\eta(X,Z)}{\partial\gamma_i\partial\gamma_j\partial\gamma_l} < v_2(X,Z), Ev_2^2 < \infty, (i,j,l = 1\cdots p+r),$$

其中$(X^\top, Z^\top)^\top = (\gamma_1\cdots\gamma_{p+r})^\top$。存在$v_3(X,Z)$满足

$$\frac{\partial^4 g(Z^\top\beta)}{\partial\theta_i\partial\theta_j\partial\theta_k\partial\theta_l}, \frac{\partial^4 v(\eta)}{\partial\eta_{i_1}\partial\eta_{j_1}\partial\eta_{k_1}\partial\eta_{l_1}} < v_3(X,Z), Ev_3^2 < \infty,$$

其中η的维数为p_1，$i,j,k,l = 1\cdots p$，$i_1, j_1, k_1, l_1 = 1\cdots p_1$；

假设4. 令η和$Z^\top\theta$的密度函数分别为$f_\eta(\eta)$和$f_{Z^\top\theta}(Z^\top\theta)$，且分别满足$0 < \inf f_\eta(\eta) \le \sup f_\eta(\eta) < \infty$和$0 < \inf f_{Z^\top\theta}(Z^\top\theta) \le \sup f_{Z^\top\theta}(Z^\top\theta) < \infty$。存在$v_4(X,Z)$满足

$$\frac{\partial^2 f_{Z^\top\theta}(Z^\top\theta))}{\partial\theta_i\partial\theta_j}, \frac{\partial^2 f_\eta(\eta)}{\partial\eta_k\partial\eta_l} < v_4(X,Z), Ev_4^2 < \infty, (i,j = 1\cdots p;\ k,l = 1\cdots p_1);$$

假设5. 核函数$K_h(\cdot)$在其支撑$[-1,1]$上是对称的且具有连续的一阶导数；

假设6. 窗宽h_i满足$\log^2(n)/(nh_i) \to 0$，$i = 1,2,3$。另外，$nh_1^4 \to \infty$, $nh_1^4 \to 0$，$h_1^4\log^2(n)/h_i \to 0$，$\log^4(n)/(nh_1h_i) \to 0$，$i = 1,2,3,\ h_2 = O(n^{-1/5})$ ，$h_3 = O(n^{-1/5})$；

假设7. 当$n \to \infty p, r \to \infty$时，有$pn^{-1/5} \to 0,\ rn^{-1/5} \to 0$；

假设8. $E\|X\|^4 < \infty$，$E\|Z\|^4 < \infty$，$E\|\varepsilon X\|^4 < \infty$，$E\|\varepsilon Z\|^4 < \infty$，$E|\varepsilon|^4 < \infty$；

假设9. 令

$$\xi_n(\beta,\theta) = w\varepsilon\left(X^\top - \frac{E(wX^\top|Z^\top\theta)}{E(w|Z_i^\top\theta)}, g'(Z^\top\theta)\left\{Z^\top - \frac{E(wZ^\top|Z^\top\theta)}{E(w|Z^\top\theta)}\right\}\right)^\top$$

，$\xi_{nl}(\beta,\theta)$为$\xi_n(\beta,\theta)$的第j个分量，$l = 1,\cdots,p+r$。当$n \to \infty$时，存在常数$C > 0$使得$E(\|\xi_n(\beta,\theta)/\sqrt{p}\|^4) < C$, $E(\|XX^\top\|^4) < C$，$E(\|XZ^\top\|^4) < \infty$，$E(\|ZX^\top\|^4) < C$成立。

如下定理给出了高维情形下估计量$(\hat\beta^\top, \hat\theta^\top)^\top$的渐近性质。

定理 4.1.1 令$(\beta_0^\top, \theta_0^\top)^\top$是参数向量$(\beta^\top, \theta^\top)^\top$的真实值，在假设1-9成立条件下，有$\sqrt{n}AV^{1/2}\{(\hat\beta^\top, \hat\theta^\top)^\top - (\beta_0^\top, \theta_0^\top)^\top\} \xrightarrow{d} N(0, G)$, 其中$\xrightarrow{d}$表示依分布

收敛，$A_1 \in \mathbb{R}^{q_1\times p}$，$A_2 \in \mathbb{R}^{q_2\times r}$，$AA^\top \to G$，$G$是$(q_1+q_2)\times(q_1+q_2)$阶矩阵，$q_1$和$q_2$固定，

$$A=\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, V=\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

$$V_{11}=E\left\{wXX^\top-\frac{E(wX|Z^\top\theta)E(wX^\top|Z^\top\theta)}{E(w|Z^\top\theta)}\right\},$$

$$V_{12}=E\left[g'(Z^\top\theta)\left\{wXZ_{-1}^\top-\frac{E(wX|Z^\top\theta)E(wZ_{-1}^\top|Z^\top\theta)}{E(w|Z^\top\theta)}\right\}\right],$$

$$V_{21}=E\left[g'(Z^\top\theta)\left\{wZ_{-1}X^\top-\frac{E(wZ_{-1}|Z^\top\theta)E(wX^\top|Z^\top\theta)}{E(w|Z^\top\theta)}\right\}\right],$$

$$V_{22}=E\left[g'(Z^\top\theta)^2\left\{wZ_{-1}Z_{-1}^\top-\frac{E(wZ_{-1}|Z^\top\theta)E(wZ_{-1}^\top|Z^\top\theta)}{E(w|Z^\top\theta)}\right\}\right],$$

接下来，我们利用半参数有效得分构造辅助向量：

$$\xi_i(\beta,\theta)=w_i\varepsilon_i\left(X_i^\top-\frac{E(w_iX_i^\top|Z_i^\top\theta)}{E(w_i|Z_i^\top\theta)}, g'(Z_i^\top\theta)\left\{Z_{-1,i}^\top-\frac{E(w_iZ_{-1,i}^\top|Z_i^\top\theta)}{E(w_i|Z_i^\top\theta)}\right\}\right)^\top.$$

当$(\beta^\top,\theta^\top)^\top$为真实参数时，对于$i=1\cdots n$，有$E\{\xi_i(\beta,\theta)\}=0$。设$q=(q_1\cdots q_n)$为概率向量，满足$\sum_{i=1}^n q_i=1$ $q_i\geq 0$，$i=1\cdots n$。关于参数$(\beta^\top,\theta^\top)^\top$的经验似然函数定义为

$$L(\beta,\theta)=\sup\{\prod_{i=1}^n(nq_i):\sum_{i=1}^n q_i=1, q_i\geq 0, \sum_{i=1}^n q_i\xi_i(\beta,\theta)=0\}. \tag{4.1.5}$$

由于(4.1.5)式中包含未知函数$w(X_i,Z_i)$、$g(Z_i^\top\theta)$、$E\{\hat{w}(X,Z)|Z_i^\top\theta\}$、$g'(Z_i^\top\theta)$、$E\{\hat{w}(X,Z)X|Z_i^\top\theta\}$ 和 $E\{\hat{w}(X,Z)Z_{-1}|Z_i^\top\theta\}$，不能直接用来对参数 $(\beta^\top,\theta^\top)^\top$ 进行统计推断。解决这个问题常用方法是首先利用估计量 $\hat{w}(X_i,Z_i)$、$\hat{g}(Z_i^\top\theta)$、$\hat{g}'(Z_i^\top\theta)$、$\hat{E}\{\hat{w}(X,Z)|Z_i^\top\theta\}$、$\hat{E}\{\hat{w}(X,Z)X|Z_i^\top\theta\}$ 和 $\hat{E}\{\hat{w}(X,Z)Z_{-1}|Z_i^\top\theta\}$ 代替上述未知函数然后再做统计推断。用估计量代替未知函数后定义关于参数$(\beta^\top,\theta^\top)^\top$的经验似然函数(称为被估计的经验似然函数)如下：

$$\tilde{L}(\beta,\theta)=\sup\{\prod_{i=1}^n(nq_i):\sum_{i=1}^n q_i=1, q_i\geq 0, \sum_{i=1}^n q_i\hat{\xi}_i(\beta,\theta)=0\}, \tag{4.1.6}$$

其中

$$\hat{\xi}_i(\beta,\theta)=\hat{w}_i\tilde{\varepsilon}_i\left(X_i^\top-\frac{\hat{E}(\hat{w}_iX_i^\top|Z_i^\top\theta)}{\hat{E}(\hat{w}_i|Z_i^\top\theta)}, \hat{g}'(Z_i^\top\theta)\left\{Z_{-1,i}^\top-\frac{\hat{E}(\hat{w}_iZ_{-1,i}^\top|Z_i^\top\theta)}{\hat{E}(\hat{w}_i|Z_i^\top\theta)}\right\}\right)^\top,$$

$\tilde{\varepsilon}_i = Y_i - X_i^\top\beta - \hat{g}(Z_i^\top\theta)$。被估计的对数经验似然比函数为

$$\tilde{l}(\beta,\theta) = -2[\log\{\tilde{L}(\beta,\theta)\} - n\log(n)]. \tag{4.1.7}$$

利用Lagrange乘数方法得到(4.1.6)中的$\{q_i\}_{i=1}^n$

$$q_i = \frac{1}{n}\frac{1}{1+\lambda^\top\hat{\xi}_i(\beta,\theta)},$$

其中λ满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\xi}_i(\beta,\theta)}{1+\lambda^\top\hat{\xi}_i(\beta,\theta)} = 0. \tag{4.1.8}$$

因此，关于参数$(\beta^\top,\theta^\top)^\top$被估计的对数经验似然比函数为

$$\tilde{l}(\beta,\theta) = -2[\log\{\tilde{L}(\beta,\theta)\} - n\log(n)] = 2\sum_{i=1}^{n} log\{1+\lambda^\top\hat{\xi}_i(\beta,\theta)\}. \tag{4.1.9}$$

下面定理给出了对数经验似然比$\tilde{l}(\beta,\theta)$的渐近分布。

定理 4.1.2 令$(\beta_0^\top,\theta_0^\top)^\top$为参数向量$(\beta^\top,\theta^\top)^\top$的真实值，在假设1-9成立条件下，$\tilde{l}(\beta_0,\theta_0)$ 具有渐近正态分布，即

$$\{2(p+r-1)\}^{-1/2}\{\tilde{l}(\beta_0,\theta_0) - (p+r-1)\} \xrightarrow{d} N(0,1).$$

一方面，定理4.1.2可以用来检验假设

$$H_0: (\beta^\top,\theta^\top)^\top = (\beta_0^\top,\theta_0^\top)^\top\ , \quad H_1: (\beta^\top,\theta^\top)^\top \neq (\beta_0^\top,\theta_0^\top)^\top.$$

另一方面，定理4.1.2还可以用来构造参数$(\beta^\top,\theta^\top)^\top$的置信域。令

$$I_\alpha(\beta,\theta) = \{(\beta^\top,\theta^\top)^\top : \tilde{l}(\beta,\theta) \leq (p+r-1) + z_\alpha\sqrt{2(p+r-1)}\},$$

其中z_α是标准正态分布的上$\alpha-$分位数。$I_\alpha(\beta,\theta)$ 为参数$(\beta^\top,\theta^\top)^\top$的置信度趋向于$1-\alpha$的置信域，即

$$P((\beta^\top,\theta^\top)^\top \in I_\alpha(\beta,\theta)) = 1-\alpha+o_p(1).$$

评论 4.1.1 定理4.1.2给出了维数发散(当$n\to\infty$时，p，$r\to\infty$)情形下参数$(\beta^\top,\theta^\top)^\top$被估计的对数经验似然比$\tilde{l}(\beta,\theta)$的渐近分布。当$p$和$r$固定，不随样本容量$n$一起趋向无穷时，$\tilde{l}(\beta_0,\theta_0)$的渐近分布为卡方分布，即，

$$\tilde{l}(\beta_0,\theta_0) \xrightarrow{d} \chi^2_{p+r-1},$$

其中χ_k^2表示自由度为$(p+r-1)$的卡方分布。

在实际应用中，在高维情形下，很少考虑整个参数向量的置信域，因为只要 $p>3$，很难想象或表示参数向量的置信域的形状。通常，我们只考虑参数向量的一个(或有限个)分量的置信区间(置信域)。假设$(\beta^\top,\theta^\top)^\top=(\beta^{(1)\top},\beta^{(2)\top},\theta^\top)^\top$，其中$\beta^{(1)}$是$\beta$的$k$-维分量($k$ 固定)。我们考虑构造 $\beta^{(1)}$ 的置信区间(置信域)。真实参数向量作相应划分：$(\beta_0^\top,\theta_0^\top)^\top=(\beta_0^{(1)\top},\beta_0^{(2)\top},\theta_0^\top)^\top$，$X_i$和$Z_i$类似地划分。

对于固定的 $\beta^{(1)}$，通过估计方程(4.1.3)可以得到估计量$\hat\beta^{(2)}$和$\hat\theta$。定义新的辅助随机向量

$$\hat{\hat{\xi}}_i(\beta^{(1)})=\hat w_i\left\{Y_i-X_i^{(1)\top}\beta^{(1)}-X_i^{(2)\top}\hat\beta^{(2)}-\hat g(Z_i^\top\hat\theta)\right\}\left(X_i^{(1)}-\frac{\hat E(\hat w_iX_i^{(1)}|Z_i^\top\hat\theta)}{\hat E(\hat w_i|Z_i^\top\hat\theta)}\right).$$

关于参数分量 $\beta^{(1)}$ 被估计的对数经验似然比为

$$\tilde l(\beta^{(1)})=2\sum_{i=1}^n\log\{1+\lambda^{(1)\top}\hat{\hat{\xi}}_i(\beta^{(1)})\}. \tag{4.1.10}$$

其中 $\lambda^{(1)}\in\mathbb{R}^k$ 是 Lagrange 乘数。下面定理给出了$\tilde l(\beta^{(1)})$的渐近分布。

定理 4.1.3 在定理4.1.2相同假设条件下，有

$$\tilde l(\beta_0^{(1)})\xrightarrow{d}\chi_k^2.$$

我们可以根据定理4.1.3构造参数分量$\beta^{(1)}$的置信区间或置信域。任意给定$0<\alpha<1$，存在c_α满足$P(\chi_k^2>c_\alpha)=\alpha$，

$$I_\alpha(\beta^{(1)})=\{\beta^{(1)}\in\mathbb{R}^k|\tilde l(\beta^{(1)})\le c_\alpha\}$$

是参数分量$\beta^{(1)}$的置信度趋向于$1-\alpha$的渐近置信区间或置信域。

§4.2 两种特性情形：部分线性模型和单指标模型

本节，我们在高维情形下讨论模型(4.1.1)的两种特殊情形的经验似然推断问题。首先考虑异方差部分线性模型，即模型(4.1.1)中单指标部分不含有参数θ，其模型如下：

$$Y_i=X_i^\top\beta+g(Z_i)+\varepsilon_i,\quad (i=1,\cdots,n). \tag{4.2.1}$$

对于模型(4.2.1)，加权估计方程(4.1.3)变为

$$n^{-1/2}\sum_{i=1}^{n}\hat{w}(X_i,Z_i)\left(Y_i-X_i^\top\beta-\hat{g}(Z_i)\right)\left[X_i-\frac{\hat{E}\{\hat{w}(X,Z)X|Z_i\}}{\hat{E}\{\hat{w}(X,Z)|Z_i\}}\right]=0. \quad (4.2.2)$$

参数β的有效估计量$\hat{\beta}$可以通过求解加权估计方程(4.2.2)得到。当$n\to\infty$，$p\to\infty$时，如下定理给出了估计量$\hat{\beta}$的理论性质。

定理 4.2.1 在假设条件1-9成立条件下，如果β_0为参数β的真实值，那么

$$\sqrt{n}A_nV_1^{-1/2}(\hat{\beta}-\beta_0)\xrightarrow{d}N(0,G),$$

其中$A_n\in\mathbb{R}^{q\times p}$满足$A_nA_n^\top\to G$，$G$是$q\times q$阶矩阵，$q$固定，

$$V_1=\left[E\left\{wXX^\top-\frac{E(wX|Z)E(wX|Z)^\top}{E(w|Z)}\right\}\right]^{-1}.$$

定理4.2.1将Ma et al. (2006)[104]中的结论推广到了样本维数p发散情形。

重新定义辅助向量

$$\hat{\xi}_i(\beta)=\hat{w}_i\left\{Y_i-X_i^\top\beta-\hat{g}(Z_i)\right\}\left\{X_i-\frac{\hat{E}(\hat{w}_iX_i|Z_i)}{\hat{E}(\hat{w}_i|Z_i)}\right\}.$$

关于β参数被估计的对数经验似然比函数为

$$\tilde{l}_1(\beta)=2\sum_{i=1}^{n}\log\{1+\lambda^\top\hat{\xi}_i(\beta)\}. \quad (4.2.3)$$

下面定理给出了$\tilde{l}(\beta)$的渐近分布。

定理 4.2.2 在假设1-9成立条件下，如果β_0是参数β的真实值，那么有

$$(2p)^{-1/2}(\tilde{l}_1(\beta_0)-p)\xrightarrow{d}N(0,1).$$

定理4.2.2可以用来构造参数β的经验似然置信域。β的置信度趋向于$(1-\alpha)100\%$的置信域为

$$I_\alpha(\beta)=\{\beta:\tilde{l}_1(\beta)\le p+z_\alpha\sqrt{2p}\}.$$

在定理4.2.2中，当维数p固定时，$\tilde{l}_1(\beta_0)$的渐近分布为卡方分布，这正是 Lu (2009)[16]中所得到的结论。

类似于定理4.1.3，我们可以构造关于参数β的分量的置信区间或置信域。令$\beta = (\beta^{(1)\top}, \beta^{(2)\top})^\top$，其中$\beta^{(1)}$是$\beta$的$k$-维分量。真实参数向量做相应划分：$\beta_0 = (\beta_0^{(1)\top}, \beta_0^{(2)\top})^\top$，$X_i$ 类似地划分。

对于固定的 $\beta^{(1)}$，通过估计方程(4.2.2)可以得到估计量$\hat{\beta}^{(2)}$。令

$$\tilde{\tilde{\xi}}_i(\beta^{(1)}) = \hat{w}_i \left\{ Y_i - X_i^{(1)\top}\beta^{(1)} - X_i^{(2)\top}\hat{\beta}^{(2)} - \hat{g}(Z_i) \right\} \left\{ X_i^{(1)} - \frac{\hat{E}(\hat{w}_i X_i^{(1)}|Z_i)}{\hat{E}(\hat{w}_i|Z_i)} \right\}.$$

关于参数$\beta^{(1)}$的被估计的对数经验似然比记为

$$\tilde{\tilde{l}}(\beta^{(1)}) = 2\sum_{i=1}^{n} \log\{1 + \lambda^{(1)\top}\tilde{\tilde{\xi}}_i(\beta^{(1)})\}. \tag{4.2.4}$$

定理 4.2.3 在定理4.2.2相同假设条件成立下，有

$$\tilde{\tilde{l}}(\beta_0^{(1)}) \xrightarrow{d} \chi_k^2.$$

根据定理4.2.3，$\beta^{(1)}$的渐近经验似然置信区间或置信域为

$$\tilde{I}_\alpha(\beta^{(1)}) = \{\beta^{(1)} | \tilde{\tilde{l}}(\beta^{(1)}) \le \chi_k^2(\alpha)\}, \quad 0 < \alpha < 1.$$

下面考虑高维情形下的异方差单指标模型

$$Y_i = g(Z_i^\top\theta) + \varepsilon_i, \quad (i = 1, \cdots, n). \tag{4.2.5}$$

的经验似然推断。上述模型相当于模型(4.1.1)中参数β为零。相应加权估计方程为

$$n^{-1/2}\sum_{i=1}^{n} \hat{w}\{Y_i - \hat{g}(Z_i^\top\theta)\}\hat{g}'(Z_i^\top\tilde{\theta}) \left[Z_{-1,i} - \frac{\hat{E}\{\hat{w}Z_{-1}|Z_i^\top\tilde{\theta}\}}{\hat{E}\{\hat{w}|Z_i^\top\tilde{\theta}\}} \right] = 0, \tag{4.2.6}$$

其中$\tilde{\theta}$是如下方程的解

$$n^{-1/2}\sum_{i=1}^{n} \hat{w}\{Y_i - f(Z_i^\top\theta)\}[a(Z_i) - \tilde{E}\{a(Z_i)|Z_i^\top\theta)\}] = 0.$$

参数θ的有效估计$\hat{\theta}$可通过求解估计方程(4.2.6)得到。下面定理给出了$\hat{\theta}$的渐近性质。

定理 4.2.4 在假设1-9成立下，$\sqrt{n}A_nV_2^{-1/2}(\hat{\theta} - \theta) \xrightarrow{d} N(0, G)$，其中$A_n \in \mathbb{R}^{q\times p}$满足$A_nA_n^\top \to G$，$G$是$q \times q$阶矩阵，$q$固定，

$$V_2 = E\left[g'(Z^\top\theta)^2 \left\{ wZ_{-1}Z_{-1}^\top - \frac{E(wZ_{-1}|Z^\top\theta)E(wZ_{-1}^\top|Z^\top\theta)}{E(w|Z^\top\theta)} \right\} \right].$$

类似地，重新购置相应的辅助向量

$$\hat{\xi}_i(\theta) = \hat{w}_i\{Y_i - \hat{g}(Z_i^\top\theta)\}\hat{g}'(Z_i^\top\theta)\left\{Z_{-1,i} - \frac{\hat{E}(\hat{w}_i Z_{-1,i}|Z_i^\top\theta)}{\hat{E}(\hat{w}_i|Z_i^\top\theta)}\right\}.$$

被估计的经验似然比为

$$\tilde{l}_2(\theta) = 2\sum_{i=1}^{n}\log\{1+\lambda^\top\hat{\xi}_i(\theta)\}.$$

定理 4.2.5 在假设1-9成立条件下，如果θ_0是参数θ的真实值，那么有

$$\{2(r-1)\}^{-1/2}\{\tilde{l}_2(\theta_0)-(r-1)\} \xrightarrow{d} N(0,1).$$

当r固定时，有$\tilde{l}_2(\theta_0) \xrightarrow{d} \chi^2_{r-1}$。根据定理4.2.5，参数$\theta$的置信度趋向于$(1-\alpha)100\%$的渐近置信域为

$$I_\alpha(\theta) = \{\theta : \tilde{l}_2(\theta) \le (r-1) + z_\alpha\sqrt{2(r-1)}\}.$$

§4.3 定理的证明

本节将给出本章几个定理的证明。在整个证明过程中，不同地方出现的 C 的具体的值可能不同，但都表示一个正的常数。令$\tilde{l}(\lambda,\beta,\theta) = n^{-1}\sum_{i=1}^n \log\{1+\lambda^\top\hat{\xi}_i(\beta,\theta)\}$，$\bar{\hat{\xi}}(\beta,\theta) = n^{-1}\sum_{i=1}^n \hat{\xi}_i(\beta,\theta)$，$a_n = O_p\{(p/n)^{1/2}\}$。另外。我们用$\|A\|$表示矩阵$A$的Frobenius范数，$\|A\| = \{tr(A^\top A)\}^{\frac{1}{2}}$。

定理4.1.1的证明： 先将下式展开，得

$$
\begin{aligned}
0 =&\frac{1}{\sqrt{n}}A_2\sum_{i=1}^{n}\left\{Y_i - X_i^\top\hat{\beta} + \hat{g}\prime(Z_i^\top\hat{\theta})\right\}\hat{w}_i\hat{g}'(Z_i^\top\hat{\theta})\left\{Z_i - \frac{\hat{E}(\hat{w}Z|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}|Z_i^\top\hat{\theta})}\right\}\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{w}_i\hat{g}'(Z_i^\top\hat{\theta})A_2\left\{Z_i - \frac{E(wZ|Z_i^\top\theta_0)}{E(w|Z_i^\top\theta_0)}\right\}X_i^\top(\beta_0-\hat{\beta})\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{w}_i\hat{g}'(Z_i^\top\hat{\theta})A_2\left\{\frac{E(wZ|Z_i^\top\theta_0)}{E(w|Z_i^\top\theta_0)} - \frac{\hat{E}(\hat{w}Z|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}|Z_i^\top\hat{\theta})}\right\}X_i^\top(\beta_0-\hat{\beta})\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{g(Z_i^\top\theta_0)-\hat{g}(Z_i^\top\theta_0)\right\}\hat{w}_i\hat{g}'(Z_i^\top\hat{\theta})A_2\left\{Z_i - \frac{E(wZ|Z_i^\top\theta_0)}{E(w|Z_i^\top\theta_0)}\right\}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{g(Z_i^\top\theta_0)-\hat{g}(Z_i^\top\theta_0)\right\}\hat{w}_i\hat{g}'(Z_i^\top\hat{\theta})A_2\left\{\frac{E(wZ|Z_i^\top\theta_0)}{E(w|Z_i^\top\theta_0)} - \frac{\hat{E}(\hat{w}Z|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}|Z_i^\top\hat{\theta})}\right\}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{g(Z_i^\top\theta_0)-\hat{g}(Z_i^\top\hat{\theta})\right\}\hat{w}_i\hat{g}'(Z_i^\top\hat{\theta})A_2\left\{Z_i - \frac{\hat{E}(\hat{w}Z|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}|Z_i^\top\hat{\theta})}\right\}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\hat{w}_i\hat{g}'(Z_i^\top\hat{\theta})A_2\left\{Z_i - \frac{\hat{E}(\hat{w}Z|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}|Z_i^\top\hat{\theta})}\right\}. \qquad (4.3.1)
\end{aligned}
$$

类似于[102]中命题2的证明，由(4.1.3)式的第二个方程可得

$$
\begin{aligned}
&A_2E\left[wg'(Z^\top\theta_0)\left\{Z - \frac{E(wZ|Z^\top\theta_0)}{E(w|Z^\top\theta_0)}\right\}X^\top\right]\sqrt{n}(\hat{\beta}-\beta_0)\\
&+A_2E\left[w\{g'(Z^\top\theta_0)\}^2\left\{Z - \frac{E(wZ|Z^\top\theta_0)}{E(w|Z^\top\theta_0)}\right\}Z^\top\right]\sqrt{n}(\hat{\theta}-\theta_0)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i w_i g'(Z_i^\top\theta_0)A_2\left\{Z_i - \frac{E(wZ|Z_i^\top\theta_0)}{E(w|Z_i^\top\theta_0)}\right\}+o_p(1). \qquad (4.3.2)
\end{aligned}
$$

类似地，由(4.1.3)式的第一个方程可得

$$
\begin{aligned}
&A_1E\left[w\left\{X - \frac{E(wX|Z^\top\theta_0)}{E(w|Z^\top\theta_0)}\right\}X^\top\right]\sqrt{n}(\hat{\beta}-\beta_0)\\
&+A_1E\left[wg'(Z^\top\theta_0)\left\{X - \frac{E(wX|Z^\top\theta_0)}{E(w|Z^\top\theta_0)}\right\}Z^\top\right]\sqrt{n}(\hat{\theta}-\theta_0)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i w_i A_1\left\{X_i - \frac{E(wX|Z_i^\top\theta_0)}{E(w|Z_i^\top\theta_0)}\right\}+o_p(1). \qquad (4.3.3)
\end{aligned}
$$

联合(4.3.2)和(4.3.3)，有

$$
AV^{1/2}\begin{pmatrix}\hat{\beta}-\beta_0\\ \hat{\theta}-\theta_0\end{pmatrix} = AV^{-1/2}\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i w_i\left\{X_i - \frac{E(wX|Z_i^\top\theta_0)}{E(w|Z_i^\top\theta_0)}\right\}\\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i w_i g'(Z_i^\top\theta_0)\left\{Z_i - \frac{E(wZ|Z_i^\top\theta_0)}{E(w|Z_i^\top\theta_0)}\right\}\end{pmatrix}+o_p(1).
$$

应用Lindeberg-Feller中心极限定理可得

$$\sqrt{n}AV^{1/2}\left\{(\hat{\beta}^\top,\hat{\theta}^\top)^\top-(\beta_0^\top,\theta_0^\top)^\top\right\}\xrightarrow{d}N(0,G).$$

综上所述，定理4.1.1证明完毕。 □

证明定理4.1.2之前，先给出下面引理。

引理 4.3.1 在定理4.1.2相同假设条件成立下，对于任意的$\lambda\in\mathbb{R}^p$，如果$\lambda=O_p(a_n)$，那么有

$$\max_{1\le i\le n}\|\hat{\xi}_i(\beta,\theta)\|=o_p(n^{1/4}\sqrt{p}),\quad \max_{1\le i\le n}|\lambda^\top\hat{\xi}_i(\beta,\theta)|=o_p(1).$$

证明： 由假设8和9，对于任给的$\epsilon>0$，有

$$\begin{aligned}P\{\max_{1\le i\le n}\|\xi_i(\beta,\theta)\|\le n^{1/4}\sqrt{p}\epsilon\}&\le\sum_{i=1}^nP\{\|\xi_i(\beta,\theta)\|\le n^{1/4}\sqrt{p}\epsilon\}\\&\le\frac{1}{np^2\epsilon^4}\sum_{i=1}^nE\|\xi_i(\beta,\theta)\|^4\\&=\frac{1}{\epsilon^k}E\|\xi_1(\beta,\theta)/\sqrt{p}\|^4.\end{aligned}\tag{4.3.4}$$

利用柯西-施瓦茨不等式得

$$\|\xi_1(\beta,\theta)/\sqrt{p}\|^4\le 1/p\sum_{l=1}^{p+r}|\xi_{1l}(\beta,\theta)|^4,$$

其中 $\xi_{1l}(\beta,\theta)$ 是 $\xi_1(\beta,\theta)$的第l个分量。由(4.3.4)式，我们有

$$\max_{1\le i\le n}\|\xi_i(\beta,\theta)\|=o_p(n^{1/4}\sqrt{p}).$$

类似于(4.3.2)和(4.3.4)式的证明，容易得出

$$\|\hat{\xi}_i(\beta,\theta)\|=\|\xi_i(\beta,\theta)\|+O_p(p).$$

由假设7，有

$$\|\hat{\xi}_i(\beta,\theta)\|=o_p(n^{1/4}\sqrt{p})+O_p(p)=o_p(n^{1/4}\sqrt{p}),$$

以及对于任意给定的$\lambda\in\mathbb{R}^p$，如果$\lambda=O_p(a_n)$，那么有

$$\max_{1\le i\le n}|\lambda^\top\hat{\xi}_i(\beta,\theta)|=o_p(1).$$

综上所述，引理4.3.1证明完毕。 □

引理 4.3.2 在定理4.2.2假设条件成立下，有$\|S_n - V\| = O_p(p/\sqrt{n})$，其中$S_n = 1/n\sum_{i=1}^n \hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top$。

证明： 类似于Chen et al.(2009)[56]中引理4的证明，容易得出

$$tr\{(S_n - V)^2\} = O_p(p^2/n).$$

因此，由Frobenius范数定义，有

$$\|S_n - V\| = \{tr[(S_n - V)^\top(S_n - V)]\}^{1/2} = O_p(p/\sqrt{n}).$$

引理4.3.2证明完毕。 □

引理 4.3.3 在定理4.1.2假设条件成立下，有$\|\lambda\| = O_p(a_n)$，其中λ为方程(4.1.8)的根。

证明： 由(4.1.8)，有$\lambda \in \mathbb{R}^{p+r}$ 满足

$$0 = \frac{1}{n}\sum_{i=1}^n \frac{\hat{\xi}_i(\beta,\theta)}{1+\lambda^\top\hat{\xi}_i(\beta,\theta)} =: \psi(\lambda).$$

令$\lambda = \rho\alpha$, 其中$\rho \geq 0$，$\alpha \in \mathbb{R}^{p+r}$，$\|\alpha\| = 1$。将$1/(1+\lambda^\top\hat{\xi}_i(\beta,\theta)) = 1 - \lambda^\top\hat{\xi}_i(\beta,\theta)/(1+\lambda^\top\hat{\xi}_i(\beta,\theta))$代入$\alpha^\top\psi(\lambda) = 0$，得

$$|\alpha^\top\bar{\hat{\xi}}_i(\beta,\theta)| \geq \frac{\rho}{1+\rho\max\limits_{1\leq i\leq n}\|\xi_i(\beta,\theta)\|}\alpha^\top S_n\alpha,$$

其中：$S_n = \frac{1}{n}\sum\limits_{i=1}^n \hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top$。因为

$$0 < 1+\lambda^\top\hat{\xi}_i(\beta,\theta) \leq 1+\rho\max_{1\leq i\leq n}\|\xi_i(\beta,\theta)\|,$$

我们有

$$\rho[\alpha^\top S_n\alpha - \alpha^\top\hat{\xi}_i(\beta,\theta)\max_{1\leq i\leq n}\|\xi_i(\beta,\theta)\|] \leq |\alpha^\top\bar{\hat{\xi}}_i(\beta,\theta)|. \tag{4.3.5}$$

又$|\alpha^\top\bar{\hat{\xi}}_i(\beta,\theta)| \leq \|\bar{\hat{\xi}}_i(\beta,\theta)\| = O_p(\sqrt{p/n})$和引理4.3.1，因此

$$\max_{1\leq i\leq n}\|\xi_i(\beta,\theta)\||\alpha^\top\bar{\hat{\xi}}_i(\beta,\theta)| = o_p(1). \tag{4.3.6}$$

联合(4.3.5)和(4.3.6)得

$$|\rho[\alpha^\top S_n\alpha + o_{p(1)}]| = O_p(\sqrt{p/n}).$$

由引理4.3.2知，存在常数$C_1>0$，当$n\to\infty$时，有$P(\alpha^\top S_n\alpha\ge\frac{1}{2}C_1)\to 1$。因此，$\rho=O_p(\sqrt{p/n})$，即$\|\lambda\|=\rho=O_p(\sqrt{p/n})$。引理4.3.3证明完毕。 □

引理 4.3.4 在定理4.1.2假设条件成立下，当$n\to\infty$时，有

$$\{2(p+r-1)\}^{-1}\{(\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{\xi}_i(\beta,\theta))^\top V^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{\xi}_i(\beta,\theta))-(p+r-1)\}\xrightarrow{d}N(0,1).$$

证明： 本引理可以直接利用Hall和Hyde(1980)[105]中的鞅中心极限定理证明，因此省略。 □

引理 4.3.5 在定理4.1.2假设条件成立下，有

$$\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{\xi}_i(\beta,\theta)\}^\top(S_n^{-1}-V^{-1})\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{\xi}_i(\beta,\theta)\}=o_p(\sqrt{p}).$$

证明： 令$D_n=V^{-1/2}S_nV^{-1/2}-I_{p+r}$，其中$I_{p+r}$为$p+r$阶单位矩阵。

$$\begin{aligned}S_n^{-1}-V^{-1}&=V^{-1/2}(V^{1/2}S_n^{-1}V^{1/2}-I_{p+r})V^{-1/2}\\&=V^{-1/2}\{-D_n+D_n^2+D_n^2(V^{1/2}S_n^{-1}V^{1/2}-I_{p+r})\}V^{-1/2}.\end{aligned}$$

容易验证

$$\begin{aligned}tr(S_n-V)&=tr(V^{1/2}(V^{-1/2}S_nV^{-1/2}-I_{p+r})V^{1/2})\\&=tr(D_nVD_nV)\ge\gamma_1^2(V)tr(D_n^2),\end{aligned}$$

其中$\gamma_1(V)$是V最大特征值。类似于Chen et al.(2009)[56]中引理4的证明，有

$$tr(D_n^2)\le tr\{(S_n-V)^2\}=O_p(p^2/n).$$

因此

$$\begin{aligned}tr(S_n^{-1}-V^{-1})^2&\le 2tr\{V^{-2}(-D_n+D_n^2)^2\}+2tr\{D_n^4(S_n^{-1}-V^{-1})^2\}\\&\le 2tr\{V^{-2}(-D_n+D_n^2)^2\}+2\{tr(D_n^2)\}^2tr\{(S_n^{-1}-V^{-1})^2\}\\&=2tr\{V^{-2}(-D_n+D_n^2)^2\}+o_p(tr\{(S_n^{-1}-V^{-1})^2\})\\&=o_p(p^2/n).\end{aligned}$$

由于$\|\frac{1}{n}\sum_{i=1}^n\hat{\xi}_i(\beta,\theta)\|=O_p(\sqrt{p/n})$，我们有

$$\begin{aligned}&\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{\xi}_i(\beta,\theta)\}^\top(S_n^{-1}-V^{-1})\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{\xi}_i(\beta,\theta)\}\\&\le n\|\frac{1}{n}\sum_{i=1}^n\hat{\xi}_i(\beta,\theta)\|^2\sqrt{tr(S_n^{-1}-V^{-1})^2}=o_p(\sqrt{p}).\end{aligned}$$

综上所述，引理4.3.5证明完毕。 □

定理4.1.2的证明： 令$W_i = \lambda^\top \hat{\xi}_i(\beta,\theta), i = 1,\cdots,n$。将(4.1.8)式展开得

$$0 = \sum_{i=1}^n \frac{\hat{\xi}_i(\beta,\theta)}{1+\lambda^\top \hat{\xi}_i(\beta,\theta)} = \sum_{i=1}^n \hat{\xi}_i(\beta,\theta) - \sum_{i=1}^n \{\hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top\}\lambda + R_n, \qquad (4.3.7)$$

其中

$$R_n = \sum_{i=1}^n \frac{\hat{\xi}_i(\beta,\theta)(\lambda^\top \hat{\xi}_i(\beta,\theta))^2}{(1+\vartheta_i)^3}, \quad |\vartheta_i| \le |\lambda^\top \hat{\xi}_i(\beta,\theta)|.$$

由引理4.3.1，有$\max_{1\le i\le n}|\vartheta_i| = o_p(1)$。因此，

$$R_n = R_{n1}\{1+o_p(1)\}, \quad R_{n1} = \sum_{i=1}^n \hat{\xi}_i(\beta,\theta)(\lambda^\top \hat{\xi}_i(\beta,\theta))^2.$$

利用引理4.3.1和引理4.3.3可得

$$\|n^{-1}R_n\| \le C\|\lambda\|^2 \max_{1\le i\le n}\|\hat{\xi}_i(\beta,\theta)\| n^{-1}\sum_{i=1}^n \|\hat{\xi}_i(\beta,\theta)\|^2 = o_p(a_n). \qquad (4.3.8)$$

整理(4.3.7)得出

$$\lambda = \{\sum_{i=1}^n \hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top\}^{-1}\sum_{i=1}^n \hat{\xi}_i(\beta,\theta) + \{\sum_{i=1}^n \hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top\}^{-1}R_n.$$

将 $\log(1+\lambda^\top \hat{\xi}_i(\beta,\theta))$ 在 $\lambda^\top \hat{\xi}_i(\beta,\theta)$ 处泰勒展开知，存在满足 $|\zeta_i| \le |\lambda^\top \hat{\xi}_i(\beta,\theta)|$ 的 ζ_i，使得下式成立

$$\log(1+\lambda^\top \hat{\xi}_i(\beta,\theta)) = \lambda^\top \hat{\xi}_i(\beta,\theta) - \frac{\{\lambda^\top \hat{\xi}_i(\beta,\theta)\}^2}{2} + \frac{\{\lambda^\top \hat{\xi}_i(\beta,\theta)\}^3}{3(1+\zeta_i)^4}.$$

将λ的表达式代入上式整理得

$$\begin{aligned}\tilde{l}(\beta,\theta) &= (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta))^\top\{\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top\}^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)) \\ &\quad - \frac{1}{n}R_n^\top\{\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top\}^{-1}R_n + \sum_{i=1}^{n}\frac{2\{\lambda^\top\hat{\xi}_i(\beta,\theta)\}^3}{3(1+\zeta_i)^4} \\ &= (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta))^\top V^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)) \\ &\quad + (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta))^\top[\{\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top\}^{-1} - V^{-1}](\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)) \\ &\quad - \frac{1}{n}R_n^\top\{\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top\}^{-1}R_n + \frac{2}{3}\sum_{i=1}^{n}\{\lambda^\top\hat{\xi}_i(\beta,\theta)\}^3\{1+o_p(1)\}.\end{aligned} \tag{4.3.9}$$

由引理4.3.5可得

$$(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta))^\top[\{\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top\}^{-1} - V^{-1}](\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)) = o_p(1). \tag{4.3.10}$$

利用引理4.3.1-引理4.3.3以及(4.3.8)式，我们有

$$\frac{1}{n}R_n^\top\{\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top\}^{-1}R_n = o_p(1), \tag{4.3.11}$$

和

$$\frac{2}{3}\sum_{i=1}^{n}\{\lambda^\top\hat{\xi}_i(\beta,\theta)\}^3\{1+o_p(1)\} = o_p(\sqrt{p}). \tag{4.3.12}$$

联合(4.3.9)-(4.3.12)式得

$$\tilde{l}(\beta,\theta) = (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta))^\top V^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i(\beta,\theta)) + o_p(\sqrt{p}).$$

最后，由引理4.3.4和引理4.3.5知，定理4.1.2结论成立。证毕。 □

定理4.1.3的证明： 首先，我们证明：$\max_{1\le i\le n}\|\hat{\hat{\xi}}_i(\beta^{(1)})\| = o_p(n^{1/2})$。容易

验证

$$\begin{aligned}
\hat{\hat{\xi}}_i(\beta) =& \hat{w}_i\{Y_i - X_i^{(1)\top}\beta^{(1)} - X_i^{(2)\top}\hat{\beta}^{(2)} - \hat{g}(Z_i^\top\hat{\theta})\}\left\{X_i^{(1)} - \frac{\hat{E}(\hat{w}_i X_i^{(1)}|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}_i|Z_i^\top\hat{\theta})}\right\} \\
=& \{w_i(1+o_p(1))\}\left\{\varepsilon_i + X_i^\top(\beta-\hat{\beta}) + X_i^{(1)\top}(\hat{\beta}^{(1)}-\beta^{(1)}) + (g(Z_i^\top\hat{\theta}) - \hat{g}(Z_i^\top\hat{\theta}))\right\} \\
&\times\left\{\left(X_i^{(1)} - \frac{E(w_i X_i^{(1)}|Z_i^\top\hat{\theta})}{E(w_i|Z_i^\top\hat{\theta})}\right) + \left(\frac{E(w_i X_i^{(1)}|Z_i^\top\hat{\theta})}{E(w_i|Z_i^\top\hat{\theta})} - \frac{\hat{E}(\hat{w}_i X_i^{(1)}|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}_i|Z_i^\top\hat{\theta})}\right)\right\} \\
=& M_{i1} + M_{i2} + M_{i3} + M_{i4} + M_{i5} + M_{i6} + M_{i7} + M_{i8}, \qquad (4.3.13)
\end{aligned}$$

其中

$$\begin{aligned}
M_{i1} =& \{w_i(1+o_p(1))\}\varepsilon_i\left\{X_i^{(1)} - \frac{E(w_i X_i^{(1)}|Z_i^\top\hat{\theta})}{E(w_i|Z_i^\top\hat{\theta})}\right\}, \\
M_{i2} =& \{w_i(1+o_p(1))\}\varepsilon_i\left\{\frac{E(w_i X_i^{(1)}|Z_i^\top\hat{\theta})}{E(w_i|Z_i^\top\hat{\theta})} - \frac{\hat{E}(\hat{w}_i X_i^{(1)}|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}_i|Z_i^\top\hat{\theta})}\right\}, \\
M_{i3} =& \{w_i(1+o_p(1))\}X_i^\top(\beta-\hat{\beta})\left\{X_i^{(1)} - \frac{E(w_i X_i^{(1)}|Z_i^\top\hat{\theta})}{E(w_i|Z_i^\top\hat{\theta})}\right\}, \\
M_{i4} =& \{w_i(1+o_p(1))\}X_i^\top(\beta-\hat{\beta})\left\{\frac{E(w_i X_i^{(1)}|Z_i^\top\hat{\theta})}{E(w_i|Z_i^\top\hat{\theta})} - \frac{\hat{E}(\hat{w}_i X_i^{(1)}|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}_i|Z_i^\top\hat{\theta})}\right\}, \\
M_{i5} =& \{w_i(1+o_p(1))\}X_i^{(1)\top}(\hat{\beta}^{(1)}-\beta^{(1)})\left\{X_i^{(1)} - \frac{E(w_i X_i^{(1)}|Z_i^\top\hat{\theta})}{E(w_i|Z_i^\top\hat{\theta})}\right\}, \\
M_{i6} =& \{w_i(1+o_p(1))\}X_i^{(1)\top}(\hat{\beta}^{(1)}-\beta^{(1)})\left\{\frac{E(w_i X_i^{(1)}|Z_i^\top\hat{\theta})}{E(w_i|Z_i\hat{\theta})} - \frac{\hat{E}(\hat{w}_i X_i^{(1)}|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}_i|Z_i^\top\hat{\theta})}\right\}, \\
M_{i7} =& \{w_i(1+o_p(1))\}(g(z_i^\top\hat{\theta}) - \hat{g}(Z_i^\top\hat{\theta}))\left\{X_i^{(1)} - \frac{E(w_i X_i^{(1)}|Z_i^\top\hat{\theta})}{E(w_i|Z_i^\top\hat{\theta})}\right\}, \\
M_{i8} =& \{w_i(1+o_p(1))\}(g(z_i^\top\hat{\theta}) - \hat{g}(Z_i^\top\hat{\theta}))\left\{\frac{E(w_i X_i^{(1)}|Z_i^\top\hat{\theta})}{E(w_i|Z_i^\top\hat{\theta})} - \frac{\hat{E}(\hat{w}_i X_i^{(1)}|Z_i^\top\hat{\theta})}{\hat{E}(\hat{w}_i|Z_i^\top\hat{\theta})}\right\}.
\end{aligned}$$

由(4.3.13), 我们有

$$\begin{aligned}
\max_{1\le i\le n}\|\hat{\hat{\xi}}_i(\beta^{(1)})\| \le& \max_{1\le i\le n}\|M_{i1}\| + \max_{1\le i\le n}\|M_{i2}\| + \max_{1\le i\le n}\|M_{i3}\| + \max_{1\le i\le n}\|M_{i4}\| \\
&+ \max_{1\le i\le n}\|M_{i5}\| + \max_{1\le i\le n}\|M_{i6}\| + \max_{1\le i\le n}\|M_{i7}\| + \max_{1\le i\le n}\|M_{i8}\|.
\end{aligned}$$

类似于Ma和Zhu(2013)[103]中命题2的证明可得

$$\max_{1\le i\le n}\|M_{l1}\| = o_p(n^{1/2}), l = 1, \cdots, 8.$$

于是，有$\max_{1\le i\le n}\|\hat{\tilde{\xi}}_i(\beta^{(1)})\| = o_p(n^{1/2})$。另外，类似于Li和Wang(2003)[44]中定理3.1的证明过程，容易得出：当$n\to\infty$时，有

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\tilde{\xi}}_i(\beta^{(1)}) \xrightarrow{d} N(0, V_1(\beta^{(1)})), \tag{4.3.14}$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\tilde{\xi}}_i(\beta^{(1)})\hat{\tilde{\xi}}_i^{\top}(\beta^{(1)}) \xrightarrow{p} V_1(\beta^{(1)}), \tag{4.3.15}$$

其中

$$V_1(\beta^{(1)}) = E\left\{wX^{(1)}X^{(1)\top} - \frac{E(wX^{(1)}|Z^{\top}\hat{\theta})E(wX^{(1)\top}|Z^{\top}\hat{\theta})}{E(w|Z^{\top}\hat{\theta})}\right\}.$$

由$\max_{1\le i\le n}\|\hat{\tilde{\xi}}_i(\beta^{(1)})\| = o_p(n^{1/2})$以及(4.1.10)的泰勒展开式得

$$\tilde{l}(\beta^{(1)}) = 2\sum_{i=1}^{n}\lambda^{(1)\top}\hat{\tilde{\xi}}_i(\beta^{(1)}) - \sum_{i=1}^{n}\{\lambda^{(1)\top}\hat{\tilde{\xi}}_i(\beta^{(1)})\}^2 + o_p(1). \tag{4.3.16}$$

类似于Owen(1990)[3]中定理1的证明，有

$$\sum_{i=1}^{n}\{\lambda^{(1)\top}\hat{\tilde{\xi}}_i(\beta^{(1)})\}^2 = \sum_{i=1}^{n}\lambda^{(1)\top}\hat{\tilde{\xi}}_i(\beta^{(1)}) + o_p(1), \tag{4.3.17}$$

$$\lambda^{(1)} = \{\sum_{i=1}^{n}\hat{\tilde{\xi}}_i(\beta^{(1)})\hat{\tilde{\xi}}_i(\beta^{(1)})^{\top}\}^{-1}\sum_{i=1}^{n}\hat{\tilde{\xi}}_i(\beta^{(1)}) + o_p(n^{-1/2}). \tag{4.3.18}$$

联合(4.3.16)-(4.3.18)式得

$$\tilde{l}(\beta^{(1)}) = \{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\tilde{\xi}}_i(\beta^{(1)})\}^{\top}\{\frac{1}{n}\sum_{i=1}^{n}\hat{\tilde{\xi}}_i(\beta^{(1)})\hat{\tilde{\xi}}_i^{\top}(\beta^{(1)})\}^{-1}\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\tilde{\xi}}_i(\beta^{(1)})\} + o_p(1).$$

因此，由(4.3.14)和(4.3.15)，我们有$\tilde{l}(\beta^{(1)}) \xrightarrow{d} \chi_k^2$。综上所述，定理4.1.3证明完毕。 □

由于部分线性模型和单指标模型都是部分线性单指标模型的特殊情形，因此，我们可以像证明定理4.1.1-定理4.1.3一样证明定理4.2.1-定理4.2.5，因此省略。

§4.4 数值模拟

表 4.4.1: 根据经验似然方法和双重稳健有效方法(DRE)构造的参数分量β_1的置信区间的覆盖概率(CP)和区间长度(AL)的比较

(p,r)	n	EL		DRE	
		CP	AL	CP	AL
(10,10)	200	0.946	0.357	0.935	0.375
	400	0.953	0.303	0.941	0.347
	600	0.958	0.285	0.954	0.316
(20,10)	400	0.921	0.412	0.913	0.443
	600	0.935	0.396	0.926	0.417
	800	0.942	0.379	0.934	0.394
(30,10)	600	0.907	0.451	0.891	0.472
	800	0.914	0.423	0.912	0.448
	900	0.938	0.394	0.928	0.419

本节，我们通过数值模拟验证本章所给出的理论结果。整个模拟过程中，我们使用Epanechnikov核函数：$K(t) = \frac{3}{4}(1-t^2)_+$，并利用交叉验证方法选择最优窗宽$h_{opt}$。在异方差部分线性单指标模型(4.1.1)中，假设$(X_1,\cdots,X_p,Z_1,\cdots,Z_r)^\top$ 服从均值0协方差矩阵为$(\sigma_{ij})_{(p+r)\times(p+r)}$的正态分布，其中$\sigma_{ij}=0.5^{|i-j|}$；$Y$服从均值为$X^\top\beta+\exp(Z^\top\theta)$，方差为$|Z^\top\theta|$ 的正态分布。令：$\beta=(u,\cdots,u,1,0.5)^\top$，$u=(1,0.5,-1)$，$\theta=(4,1,-1,1,0,\cdots,0)^\top/4$。考虑维数$p=10,20,30$；$r=10$。

首先，我们考虑利用经验似然方法构造参数分量β_1的置信区间，其他参数分量的置信区间或置信域可以同样方法构造。同时，为了比较，我们也利用[102] 中双重稳健且有效方法构造β_1的置信区间。令名义置信度$1-\alpha$为0.95，如此模拟重复1000次，其结果总结在表4.4.1。

从表4.4.1可以看出，根据经验似然方法和双重稳健有效方法构造的参数分量β_1的置信区间的覆盖概率都随样本容量增加而增加，当样本容量达到一定程度时，覆盖概率开始接近名义置信水平。另外，基于经验似然方法的置信区间长度一致小于基于双重稳健有效方法的置信区间长度，而前者的置信度却高于后者。

在第二个模拟中，我们进一步考虑比较分别利用经验似然方法和双

表 4.4.2: 根据经验似然方法和双重稳健有效方法(DRE)构造的参数分量(β_1, β_2)的置信域的覆盖概率的比较

p	n	1-α=0.90		1-α=0.95	
		EL	DRE	EL	DRE
(10,10)	200	0.869	0.867	0.923	0.917
	400	0.885	0.879	0.936	0.924
	600	0.903	0.892	0.942	0.931
(20,10)	400	0.863	0.853	0.917	0.906
	600	0.876	0.862	0.925	0.908
	800	0.894	0.875	0.932	0.922
(30,10)	600	0.857	0.848	0.894	0.883
	800	0.874	0.856	0.908	0.901
	900	0.882	0.864	0.921	0.916

表 4.4.3: 根据经验似然方法和双重稳健有效方法(DRE)构造的参数$(\beta^\top, \theta^\top)^\top$的置信域的覆盖概率的比较

p	n	1-α=0.90		1-α=0.95	
		EL	DRE	EL	DRE
(10,10)	1500	0.856	0.851	0.913	0.909
	2000	0.862	0.857	0.926	0.915
	2500	0.883	0.874	0.932	0.927
(20,10)	2000	0.845	0.836	0.894	0.886
	2500	0.859	0.853	0.908	0.894
	3000	0.878	0.869	0.921	0.913
(30,10)	2500	0.847	0.835	0.889	0.872
	3000	0.852	0.848	0.905	0.898
	3500	0.873	0.862	0.917	0.904

重稳健方法构造参数分量(β_1, β_2) 的置信域的覆盖概率，其结果总结在表4.4.2。最后，我们还比较了分别利用经验似然方法和双重稳健方法构造参数$(\beta^\top, \theta^\top)^\top$的置信域的覆盖概率，其结果总结在表4.4.3。从表4.4.2和表4.4.3中可以看出，基于经验似然方法的置信域的置信度一致高于基于双重稳健有效方法的置信域的置信度。模拟结果进一步验证了本章提出的在高维情形下利用经验似然方法对异方差部分线性单指标模型进行推断的优点。

§4.5 实例分析

本节将利用异方差部分线性单指标模型分析ACTG175数据集(AIDS Clinical Trials Group Protocol 175)，Hammer et al.(1996)和Davidian et al. (2005)分别在[106]和[107]中分析过这些数据。CD4是一种辅助T细胞受体（TCR）与抗原呈递细胞的共受体，许多HIV临床试验都是分析指定时期治疗方法对CD4 细胞计数的影响。ACTG175数据集包括2139位HIV患者随机等比例分配到四种治疗方法(齐多夫定(ZDV)单药治疗，ZDV+去羟肌苷(DDI)治疗，ZDV+扎西他宾治疗和DDI治疗)对应的实验组接受治疗的结果。ACTG 175结果表明，抗逆转录病毒疗法可以减少中期HIV疾病且无症状患者的风险。

下面根据ACTG175数据集建立异方差部分线性单指标模型对HIV患者接受ZDV单药治疗结果进行分析。CD4 计数都在 96 ± 5 周之间进行(CD496)，数据集给出了532例患者200至500毫升剂量的CD4细胞计数。因为在 ACTG175 数据集中某些样本的CD496值缺失，只有321个样本 CD496 记录完整，所以我们只考虑这321个样本。设变量 CD496 为响应变量 Y，变量 age， wtkg， hemo， homo，drugs， karnof， race， gender， str2， symptom， CD40， CD420， CD80 和 CD820 为解释变量。进一步考虑将离散型变量(hemo, homo, drugs, karnof , race, gender, str2, symptom)指定为线性部分的解释变量，标准化的连续型(age,wtkg, cd80, cd820, cd40, cd420)指定为单指标部分的解释变量。将响应变量Y(CD496)标准化后建立如下异方差部分线性单指标模型

$$Y = X^\top \beta + g(Z^\top \theta) + \varepsilon,$$

其中：$X^\top = (\mathrm{hemo}, \mathrm{homo}, \mathrm{drugs}, \mathrm{karnof}, \mathrm{race}, \mathrm{gender}, \mathrm{str2}, \mathrm{symptom})$， $Z^\top = (\mathrm{age}, \mathrm{wtkg}, \mathrm{cd80}, \mathrm{cd820}, \mathrm{cd40}, \mathrm{cd420})$, $Y = CD496$，$g(\cdot)$为未知函数。从残差估

表 4.5.1: 参数$(\beta^\top,\theta^\top)^\top$的估计值以及根据经验似然和DRE方法构造的每个参数置信区间(置信水平为95%)

	$(\hat{\beta}^\top,\hat{\theta}^\top)^\top$	Confidence intervals	
		EL	DRE
hemo	0.0044	[-0.0507, 0.0932]	[-0.3218, 0.3305]
homo	0.0054	[-0.0526, 0.0103]	[-0.1784, 0.1892]
drugs	0.0010	[-0.0543, 0.0597]	[-0.2795, 0.2814]
karnof	0.0001	[-0.0482, 0.0504]	[-0.0136, 0.0138]
race	-0.0019	[-0.0685, 0.0306]	[-0.1890, 0.1851]
gender	-0.0039	[-0.0881, 0.0109]	[-0.2274, 0.2196]
str2	-0.0089	[-0.0138, -0.0405]	[-0.1697, 0.1520]
symptom	-0.0088	[-0.0137, -0.0394]	[-0.2290, 0.2115]
age	0.5687	[0.2687, 0.9687]	[-0.6460, 0.7597]
wtkg	-0.5925	[-0.9925, -0.2925]	[-1.1214, 1.0029]
cd80	0.0447	[-0.1553, 0.3947]	[-0.5243, 0.5332]
cd820	-0.0989	[-0.1489, 0.1511]	[-0.5509, 0.5311]
cd40	-0.4832	[-0.8332, -0.1832]	[-0.6623, 0.5657]
cd420	0.3701	[0.0701, 0.5701]	[-0.5681, 0.6421]

计的散点图判断ε为异方差。使用本章给出的方法，我们得到了每个变量系数的估计值以及相应的置信区间，其结果总结在表4.5.1。从表4.5.1发现，基于经验似然方法的置信区间长度一致小于基于双重稳健有效方法的置信区间长度。

第五章 高维情形下两样本问题的经验似然推断

本章讨论两样本两样本问题的经验似然推断。当$n \to \infty$，$p \to \infty$时，首先，我们构造两样本均值之差的经验似然比统计量，在一定条件下证明了经验似然比统计量的极限分布为渐近正态分布，并构造了均值之差的经验似然置信域；其次，我们利用经验似然方法对两样本线性模型系数之差进行统计推断，得到了其经验似然比统计量的渐近分布，并在一定条件下证明了根据经验似然方法得到的估计量渐近有效。

本章内容结构如下：第一节介绍两样本均值经验似然推断方法及其主要结论；第二节利用经验似然方法对两样本线性模型系数之差进行统计推断；第三节为主要定理的证明；第四和第五节分别为数值模拟和实例分析。

§5.1 两样本均值的经验似然推断

考虑独立同分布样本：$\{X_{i1}, X_{i2}, \ldots, X_{in_i}\} \sim F_i$，$i = 1, 2$，其中：$X_{ij} \in \mathbb{R}^p$，$F_i \in \mathbb{R}^p$是均值为$\mu_i$协方差矩阵为$\Sigma_i$的分布函数。在高维数据分析中，一个常见问题是检验两个高维样本均值是否相等，即

$$H_0 : \delta = \mu_1 - \mu_2 = \delta_0 \ , \quad H_1 : \delta = \mu_1 - \mu_2 \neq \delta_0. \tag{5.1.1}$$

当样本维数p固定时，Liu et al.(2008)[48]讨论了两样本均值之差的经验似然推断，并在一定条件下证明了两样本均值之差的经验似然比的极限分布为χ^2_p。本节我们将Liu et al.(2008)[48]的结果推广到维数发散情形，即，当$n \to \infty$时，$p \to \infty$。

假设有如下多元模型：

$$X_{ij} = \Gamma_i Z_{ij} + \mu_i, \quad j = 1, 2, \ldots, n_i, \quad i = 1, 2, \tag{5.1.2}$$

其中：Γ_i为$p \times m$阶矩阵，$m \geq p$，$\Gamma_i \Gamma_i^\top = \Sigma_i$，$\{Z_{ij}\}_{j=i}^{n_i}$为$m-$维独立同分布随机向量。令$\gamma_1(\Sigma_i) \leq \ldots \leq \gamma_p(\Sigma_i)$为矩阵$\Sigma_i$的特征值。我们给出如下假设：

假设1. $E(Z_{ij}) = 0$，$\mathrm{Var}(Z_{ij}) = I_m$，$E(Z_{ijl}^{4k}) = m_{i4k} \in (0, \infty)$，$E(Z_{ijl_1}^{\alpha_1} \ldots Z_{ijl_q}^{\alpha_q}) = E(Z_{ijl_1}^{\alpha_1}) \ldots E(Z_{ijl_q}^{\alpha_q})$，$\sum_{l=1}^{q} \alpha_l \leq 4k$和$l_1 \neq l_2 \neq \ldots \neq l_q$。这里$k$为正整数，$I_m$为$m-$阶单位矩阵；

假设2. 存在常数$C_{2i} > C_{1i} > 0$，$i = 1, 2$，Σ_i的特征值满足$C_{1_i} \le \gamma_1(\Sigma_i) \le \ldots \le \gamma_p(\Sigma_i) \le C_{2_i}$；

假设3. 当$n \to \infty$时，$p \to \infty$，$pn^{-(1/5)} \to 0$。

令$Y_j = X_j + \delta_0$，$n = n_1 + n_2$，当$n \to \infty$ 时，$\theta = n_1/(n_1 + n_2) \to \theta_0 \in (0, 1)$。关于$\delta$的经验似然函数可以表示为

$$L(\delta_0) = \sup\{\prod_{i=1}^{n_1} p_i \prod_{j=n_1+1}^{n} q_j | \sum_{i=1}^{n_1} p_i(X_i - \mu_1) = \sum_{j=n_1+1}^{n} q_j(Y_j - \mu_1) = 0\}, \quad (5.1.3)$$

其中，$(p_1, p_2, \ldots, p_{n_1})$和$(q_1, q_2, \ldots, q_{n_2})$为概率向量，满足

$$\sum_{i=1}^{n_1} p_i = \sum_{j=1}^{n_2} q_j = 1, \quad p_i \ge 0, q_j \ge 0.$$

由Lagrange乘数法，有

$$p_i = \frac{1}{n\theta} \frac{1}{1 + \theta^{-1}\lambda_1^\top (X_i - \mu_1)}, \qquad q_j = \frac{1}{n(1-\theta)} \frac{1}{1 - (1-\theta)^{-1}\lambda_2^\top (Y_j - \mu_1)}.$$

类似于Jing (1995)[47]和Liu et al.(2008)[48]，为了计算方便，对Lagrange乘数λ_1和λ_2增加如下约束：

$$(\lambda_1, \lambda_2) \in \Omega, \quad \Omega = \{(\lambda_1, \lambda_2) : \lambda_1 = -\lambda_2 = \lambda\}.$$

关于δ的经验似然比为

$$L(\lambda, \mu_1) = 2\{\sum_i \log(1 + \theta^{-1}\lambda^\top (X_i - \mu_1)) - \sum_j \log(1 + (1-\theta)^{-1}\lambda^\top (Y_j - \mu_1))\}. \quad (5.1.4)$$

(5.1.4)式分别对λ_1和λ_2求偏导数并令其为零得

$$\begin{cases} \frac{1}{\theta}\sum_i \frac{X_i - \mu_1}{1+\theta^{-1}\lambda^\top (X_i - \mu_1)} - \frac{1}{1-\theta}\sum_j \frac{Y_j - \mu_1}{1+(1-\theta)^{-1}\lambda^\top (Y_j - \mu_1)} = 0, \\ \frac{1}{\theta}\sum_i \frac{\lambda}{1+\theta^{-1}\lambda^\top (X_i - \mu_1)} - \frac{1}{1-\theta}\sum_j \frac{\lambda}{1+(1-\theta)^{-1}\lambda^\top (Y_j - \mu_1)} = 0. \end{cases} \quad (5.1.5)$$

令$(\hat{\lambda}, \hat{\mu}_1)$为方程组(5.1.5)的解，有$l(\delta_0) = L(\hat{\lambda}, \hat{\mu}_1)$。

下面定理给出了维数发散情形下两样本均值之差的经验似然比的渐近分布。

定理 5.1.1 假设1($k = 1$)，假设2-3以及原假设成立条件下，当$n \to \infty$时，有

$$(2p)^{-\frac{1}{2}}\{l(\delta_0) - p\} \xrightarrow{d} N(0, 1). \quad (5.1.6)$$

定理5.1.1将Liu et al.(2008)[48]的结果推广到维数发散情形。特别地，定理5.1.1可用来构造δ的经验似然置信域。δ的渐近置信水平为$1-\alpha$的渐近经验似然置信域$I_\alpha(\delta)$为

$$I_\alpha(\delta)=\{\delta:\ l(\delta)\le p+z_\alpha\sqrt{2p}\}.$$

§5.2 两样本线性模型系数之差的经验似然推断

考虑如下线性模型：

$$Y_k=\begin{cases} X_k^\top\beta+\varepsilon_k, 1\le k\le n_1, \\ X_k^\top\beta_1+\varepsilon_k, n_1+1\le k\le n, \end{cases} \tag{5.2.1}$$

其中：$X_k\in\mathbb{R}^p$为独立同分随机向量，$Y_k\in\mathbb{R}$为相应的响应变量，β和β_1为$p-$维系数。不失一般性，我们假设$E(X_k)=0$，$\mathrm{Cov}(X_k)=\Sigma$，误差$\varepsilon_k\in\mathbb{R}$为相互独立的随机变量，$\{\varepsilon_i, 1\le i\le n_1\}$服从均值为零方差为$\sigma_1^2$的分布$F$，$\{\varepsilon_j, n_1+1\le j\le n\}$服从均值为零方差为$\sigma_2^2$的分布$G$。令$\delta=\beta-\beta_1$，考虑如下检验假设

$$H_0:\delta=\delta_0\ ,\qquad H_1:\delta\ne\delta_0.$$

当p固定且$X_k(1\le k\le n)$为确定变量时，在原假设成立条件下，Zi et al.(2012)[50]证明了关于δ 的经验似然比统计量的极限分布为自由度为p的渐近卡方分布。本节在维数发散情形下对δ进行经验似然推断。在模型(5.2.1)中，假设β_0为原假设成立下的真实参数。令

$$Y_j^*=Y_j+X_j^\top\delta_0,$$

$$Z_i(\beta)=X_i(Y_i-X_i^\top\beta),$$

$$Z_j(\beta)=X_j(Y_j^*-X_j^\top\beta),$$

其中$i=1,2,\ldots,n_1$，$j=n_1+1,n_1+2,\ldots,n$。

关于两样本线性模型系数之差δ的经验似然函数为

$$\tilde{L}(\delta_0)=\sup\{\prod_{i=1}^{n_1}p_i\prod_{j=n_1+1}^{n}q_j|\sum_{i=1}^{n_1}p_iZ_i=\sum_{j=n_1+1}^{n}q_jZ_j=0\}. \tag{5.2.2}$$

由Lagrange乘数法，有

$$p_i = \frac{1}{n\theta}\frac{1}{1+\theta^{-1}\lambda_1^\top Z_i(\beta)}, \quad q_j = \frac{1}{n(1-\theta)}\frac{1}{1-(1-\theta)^{-1}\lambda_2^\top Z_j(\beta)}.$$

为了计算方便，类似于本章第一节对Lagrange乘数λ_1和λ_2施加如下约束：

$$(\lambda_1, \lambda_2) \in \Omega, \quad \Omega = \{(\lambda_1, \lambda_2) : \lambda_1 = -\lambda_2 = \lambda\}.$$

关于δ的经验似然比可以表示为

$$\tilde{l}(\lambda, \beta) = 2\{\sum_i \log(1+\theta^{-1}\lambda^\top Z_i(\beta)) - \sum_j \log(1+(1-\theta)^{-1}\lambda^\top Z_j(\beta))\}. \quad (5.2.3)$$

使(5.2.3)式达到最小的$\hat{\beta}$为参数β的经验似然估计。由Lagrange乘数法知，求使(5.2.3)式达到最小的$(\hat{\lambda}, \hat{\beta})$等价于求解如下方程组

$$\begin{cases} \frac{1}{\theta}\sum_i \frac{Z_i(\beta)}{1+\theta^{-1}\lambda^\top Z_i(\beta)} - \frac{1}{1-\theta}\sum_j \frac{Z_j(\beta)}{1+(1-\theta)^{-1}\lambda^\top Z_j(\beta)} = 0, \\ \frac{1}{\theta}\sum_i \frac{X_i X_i^\top \lambda}{1+\theta^{-1}\lambda^\top Z_i(\beta)} - \frac{1}{1-\theta}\sum_j \frac{X_j X_j^\top \lambda}{1+(1-\theta)^{-1}\lambda^\top Z_j(\beta)} = 0. \end{cases} \quad (5.2.4)$$

因此，我们有$\tilde{L}(\delta_0) = \tilde{l}(\hat{\lambda}, \hat{\beta})$。

在模型(5.2.1)中，当X_k为随机变量时，类似于本章第二节假定X_k满足模型(5.1.2)以及相应的正则性条件。另外，假设随机误差满足：

假设4. 随机误差$\{\varepsilon_i, 1 \le i \le n_1\}$和$\{\varepsilon_j, n_1+1 \le j \le n\}$为相互独立随机变量，均值为零，$E(\varepsilon_i^{4k}) < \infty$，$E(\varepsilon_j^{4k}) < \infty$，其中$k$为正整数。

下面定理给出了两样本线性模型系数之差的经验似然比的渐近分布。

定理 5.2.1 假设1($k = 1$)，假设2-4以及原假设成立条件下，当$n \to \infty$时，有

$$(2p)^{-\frac{1}{2}}\{\tilde{L}(\delta_0) - p\} \xrightarrow{d} N(0, 1). \quad (5.2.5)$$

定理5.2.1可用来检验两样本线性模型系数是否相等的假设，同时，也可以用来构造δ的经验似然置信域。δ的渐近置信水平为$1-\alpha$的渐近经验似然置信域$\tilde{I}_\alpha(\delta)$为

$$\tilde{I}_\alpha(\delta) = \{\delta : \ \tilde{L}(\delta) \le p + z_\alpha\sqrt{2p}\}.$$

评论 5.2.1 定理5.2.1还可以用来检验高维线性模型的稀疏性假设和对高维线性模型进行变量选择。例如，在模型(5.2.1)中令 $\beta = (\beta_{01}, \beta_{02})$ 和

$\beta_1 = (\beta_{01}, 0)$，其中$\beta_{01} \in \mathbb{R}^q$为$q-$维向量，$\beta_{02} \in \mathbb{R}^{(p-q)}$为$(p-q)-$维向量。检验假设

$$H_0 : \beta_{02} = 0 \ , \qquad H_1 : \beta_{02} \neq 0$$

等价于检验假设

$$H_0 : \delta = \beta - \beta_1 = 0 \ , \qquad H_1 : \delta = \beta - \beta_1 \neq 0.$$

下面定理给出经验似然估计$\hat{\beta}$的渐近性质。

定理 5.2.2 在定理5.2.1相同假设条件下，当$n \to \infty$时，有

$$\sqrt{n} W_n \Phi^{-\frac{1}{2}} (\hat{\beta} - \beta_0) \xrightarrow{d} N(0, G),$$

其中$W_n \in \mathbb{R}^{q\times p}$，$W_n W_n^\top \to G$，$G \in \mathbb{R}^{q\times q}$，$q$固定，$\Phi = (\sigma_1^2 + \sigma_2^2) B^{-1} \Sigma (B^{-1})^\top$，$B = \frac{1}{1-\theta} S_{2n} - \frac{1}{\theta} S_{1n}$，$S_{1n} = \frac{1}{n\theta} \sum_{i=1}^{n_1} X_i X_i^\top$，$S_{2n} = \frac{1}{n(1-\theta)} \sum_{j=n_1+1}^{n} X_j X_j^\top$。

在定理5.2.2中，$W_n(\hat{\beta} - \beta_0)$表示维数发散的向量$W_n(\hat{\beta} - \beta_0)$在维数固定的$q$维空间上的投影，其投影向量的极限分布为多元正态分布。

§5.3 定理的证明

证明定理之前，我们先给出如下引理。

引理 5.3.1 在假设1($k = 1$)和假设2-3成立条件下，有$\|\hat{\mu}_1 - \mu_0\| = O_p\{(p/n)^{1/2}\}$，$\|\hat{\lambda}\| = O_p\{(p/n)^{1/2}\}$，其中$\|\cdot\|$表示欧几里得范数。

证明： 本引理证明类似于Chen et al.(2009)[56]中定理1的证明，因此省略。 □

引理 5.3.2 令$D_n = \{\beta : \|\hat{\beta} - \beta_0\| \leq C\frac{p}{n}\}$，其中常数$C > 0$。在假设1($k = 1$)和假设2-3成立条件下，有

(1) $\|\lambda(\beta)\|^2 = O_p(p/n)$，$\beta \in D_n$；

(2) 以概率1，$\tilde{l}(\lambda, \beta)$在$D_n$内最小值存在。

证明： 本引理证明类似于Tang和Leng(2010)[83]中引理3和引理4的证明，因此省略。 □

定理5.1.1的证明： 本引理证明类似于Chen et al.(2009)[56]中定理2的证明，因此省略。 □

定理5.2.1的证明： 令 $U_i=\lambda^\top Z_i(\beta)$，$V_j=\lambda^\top Z_j(\beta)$，$i=1,2,\ldots,n_1$，$j=n_1+1,n_1+2,\ldots,n$。由引理5.3.2，有

$$\max_{i=1,2,\ldots,n_1}|U_i|\leq\|\lambda^\top\|\max_{i=1,2,\ldots,n_1}\|Z_i(\beta)\|=O_p\{(\frac{p}{n})^{\frac{1}{2}}\}O_p\{p\}=o_p(1),$$

和

$$\max_{j=n_1+1,n_1+2,\ldots,n}|V_j|\leq\|\lambda^\top\|\max_{j=n_1+1,n_1+2,\ldots,n}\|Z_j(\beta)\|=O_p\{(\frac{p}{n})^{\frac{1}{2}}\}O_p\{p\}=o_p(1).$$

将(5.2.4)式第一个方程左边展开得

$$0=g(\lambda)=\frac{1}{n\theta}\sum_{i=1}^{n_1}Z_i(\beta)-\frac{1}{n(1-\theta)}\sum_{j=n_1+1}^{n}Z_j(\beta)-(S_1+S_2)\lambda+\beta_{1n}+\beta_{2n}.$$

令

$$\bar{Z}_1(\beta)=\frac{1}{n\theta}\sum_{i=1}^{n_1}Z_i(\beta),\qquad \bar{Z}_2(\beta)=\frac{1}{n(1-\theta)}\sum_{j=n_1+1}^{n}Z_j(\beta),$$

$$S_1=\frac{1}{n\theta}\sum_{i=1}^{n_1}Z_i(\beta)Z_i(\beta)^\top,\quad S_2=\frac{1}{n(1-\theta)}\sum_{j=n_1+1}^{n}Z_j(\beta)Z_j(\beta)^\top,$$

$$\beta_{1n}=\frac{1}{n\theta}\sum_{i=1}^{n_1}Z_i(\beta)\frac{U_i^2}{(1+\xi_i)},\quad |\xi_i|\leq|\lambda^\top Z_i(\beta)|,$$

$$\beta_{2n}=\frac{1}{n(1-\theta)}\sum_{j=n_1+1}^{n}Z_j(\beta)\frac{V_j^2}{(1+\xi_j)},\quad |\xi_j|\leq|\lambda^\top Z_j(\beta)|.$$

因此

$$g(\lambda)=\bar{Z}_1(\beta)+\bar{Z}_2(\beta)-(S_1+S_2)\lambda+\beta_{1n}+\beta_{2n}. \tag{5.3.1}$$

由

$$\max_{i=1,2,\ldots,n_1}|U_i|=o_p(1),\quad \max_{j=n_1+1,n_1+2,\ldots,n}|V_j|=o_p(1),$$

得

$$\max_{i=1,2,\ldots,n_1}|\xi_i|=0_p(1),\quad \max_{j=n_1+1,n_1+2,\ldots,n}|\xi_j|=0_p(1).$$

根据引理5.3.2，我们有

$$\beta_{1n}=\max_{i=1,2,\ldots,n_1}|U_i|\|\lambda\|^2\{1+o_p(1)\}=o_p(\|\lambda\|^2)$$

和

$$\beta_{2n}=\max_{j=n_1+1,n_1+2,\ldots,n}|V_j|\|\lambda\|^2\{1+o_p(1)\}=o_p(\|\lambda\|^2).$$

由(5.3.1)式可得

$$\lambda=(S_1+S_2)^{-1}(\bar{Z}_1(\beta)-\bar{Z}_2(\beta))+(S_1+S_2)^{-1}(\beta_{1n}+\beta_{2n}). \tag{5.3.2}$$

利用Taylor公式，有

$$\log(1+U_i)=U_i+\frac{U_i^2}{2}+\frac{U_i^3}{3(1+\xi_i)^3},\quad |\xi_i|\le|U_i|,$$

和

$$\log(1+V_j)=V_j+\frac{V_j^2}{2}+\frac{V_j^3}{3(1+\xi_j)^3},\quad |\xi_j|\le|V_j|.$$

因此，我们有

$$\tilde{l}(\lambda,\beta)=R_1+R_2+R_3,$$

其中

$$\begin{aligned}
R_1&=n(\bar{Z}_1(\beta)-\bar{Z}_2(\beta))^\top(S_1+S_2)^{-1}(\bar{Z}_1(\beta)-\bar{Z}_2(\beta)),\\
R_2&=n(\beta_{1n}+\beta_{2n})^\top(S_1+S_2)^{-1}(\beta_{1n}+\beta_{2n}),\\
R_3&=\frac{2}{3}\sum_{i=1}^{n_1}\{\lambda^\top Z_i(\beta)\}^3\{1+o_p(1)\}+\frac{2}{3}\sum_{j=n_1+1}^{n}\{\lambda^\top Z_j(\beta)\}^3\{1+o_p(1)\}.
\end{aligned}$$

由引理5.3.1和假设3，有

$$\begin{aligned}
n(\beta_{1n})^\top(S_1+S_2)^{-1}(\bar{Z}_1(\beta_{1n})&=O_p(n\|\beta_{1n}\|^2)=o_p\{n(\frac{p}{n})^2\}=o_p(1),\\
n(\beta_{2n})^\top(S_1+S_2)^{-1}(\bar{Z}_1(\beta_{2n})&=O_p(n\|\beta_{2n}\|^2)=o_p\{n(\frac{p}{n})^2\}=o_p(1),\\
n(\beta_{1n})^\top(S_1+S_2)^{-1}(\bar{Z}_1(\beta_{2n})&=n(\beta_{2n})^\top(S_1+S_2)^{-1}(\bar{Z}_1(\beta_{1n})\\
&=O_p(n\|\beta_{1n}\|\|\beta_{2n}\|)=o_p\{n(\frac{p}{n})^2\}=o_p(1),\\
\sum_{i=1}^{n_1}\{\lambda^\top Z_i(\beta)\}^3&=O_p\{n\theta(\frac{p}{n})^{\frac{3}{2}}p\}=o_p(p^{\frac{1}{2}}),\\
\sum_{j=n_1+1}^{n}\{\lambda^\top Z_j(\beta)\}^3&=O_p\{n(1-\theta)(\frac{p}{n})^{\frac{3}{2}}p\}=o_p(p^{\frac{1}{2}}).
\end{aligned}$$

类似于Chen et al.(2009)[56]中引理5和引理6的证明，容易得出

$$(2p)^{-\frac{1}{2}}\{R_1 - p\} \xrightarrow{d} N(0,1)$$

和

$$\tilde{l}(\delta_0) = \tilde{l}(\lambda, \beta) = R_1 + o_p(p^{\frac{1}{2}}).$$

综上可得

$$(2p)^{-\frac{1}{2}}\{\tilde{l}(\delta_0) - p\} = (2p)^{-\frac{1}{2}}\{R_1 - p\} + o_p(1) \xrightarrow{d} N(0,1).$$

定理5.2.1证毕。 □

定理5.2.2的证明： 由条件极值方法知，$(\hat{\beta}, \hat{\lambda})$满足

$$\begin{cases} Q_{1n}(\beta, \lambda) = 0, \\ Q_{2n}(\beta, \lambda) = 0, \end{cases}$$

其中

$$Q_{1n}(\beta, \lambda) = \frac{1}{n}\sum_{i=1}^{n_1} \frac{Z_i(\beta)}{\theta + \lambda^\top Z_i(\beta)} - \frac{1}{n}\sum_{j=n_1+1}^{n} \frac{Z_j(\beta)}{1-\theta+\lambda^\top Z_j(\beta)}, \tag{5.3.3}$$

$$Q_{2n}(\beta, \lambda) = -\frac{1}{n\theta}\sum_{i=1}^{n_1} \frac{X_i X_i^\top \lambda}{\theta + \lambda^\top Z_i(\beta)} + \frac{1}{n(1-\theta)}\sum_{j=n_1+1}^{n} \frac{X_j X_j^\top \lambda}{1-\theta+\lambda^\top Z_j(\beta)}. \tag{5.3.4}$$

令

$$\hat{\psi} = (\hat{\beta}, \hat{\lambda})^\top, \quad \psi_0 = (\beta_0, 0)^\top.$$

将方程(5.3.3)和(5.3.4)在$(\beta_0, 0)^\top$处展开得

$$0 = Q_{kn}(\beta, \lambda) = Q_{kn}(\beta_0, 0) + \frac{\partial Q_{kn}(\beta_0, 0)}{\partial \beta}(\hat{\beta} - \beta_0) + \frac{\partial Q_{kn}(\beta_0, 0)}{\partial \lambda}\hat{\lambda} + R_{kn}, \tag{5.3.5}$$

其中，$k = 1, 2$，余项R_{1n}和R_{2n}都是$p-$维向量，第$d-$个分量为

$$R_{jn,d} = \frac{1}{2}(\hat{\psi} - \psi_0)^\top \nabla^2\{Q_{jn,d}(\psi^*)\}(\hat{\psi} - \psi_0),$$

算子∇^2 表示ψ求二阶偏导数，ψ^*满足$\|\psi^* - \psi_0\| \leq \|\hat{\psi} - \psi_0\|$。

令

$$S_{1n} = \frac{1}{n\theta}\sum_{i=1}^{n_1} X_i X_i^\top,$$

$$S_{2n} = \frac{1}{n(1-\theta)} \sum_{i=1}^{n_1} X_j X_j^{\top}.$$

由(5.3.3)和(5.3.4)式，有

$$Q_{1n}(\beta_0, 0) = \bar{Z}_1(\beta) - \bar{Z}_2(\beta),$$

$$Q_{2n}(\beta_0, 0) = 0,$$
$$\frac{\partial Q_{1n}(\beta_0, 0)}{\partial \lambda} = \frac{-1}{\theta} S_1 + \frac{1}{1-\theta} S_2,$$
$$\frac{\partial Q_{1n}(\beta_0, 0)}{\partial \beta} = \frac{\partial Q_{2n}(\beta_0, 0)}{\partial \lambda} = \frac{-1}{\theta} S_{1n} + \frac{1}{1-\theta} S_{2n},$$
$$\frac{\partial Q_{2n}(\beta_0, 0)}{\partial \beta} = 0,$$

其中$\bar{Z}_1(\beta)$，$\bar{Z}_2(\beta)$，S_1和S_2 表达式见定理5.2.1证明过程。令

$$W = \begin{pmatrix} \frac{\partial Q_{1n}(\beta_0,0)}{\lambda} & \frac{\partial Q_{1n}(\beta_0,0)}{\beta} \\ \frac{\partial Q_{2n}(\beta_0,0)}{\lambda} & \frac{\partial Q_{2n}(\beta_0,0)}{\beta} \end{pmatrix},$$

容易得出

$$\begin{pmatrix} Q_{1n}(\beta_0, 0) \\ Q_{2n}(\beta_0, 0) \end{pmatrix} = W \begin{pmatrix} \hat{\lambda} \\ \hat{\beta} - \beta_0 \end{pmatrix} + \begin{pmatrix} R_{1n} \\ R_{2n} \end{pmatrix}. \tag{5.3.6}$$

(5.3.6)式两边都左乘W的逆，整理得

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\beta} - \beta_0 \end{pmatrix} = W^{-1} \begin{pmatrix} Q_{1n}(\beta_0, 0) \\ 0 \end{pmatrix} + W^{-1} \begin{pmatrix} R_{1n} \\ R_{2n} \end{pmatrix},$$

利用分块矩阵求逆的方法得W^{-1}和B^{-1}

$$W^{-1} = \begin{pmatrix} 0 & B^{-1} \\ B^{-1} & -B^{-1}AB^{-1} \end{pmatrix},$$

其中

$$A = \frac{1}{1-\theta} S_2 - \frac{-1}{\theta} S_1, B = \frac{1}{1-\theta} S_{2n} - \frac{-1}{\theta} S_{1n},$$
$$B^{-1} = (1-\theta) S_{2n}^{-1} (I - ((1-\theta) S_{2n}^{-1} - \theta S_{1n}^{-1})^{-1} (1-\theta) S_{2n}^{-1}).$$

那么

$$\hat{\beta} - \beta_0 = B^{-1}(\bar{Z}_1(\beta) - \bar{Z}_2(\beta)) + B^{-1} R_{1n} - B^{-1} A B^{-1} R_{2n}. \tag{5.3.7}$$

令$W_n \in \mathbb{R}^{q\times p}$，$W_n W_n^\top \to G$，$G$为$q\times q$阶矩阵，$q$固定，

$$\Phi_n = (\sigma_1^2+\sigma_2^2)B^{-1}\Sigma(B^{-1})^\top,$$

$$T_{n_i} = n^{-\frac{1}{2}}\xi_{n_i},\quad i=1,2,\ldots,n_1,$$

$$\xi_{n_i} = W_n\Phi^{-\frac{1}{2}}B^{-1}Z_i(\beta),$$

$$T_{n_j} = n^{-\frac{1}{2}}\xi_{n_j},\quad j=n_1+1,n_2+2,\ldots,n,$$

$$\xi_{n_j} = W_n\Phi^{-\frac{1}{2}}B^{-1}Z_j(\beta).$$

显然

$$E\|T_{n_i}\|^4 = n^{-2}E(\xi_{n_i}^\top\xi_{n_i}) = O_p(\frac{p^2}{n^2}),$$

$$E\|T_{n_j}\|^4 = n^{-2}E(\xi_{n_j}^\top\xi_{n_j}) = O_p(\frac{p^2}{n^2}),$$

$$Pr(\|T_{n_i}\|\geq\varepsilon) \leq n^{-1}\varepsilon^{-2}E\|\xi_{n_i}\|^2 = O_p(\frac{p^2}{n^2}),$$

$$Pr(\|T_{n_j}\|\geq\varepsilon) \leq n^{-1}\varepsilon^{-2}E\|\xi_{n_j}\|^2 = O_p(\frac{p^2}{n^2}).$$

因此，我们有

$$\begin{aligned}\sum_{i=1}^{n_1}E\|T_{n_i}\|^2I(\|T_{n_i}\|^2>\varepsilon) &= n\theta E\|T_{n_1}\|^2I(\|T_{n_i}\|>\varepsilon)\\ &\leq n\theta\{E\|T_{n_1}\|^4\}^{\frac{1}{2}}\}\{Pr(\|T_{n_i}\|>\varepsilon)\}^{\frac{1}{2}}\to 0,\\ \sum_{j=n_1+1}^{n}E\|T_{n_j}\|^2I(\|T_{n_j}\|^2>\varepsilon) &= n\theta E\|T_{n_n}\|^2I(\|T_{n_j}\|>\varepsilon)\\ &\leq n\theta\{E\|T_{n_n}\|^4\}^{\frac{1}{2}}\}\{Pr(\|T_{n_j}\|>\varepsilon)\}^{\frac{1}{2}}\to 0,\end{aligned}$$

和

$$\begin{aligned}\sum_{k=1}^{n}E\|T_{n_k}\|^2I(\|T_{n_k}\|^2>\varepsilon) &= \sum_{i=1}^{n_1}E\|T_{n_i}\|^2I(\|T_{n_i}\|^2>\varepsilon)\\ &\quad+\sum_{j=n_1+1}^{n}E\|T_{n_j}\|^2I(\|T_{n_j}\|^2>\varepsilon)\to 0.\end{aligned}$$

由Lindeberg-Feller中心极限定理可得

$$\sqrt{n}W_n\Phi^{-\frac{1}{2}}B^{-1}(\bar{Z}_1(\beta)-\bar{Z}_2(\beta)) \xrightarrow{d} N(0,G). \tag{5.3.8}$$

根据引理5.3.2可以推出

$$\|\hat{\psi}-\psi_0\|^2 = O_p(\frac{p}{n}),$$

和

$$\|R_{kn}\| \leq \frac{\|\hat{\psi}-\psi_0\|^4}{n^{-2}} \sum_{i,j,d=1}^{2p} \frac{n^2\partial^2 Q_{kn,d}(\psi^*)}{\partial\psi_i\partial\psi_j} = O_p(\frac{p^5}{n^2}), \quad k=1,2.$$

所以，有

$$\|\sqrt{n}W_n\Phi_n^{-\frac{1}{2}}B^{-1}R_{1n}\|^2 = O_p(\frac{(n\theta)p^5}{n^2}) = o_p(1), \tag{5.3.9}$$

和

$$\|\sqrt{n}W_n\Phi_n^{-\frac{1}{2}}B^{-1}R_{2n}\|^2 = O_p(\frac{(n(1-\theta))p^5}{n^2}) = o_p(1). \tag{5.3.10}$$

联合(5.3.8)-(5.3.10)得

$$\sqrt{n}W_n\Phi_n^{-\frac{1}{2}}(\hat{\beta}-\beta_0) \xrightarrow{d} N(0,G).$$

综上所述，定理5.2.2证明完毕。 □

§5.4 数值模拟

首先，我们对两样本均值之差的经验似然推断方法进行模拟研究。令$n_1=150$，$n_2=100$，$p=40$。样本观测值$\{X_{ijk}\}$由如下移动平均模型产生，

$$X_{ijk} = U_{ijk} + \rho U_{i,(j+1),k} + \mu_{jk}, (j=1,2\ldots,p, i=1,2,\ldots,n_k, k=1,2), \tag{5.4.1}$$

其中ρ和μ_{jk}为常数，U_{ijk}是独立并服从形状参数为4尺度参数为1的$\Gamma-$分布的随机变量。在模型(5.4.1)中，$\Sigma=(\sigma_{ij})$，$\sigma_{ii}=4(1+\rho^2)$，$\sigma_{i,(i+1)}=4\rho$，$\sigma_{ij}=0$，$|i-j|>1$。检验显著性水平为$\alpha=0.05$，$\rho=0$，0.5，1。在模拟中，我们比较了Bai和Arandasa(1996)[108]中的Hotelling检验、BS检验和经验似然比检验的势函数。所有模拟重复1000次，其结果总结在表(5.4.1)。

表 5.4.1: Hotelling检验、BS检验和经验似然比检验的势函数比较($\eta = \|\mu_1 - \mu_2\|^2/p$,$n_1 = 150, n_2 = 100, p = 40, \alpha = 0.05$)

	$\rho = 0$			$\rho = 0.5$			$\rho = 1$		
η	H	BS	ELR	H	BS	ELR	H	BS	ELR
0.00	0.035	0.048	0.063	0.039	0.062	0.054	0.038	0.045	0.063
0.10	0.124	0.112	0.360	0.064	0.106	0.376	0.090	0.081	0.240
0.15	0.331	0.357	0.624	0.225	0.389	0.385	0.152	0.204	0.314
0.20	0.617	0.671	0.862	0.324	0.512	0.587	0.208	0.383	0.591
0.25	0.896	0.919	0.935	0.561	0.790	0.852	0.389	0.658	0.659
0.30	0.989	0.998	0.994	0.842	0.953	0.941	0.607	0.820	0.823
0.35	0.999	1.000	1.000	0.928	0.985	0.983	0.691	0.945	0.935
0.40	1.000	1.000	1.000	0.989	0.997	1.000	0.925	0.989	0.979
0.45	1.000	1.000	1.000	0.999	1.000	1.000	0.963	1.000	1.000
0.50	1.000	1.000	1.000	1.000	1.000	1.000	0.992	1.000	1.000

从表5.4.1可以看出，当原假设成立时($\delta = \mu_1 - \mu_2 = \delta_0$)，三个检验方法拒绝原假设的频率接近名义显著性水平；当备择假设成立时BS检验和经验似然比检验的势函数一致优于Hotelling检验。δ与δ_0的差越大，拒绝原假设的频率也越大。当$\eta = \|\mu_1 - \mu_2\|^2/p$接近0.4时，检验方法拒绝原假设的频率几乎接近于1。从理论上讲，BS检验和Hotelling检验都需要假设两样本有相同协方差矩阵，但本章提出的经验似然比方法不需要这个假设。因此模拟结果进一步验证了本章提出的方法的优点。

在第二个模拟中，我们继续考虑模型(5.2.1)，令$X_i \in \mathbb{R}^p$和$X_j \in \mathbb{R}^p$为服从正态分布的随机变量，其每个分量的方差为1，当$|i-j| = 1$时，变量的第i和第j个分量的相关系数$\rho(i,j) = \rho = 0.5$，当$|i-j| > 1$时，$\rho(i,j) = 0$。真实参数$\beta = (1, 0.5, 0.5, 1, 1, 0.5, 1.5, 1.5, 2, 0.5)$，$\beta_1 = (2, 1, 1, 0.5, 2, 1, 0.5, 1, 0.5, 1)$。误差变量$\varepsilon_i$有如下两种情形：(1) $\varepsilon_i = N(0,1)$， $\varepsilon_j = N(0,1)$； (2) $\varepsilon_i = N(0,1)$， $\varepsilon_j = \chi_2^2 - 2$。对于$n_1 = 100$，$n_2 = 150$，或 $n_1 = 200$，$n_2 = 250$，通过以上方式得随机样本$\{X_i, Y_i\}_{i=1}^{n_1}$和$\{X_j, Y_j\}_{j=n_1+1}^{n}$。我们模拟基于经验似然方法的假设检验($H_0 : \delta = \mu_1 - \mu_2 = \delta_0$, $H_1 : \delta = \mu_1 - \mu_2 \neq \delta_0$)的势函数，检验显著性水平为$\alpha = 0.05$，$\rho = 0$，0.5，如此模拟重复1000次，其结果见

表 5.4.2: 经验似然比检验的势函数($\eta = \|\delta - \delta_1\|/\|\delta_1\|$)

	p=10				p=15			
	$n_1 = 100$ $n_2 = 150$		$n_1 = 200$ $n_2 = 250$		$n_1 = 100$ $n_2 = 150$		$n_1 = 200$ $n_2 = 250$	
η	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
0.00	0.070	0.089	0.049	0.071	0.079	0.081	0.064	0.070
0.03	0.080	0.113	0.082	0.142	0.202	0.198	0.083	0.125
0.06	0.120	0.140	0.181	0.159	0.251	0.136	0.235	0.183
0.09	0.181	0.156	0.363	0.248	0.424	0.259	0.543	0.364
0.12	0.362	0.301	0.576	0.265	0.400	0.463	0.726	0.441
0.15	0.585	0.346	0.832	0.484	0.685	0.525	0.912	0.617
0.18	0.676	0.476	0.945	0.673	0.790	0.560	0.957	0.738
0.21	0.833	0.653	0.990	0.752	0.853	0.654	0.986	0.912
0.24	0.960	0.697	1.000	0.939	0.946	0.798	1.000	0.945
0.27	0.965	0.835	1.000	0.954	0.967	0.867	1.000	1.000
0.30	0.983	0.878	1.000	0.993	1.000	0.956	1.000	1.000

表 5.4.3: 两样本线性模型系数经验似然估计的均方误差根

	$n_1=50$		$n_1=100$		$n_1=200$	
	$n_2=50$		$n_2=150$		$n_2=250$	
η	β	$\hat{\beta}_{ls}$	β	$\hat{\beta}_{ls}$	β	$\hat{\beta}_{ls}$
β_{01}	0.0670	0.1772	0.0693	0.1022	0.0627	0.0639
β_{02}	0.0359	0.1477	0.0340	0.1165	0.0326	0.0624
β_{03}	0.0454	0.1503	0.0478	0.1170	0.0427	0.0708
β_{04}	0.1210	0.1649	0.1221	0.1004	0.1131	0.0624
β_{05}	0.0826	0.1530	0.0824	0.1016	0.0759	0.0680
β_{06}	0.0545	0.1587	0.0559	0.1119	0.0510	0.0555
β_{07}	0.1061	0.1505	0.1097	0.1034	0.0999	0.0653
β_{08}	0.0935	0.1300	0.0950	0.1090	0.0865	0.0612
β_{09}	0.1656	0.1597	0.1713	0.1124	0.1555	0.0621
β_{10}	0.0164	0.1637	0.0173	0.1027	0.0166	0.0511

表5.4.2。

表5.4.2模拟结果表明，当原假设成立($\delta=\delta_0=\beta-\beta_1$)，以及样本容量大到一定程度($n_1=200,\ n_2=250$)时检验方法拒绝原假设的频率接近名义显著性水平。$\delta$与$\delta_0$的差越大，拒绝原假设的频率也越大，当$\eta=\|\delta-\delta_0\|/\|\delta_0\|$接近0.3时，检验方法拒绝原假设的频率非常接近于1。

最后，我们对经验似然估计 $\hat{\beta}$ 的性质进行模拟。继续考虑模型(5.2.1)，令 $X_i\in\mathbb{R}^p$ 和 $X_j\in\mathbb{R}^p$ 为服从协方差矩阵为 p 阶单位矩阵的正态分布的随机变量，真实参数 $\beta=(1,0.5,0.7,1.8,1.2,0.8,1.6,1.4,2.5,0.2)$， $\delta=(-0.8,-0.8,-0.1,1.3,-1.3,-0.2,0.7,-0.3,2.2,-1)$， ε_i 和 ε_j 为服从标准正态分布随机变量。如此模拟重复1000次，其结果见表5.4.3。表5.4.3模拟结果显示，经验似然估计$\hat{\beta}$的均方误差根一致小于最小二乘估计$\hat{\beta}_{ls}$的均方误差根，即，高维情形下经验似然估计方法比普通最小二乘估计更有效。

§5.5 实例分析

下面利用本章方法分析急性淋巴细胞性白血病 (ALL) 数据集 (§2.5

分析过这些数据)。假定$X_1, \cdots, X_n$分别表示n个急性淋巴细胞性白血病病人($n = 128$) 的 B−细胞基因表达数据，每个病人的 B−细胞基因表达数据X_q由p个基因组成($q = 12625$)。我们考虑ALL数据中79个急性淋巴细胞性白血病病人的 B−细胞基因表达数据 (37 个 BCR/ABL 分子类型和 42 个 NEG 分子类型的 B−细胞基因表达数据)。令F_{1X_q}和F_{2X_q}分别表示 BCR/ABL 分子类型和 NEG 分子类型的 B−细胞基因变量的分布函数，其均值分别为μ_{1X_q}和μ_{2X_q}。一个比较有实际意义的问题是两种分子类型病人(BCR/ABL 分子类型和 NEG 分子类型)的 B−细胞基因基因数据的均值是否相等，即检验如下假设

$$H_0 : \mu_{1X_q} = \mu_{2X_q}, \qquad H_1 : \mu_{1X_q} \neq \mu_{2X_q}, \quad \text{for} \quad q = 1, \cdots, n.$$

在进行假设检验之前，类似于Chiaretti et al. (2004)[109]和Gentleman et al.(2005)[110]，我们先对基因数据进行过滤，保留满足如下条件的基因：(1) 至少 75% 的样本被测得的强度达到 100 以上；(2) 样本间的强度变异系数处于 0.7 至 10 之间。经过基因过滤之后，每个急性淋巴细胞性白血病病人的 B−细胞基因表达数据保留431个基因。

利用本章提出的方法对上述假设进行检验，其结果显示，经验似然比检验统计量的值为3.9519，检验的p值为0.000386，其结果表明两种分子类型病人(BCR/ABL 分子类型和 NEG 分子类型)的 B−细胞基因基因数据的均值差异显著。进一步分析可得，在显著性水平$\alpha = 0.01$条件下，两种分子类型病人过滤后保留的431个基因数据中有86个基因差异显著，表5.5.1给出了差异显著的基因以及相应检验的p值。

表 5.5.1: BCR/ABL分子类型和NEG分子类型病人 B−细胞基因差异检验结果

Probe ID	p-value	Probe ID	p-value	Probe ID	p-value
106_at	0.0001	1107_s_at	0.0000	1211_s_at	0.0001
1403_s_at	0.0077	160_at	0.0095	1909_at	0.0051
2057_g_at	0.0058	31526_f_at	0.0019	316_g_at	0.0025
31856_at	0.0006	32542_at	0.0004	32562_at	0.0000
32612_at	0.0010	32649_at	0.0044	32860_g_at	0.0071
Probe ID	p-value	Probe ID	p-value	Probe ID	p-value
32977_at	0.0056	330_s_at	0.0003	33131_at	0.0052
33232_at	0.0000	33284_at	0.0060	33304_at	0.0071

33440_at	0.0000	33700_at	0.0008	33774_at	0.0011
33821_at	0.0039	33891_at	0.0042	34210_at	0.0023
34362_at	0.0028	34677_f_at	0.0013	35626_at	0.0057
35831_at	0.0000	35926_s_at	0.0075	35939_s_at	0.0013
35940_at	0.0085	36234_at	0.0098	36398_at	0.0030
36536_at	0.0005	36412_s_at	0.0010	36591_at	0.0001
6617_at	0.0001	36638_at	0.0001	37006_at	0.0002
37027_at	0.0004	37043_at	0.0001	37283_at	0.0040
37306_at	0.0022	37351_at	0.0001	37398_at	0.0000
37399_at	0.0054	37536_at	0.0037	37539_at	0.0005
37600_at	0.0005	37641_at	0.0066	37732_at	0.0059
38052_at	0.0001	38119_at	0.0000	38124_at	0.0024
38381_at	0.0004	38385_at	0.0015	38408_at	0.0002
38514_at	0.0035	38546_at	0.0000	38578_at	0.0004
38631_at	0.0001	38994_at	0.0002	39070_at	0.0000
39317_at	0.0001	39319_at	0.0001	39329_at	0.0013
39338_at	0.0005	40091_at	0.0030	40202_at	0.0035
40698_at	0.0077	40775_at	0.0088	40888_f_at	0.0019
41071_at	0.0000	41096_at	0.0021	41193_at	0.0002
41468_at	0.0009	41734_at	0.0007	41779_at	0.0085
425_at	0.0004	41827_f_at	0.0012	675_at	0.0014
879_at	0.0001	97935_3_at	0.0060		

第六章 全文总结及展望

随着“大数据”时代的来临，高维数据分析理论受到了统计学者的广泛关注。本文在样本维数发散情形下，利用经验似然方法对高维数据进行统计推断。主要研究结果概括如下：

第一，将经验似然方法推广到高维情形下半参数模型的统计推断。利用经验似然方法构造了参数的估计量及其置信域。当样本维数p和容量n一起趋向无穷情形时，在一定条件下证明了经验似然比统计量的渐近分布为正态分布，并证明了通过经验似然方法得到的参数估计量具有一致性。

第二，将经验似然方法推广到高维删失情形下可加危险率模型的统计推断。利用经验似然方法构造了参数的估计量及参数置信域或参数分量的置信区间(置信域)。当样本维数p和容量n一起趋向无穷情形时，在一定条件下证明了通过经验似然方法得到的参数估计量具有一致性，并证明了经验似然比统计量的渐近分布为正态分布(或 χ_q^2 分布)。

第三，将经验似然方法推广到高维情形下异方差部分线性单指标模型的统计推断。利用经验似然方法构造了参数的估计量及参数置信域或参数分量的置信区间(置信域)。当样本维数p和容量n一起趋向无穷情形时，在一定条件下证明了经验似然比统计量的渐近分布分别为正态分布(或 χ_q^2 分布)。

第四，将经验似然方法推广到高维情形下两样本问题的统计推断。利用经验似然方法构造了两样本均值之差和两线性模型系数之差的估计量及其置信域。在一定条件下证明了经验似然比统计量的渐近分布为正态分布，并证明了通过经验似然方法得到的参数估计量具有一致性。

第五，将惩罚经验似然方法推广到高维稀疏情形下半参数模型的变量选择和参数估计问题。当样本维数p和容量n一起趋向无穷情形时，在一定条件下证明了惩罚经验似然比统计量具有渐近 χ_q^2 分布，而且证明了惩罚经验似然方法具有Oracle性质。

第六，将惩罚经验似然方法推广到高维稀疏删失情形下可加危险率模型的变量选择和参数估计问题。当样本维数p和容量n一起趋向无穷情形时，在一定条件下给出了惩罚经验似然统计量的渐近分布 $-\ \chi_q^2$ 分布，并证明了惩罚经验似然方法具有Oracle性质。

第七，通过模拟和实例分析进一步验证本文所提出的基于经验似然方法的高维数据分析理论结果及其实际意义和应用价值。

由于水平有限，我们的研究工作还很粗浅，很多高维数据分析理论相关问题值得我们进一步研究。下面是我们对下一步研究工作的展望：

第一，在第二章和第四章将惩罚经验似然方法应用于高维半参数模型和高维删失情形下可加危险率模型的变量选择及参数估计。我们提出的结论允许样本维数p和容量n一起趋向无穷，但要求$p < n$。在超高维情形下，即$p > n$且样本维数p和容量n一起趋向无穷时，基于惩罚经验似然方法的变量选择问题值得进一步研究。

第二，本文考虑的变量选择问题是单个参数的变量选择。但在实际问题中，组结构经常出现，即几个相关变量作为一组对响应变量产生影响。例如：在基因表达分析中，某些相关基因属于同一生物途径。因此，在变量选择的时候，我们需要将这些相关基因作为一个组一起被选择进行分析。关于组选择的变量选择方法近年来受到广泛关注，但是利用惩罚经验似然方法进行组选择问题很少有人研究。在样本维数p和容量n一起趋向无穷情形下，特别是超高维$(p > n)$情形下利用惩罚经验似然方法进行组选择的问题是我们今后研究的一个方向。

第三，删失数据长期以来都是统计学者广泛关注的弱点问题之一，本文中我们仅讨论了高维删失情形下可加危险率模型的变量选择问题。因此，我们可以继续考虑将经验似然方法应用于样本维数p 和容量n 一起趋向无穷情形下其他统计模型在删失数据下的变量选择。

第四，Chen et al.(2009)[56]利用经验似然方法在样本维数p和样本容量n一起趋向无穷情形下对一般多元模型进行统计推断时给出了一个p趋向于无穷的速度的最优上界：$p = o(n^{1/2})$。本文理论结果的推导中，为了便于随机分析，限定的p趋向于无穷的速度上界比较保守(第二章中限定为$p = o(n^{1/6})$，第三至第五章中限定为$p = o(n^{1/5})$)。对于本文给出的理论结果，p趋向于无穷的速度上界可以提高到多少，以及其速度的最优上界为多少的问题有待探讨。

参考文献

[1] D. R. Thomas, G. L. Grunkemeier. Confidence intercal estimation for survival probabilities for censored data [J]. *Journal of the American Statistical Association,* 1975, 70: 865 - 871.

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

[3] A. B. Owen. Empirical likelihood ratio confidence regions [J]. *The Annals of statistics,* 1990, 18: 90-120.

[4] A. B. Owen. Empirical likelihood [M]. *Chapman and Hall/CRC*, 2001.

[5] A. B. Owen. Empirical likelihood for linear models [J]. *The Annals of statistics,* 1991, 19: 1725-1747.

[6] S. X. Chen. On the accuracy of empirical likelihood confidence regions for linear regression model [J]. *Annals of the Institute of Statistical Mathematics,* 1993, 45: 621-637.

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

[8] T. DiCiccio, P. Hall, J. Romano. Empirical likelihood is bartlett correctable [J]. *The Annals of statistics,* 1991, 19: 1053-1061.

[9] J. Chen, A. M. Variyath, B. Abraham. Adjusted empirical likelihood and its properties [J]. *Journal of Computational and Graphical Statistics,* 2008, 17: 1-18.

[10] E. D. Kolaczyk. Empirical likelihood for generalized linear models [J]. *Statistica Sinica,* 1994, 4: 199-218.

[11] S. X. Chen, H. J. Cui. An extended empirical likelihood for generalized linear models [J]. *Statistica Sinica,* 2003, 13: 69-81.

[12] J. Qin, J. Lawless. Empirical likelihood and general estimating equations [J]. *The Annals of statistics,* 1994, 22: 300-325.

[13] Q. H. Wang, B. Y. Jing. Empirical likelihood for partial linear model with fixed design [J]. *Statistics and Probability Letters,* 1999, 41: 425-433.

[14] Q. H. Wang, B. Y. Jing. Empirical likelihood for partial linear models [J]. *Annals of the Institute of Statistical Mathematics,* 2003, 55(3): 585-595.

[15] J. Shi, T-S. Lau. Empirical likelihood for partially linear models [J]. *Journal of Multivariate Analysis,* 2000, 72: 132-148.

[16] X. W. Lu. Empirical likelihood for heteroscedastic partially linear models [J]. *Journal of Multivariate Analysis,* 2009, 100: 387-396.

[17] L. G. Xue, L. X. Zhu. Empirical likelihood for single-index models [J]. *Journal of Multivariate Analysis,* 2006, 97: 1295-1312.

[18] L. X. Zhu, L. G. Xue. Empirical likelihood confidence regions in a partially linear single-index model [J]. *Journal of the Royal Statistical Society Series B,* 2006, 68: 549-570.

[19] S. X. Chen, Y. S. Qin. Empirical likelihood confidence intervals for local linear smoothers [J]. *Biometrika,* 2000, 87: 946-953

[20] J. Q. Fan, J. Zhang. Sieve empirical likelihood ratio tests for nonparametric functions [J]. *The Annals of statistics,* 2004, 32: 1858-1907.

[21] G. S. Qin, M. Tsao. Empirical likelihood based inference for the derivative of the nonparametric regression function [J]. *Bernoulli,* 2005, 11(4): 715-735.

[22] S. X. Chen, P. Hall. Smoothed empirical likelihood confidence intervals for quantiles [J]. *The Annals of statistics,* 1993, 21: 1166-1181.

[23] S. X. Chen. Empirical likelihood confidence intervals for nonparametric density estimation [J]. *Biometrika,* 1996, 83: 329-341.

[24] P. Bertail. Empirical likelihood in some semiparametric models [J]. *Bernoulli,* 2006, 12(2):299-331.

[25] S.S. Wang, H. J. Cui, R. Z. Li. Empirical likelihood inference for semi-parametric estimating equations [J]. *Science China(Mathematics),* 2013, 06:1247-1262.

[26] J. H. You, Y. Zhou. Empirical likelihood for semi-parametric varying coefficient partially linear regression models [J]. *Statistics and Probability Letters,* 2006, 76:412-422.

[27] Z. S. Huang, R. Q. Zhang. Empirical likelihood for nonparametric parts in semi-parametric varying-coefficient partially linear models [J]. *Statistics and Probability Letters,* 2009, 79:1798-1808.

[28] Z. S. Huang, R. Q. Zhang. Empirical likelihood for the varying-coefficient single-index model [J]. *The Canadian Journal of Statistics,* 2010, 38(3):434-452.

[29] L. G. Xue, Q. H. Wang. Empirical likelihood for single-index varying-coefficient models [J]. *Bernoulli,* 2012, 18(3):836-856.

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

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

[32] B. Zhong. Quantile processes in the presence of auxiliary information [J]. *Annals of the Institute of Statistical Mathematics,* 1997a, 49: 35-55.

[33] B. Zhong. Empirical likelihood confidence intervals for M-functionals in the presence of auxillary information [J]. *Statistics and Probability Letters,* 1997b, 32: 87-97.

[34] N. H. Chan, S. Q. Ling. Empirical likelihood for GARCH model [J]. *Econometric Theory,* 2006, 22: 403-428.

[35] Y. Kitamura. Empirical likelihood methods with weakly dependent processes [J]. *The Annals of statistics,* 1997, 25: 2084-2102.

[36] F. Bravo. Empirical likelihood based inference with applications to some econometric models [J]. *Econometric Theory,* 2004, 20: 231-264.

[37] G. S. Qin, B. Y. Jing. Empirical likelihood for censored linear regression models [J]. *Scanadinavian Journal of Statistics,* 2001a, 28(4):661-673.

[38] G. S. Qin, B. Y. Jing. Censored patial linear models and empirical likelihood [J]. *Journal of Multivariate Analysis,* 2001b, 78:37-61.

[39] G. S. Qin, B. Y. Jing. Empirical likelihood for the Cox regression model under random censorship [J]. *Communications in Statistics-Simulation and Computation,* 2001c, 30:79-90.

[40] Q. H. Wang, B. Y. Jing. Empirical likelihood for a class functionals of survival distribution with censored data [J]. *Annals of the Institute of Statistical Mathematics,* 2001, 53: 517-527.

[41] Q. H. Wang, G. Li. Empirical likelihood semiparametric regression analysis under random censorship [J]. *Journal of Multivariate Analysis,* 2002, 83: 469-486.

[42] X. Pan, M. Zhou. Empirical likelihood ratio in terms of cumulative hazard function for censored data [J]. *Journal of Multivariate Analysis,* 2002, 80: 166-188.

[43] G. Qin, M. Tsao. Empirical likelihood inference for median regression models for censored survival data [J]. *Journal of Multivariate Analysis,* 2003, 85: 416-430.

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

[45] W.Lu, Y. Liang. Empirical likelihood inference for linear transformation models [J]. *Journal of Multivariate Analysis,* 2003, 85: 416-430.

[46] G. Qin, Y. Zhao. Empirical likelihood inference for the mean residual life under random censorship [J]. *Statistics and Probability Letters,* 2007, 77: 549-557.

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

[48] Y. Liu, C. Zou, R. Zhang. The Bartlett correction for two-sample empirical likelihood [J]. *Statistics and Probability Letters,* 2008, 78: 548-556.

[49] Y. S. Qin, L. C. Zhao. Empirical likelihood ratio confidence intervals for various differences of two populations [J]. *Systems Science and Mathematics,* 2000, 13(1): 23-30.

[50] X. M. Zi, C. L. Z, Y. K. Liu. Two-sample empirical likelihood method for difference between coefficients in linear regression models [J]. *Statistic Papers,* 2012, 53: 83-93.

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

[52] Y. Zhou, H. Liang. Empirical likelihood-based semiparametric inference for the treatment effect in the two-sample problem with censoring [J]. *Biometrika,* 2005, 92: 271-282.

[53] J. Shen, S. He. Empirical likelihood for the difference of two survival functions under right censorship [J]. *Statistics and Probability Letters,* 2006, 76: 169-181.

[54] J. Ren. Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data [J]. *The Annals of statistics,* 2008, 36: 147-166.

[55] H. L. Hjort, I. M. Mckeague, I. V. Keilegom. Extending the scope of empirical likelihood [J]. *The Annals of statistics,* 2009, 37: 1079-1111.

[56] S. X. Chen, L. Peng, Y. L. Qin. Effect of data dimension on empirical likelihood [J]. *Biometrika,* 2009, 96: 712-722.

[57] G. R. Li, L. Lin, L. X. Zhu. Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters [J]. *Journal of Multivariate Analysis,* 2012, 105:85-111.

[58] X. Y. Tang, J. B. Li, H. Lian. Empirical likelihood for patially linear proportional hazards models with growing dimensions [J]. *Journal of Multivariate Analysis,* 2013, 121:22-32.

[59] H. Akaike. Information theory and an extension of the maximum likelihood principle [M]. *Proceedings of the 2nd International symposium on Information Theory, Ed. B. N. Petrov and F. Csaki,* 267-281, 1973. Budapest: Akademia Kiado.

[60] G. Schwartz. Estimating the dimension of a mode [J]. *The Annals of statistics,* 1978, 6: 461-464.

[61] M. Stone. Comments on model selection criteria of Akaike and Schwartz [J]. *Journal of the Royal Statistical Society Series B,* 1979, 41: 276-278.

[62] C. L. Mallows. Some comments on c_p [J]. *Technometrics,* 1973, 15:661-675.

[63] H. Akaika. Fitting autogressive models for prediction [J]. *The Annals of statistics,* 1969, 21: 243-247.

[64] D. M. Allen. The relationship between variable selection and data augmentation and a method for prediction [J]. *Technometrics,* 1974, 16:125-127.

[65] R. J. Tibshirani, K. Knight. The covariance inflation criterion for adaptive model selection [J]. *Journal of the Royal Statistical Society Series B,* 1999, 61: 529-546.

[66] A. E. Hoerl, R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems [J]. *Technometrics,* 1970, 12:55-67.

[67] I. E. Frank, J. H. Friedman. A statistical view of same chemometrics regression tools (with discussion) [J]. *Technometrics,* 1993, 35:109-148.

[68] R. J. Tibshirani. Rrgression shrikage and selection via the lasso [J]. *Journal of the Royal Statistical Society Series B,* 1996, 58: 267-288.

[69] J. Q. Fan, R. Z. Li. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. *Journal of the American Statistical Association,* 2001, 96: 1348 - 1360.

[70] H. Zou, T. Hastie. Rrgression and variable selection via the elastic net [J]. *Journal of the Royal Statistical Society Series B,* 2005, 67: 301-320.

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

[72] E. Candes, T. Tao. The dantzig selector: statistical estimation when p is much larger than n [J]. *The Annals of statistics,* 2007, 35:2313-2351.

[73] C. H. Zhang. Nearly unbiased variable selection under minimax concave penalty [J]. *The Annals of statistics,* 2010, 38:894-942.

[74] J. Q. Fan, H. Peng. Nonconcave penalized likelihood with a diverging number of parameters [J]. *The Annals of statistics,* 2004, 32(3):928-961.

[75] P. Zhao, B. Yu. On model selection consistency of Lasso [J]. *Mach. Learn. Res,* 2006, 7:2541-2567.

[76] J. Huang, J. L. Horowitz, S. Ma. Asymptotic properties of bridge estimators in sparse high-dimensional regression models [J]. *The Annals of statistics,* 2008, 36:587-613.

[77] J. Huang, S. Ma, C. Zhang. Adaptive Lasso for sparse high-dimensional regression models [J]. *Statistica Sinica,* 2008, 18:1603-1618.

[78] H. Zou, H. Zhang. On the adaptive elastic-net with a divergeing number of paramter-s [J]. *The Annals of statistics,* 2009, 37:1733-1751.

[79] W. Lin, J. C. Lv. High-dimensional sparse additive hazards regression [J]. *Journal of the American Statistical Association,* 2013, 501: 247 - 264.

[80] F. Bartolucci. A penalized version of the empirical likelihood ratio for the population mean [J]. *Statistics and Probability Letters,* 2007, 77: 104-110.

[81] T. Otsu. Penalized empirical likelihood estimation of semiparametric models [J]. *Journal of Multivariate Analysis,* 2007, 98: 1923-1954.

[82] W. Hou, L. X. Song, X. G. Wang. Penalized empirical likelihood via bridge estimator in Coxs proportional hazard model [J]. *Communications in Statistics - Theory and Methods,* 2014, 43: 426-440.

[83] X. Y. Tang, C. L. Leng. Penalized high-dimensional empirical likelihood [J]. *Biometrika,* 2010, 97: 905-920.

[84] C. L. Leng, X. Y. Tang. Penalized empirical likelihood and growing dimensional general estimationg equations [J]. *Biometrika,* 2012, 99: 706-716.

[85] S. N. Lahiri, S. Mukhopadhyay. A penalized empirical likelihood method in high dimensions [J]. *The Annals of statistics,* 2012, 40(5):2511-2540.

[86] J. Q. Fan, J. C. Lv. Sure independence screening for ultrahigh dimensional feature space [J]. *Journal of the Royal Statistical Society Series B,* 2008, 70: 894-911.

[87] Q. H. Wang, Z. Zheng. Asymptotic properties for the semiparametric regression model with randomly sensored data [J]. *Science in China, Series A,* 1997, 40: 945-957.

[88] W. Newey, R. Smith. Mukhopadhyay. Higher order properties of gmm and generalized empirical likelihood estimators [J]. *Econometrica,* 2004, 72:219-255.

[89] C. Lam, J. Q. Fan. Profile-kernel likelihood inference with diverging number of parametersOn the adaptive elastic-net with a divergeing number of paramters [J]. *The Annals of statistics,* 2008, 46:2232-2260.

[90] Q. H. Chen, P. S. Zhong, H. J. Cui. Empirical likelihood for mixed-effects error-in-variables model [J]. *Acta Mathematicae Applicatae Sinica, English Series,* 2009, 25:561-578.

[91] S. Dudoit, S. Keles, M. Vanderlaan. Multiple tests of association with biological annotation metadata [J]. *Institute of Mathematical Statistics Collections,* 2008, 2:153-218.

[92] S. X. Chen, Y. L. Qin. A two-sample test for high-dimensional data with applications to gene-set testing [J]. *The Annals of statistics,* 2010, 38:808-835.

[93] O. O. Aalen. A linear regression model for the analysis of life times [J]. *statistics in medicine,* 1989, 8:907-925.

[94] D. Lin, Z. Ying. Semiparametric analysis of the additive risk model [J]. *Biometrika,* 1994, 81:61-71.

[95] A.M. Variyath, J.H. Chen, B. Abrahamc. Empirical likelihood based variable selection [J]. *Journal of Statistical Planning and Inference,* 2010, 140:971-981;

[96] T. Sorlie, R. Partker, T. Hatie, Marron, A. Nobel, S. Deng, H. Johnson, R. Pesich, S. Geisler, J. Demeter, C. Lonning, P. Brown, A. Borresen-dale, D. Botstein. Repeated observation of breast tumor subtypes in independent gene expression data sets [J]. *Proceedings of the National Academy of Sciences,* 2000, 100:8418-8423;

[97] T. Martinussen, T.H. Scheike. Covariate selection for the semiparametric additive risk model [J]. *Scandinavian Journal of Statistics,* 2009, 36:602-619;

[98] R. J. Carroll, J. Fan, I. Gijbels, M. P. Wand. Generalized partially linear single-index models [J]. *Journal of the American Statistical Association,* 1997, 92:477-489;

[99] Y. Yu, D. Ruppert. Penalized spline estimation for partially linear single-index models [J]. *Journal of the American Statistical Association,* 2002, 97:1042-1054;

[100] Y. Xia, H. Tong, W. K. Li, L. X. Zhu. An adaptive estimation of dimension reduction space [J]. *Journal of Royal Statistical society,Series B,* 2002, 64:363-410;

[101] Y. Xia, W. Härdle. Semi-parametric estimation of partially linear single-index models [J]. *Journal of Multivariate Analysis,* 2006, 97:1162-1184;

[102] Y. Y. Ma, L. P. Zhu. Doubly robust and effcient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates [J]. *Journal of Royal Statistical society,Series B,* 2013, 75:305-322;

[103] P. Lai, Q. H. Wang. Semiparametric efficient estimation for partially linear single-index models with responses missing at random [J]. *Journal of Multivariate Analysis,* 2014, 128:33-50;

[104] Y. Ma, J. M. Chiou, N. Wang. Efficient semiparametric estimator for heteroscedastic partially linear models [J]. *Biometrika,* 2006, 943:75-84;

[105] P. G. Hall, C. C. Hyde. Martingale Central Limit Theory and its Applications [M]. *Academic Press, New York*, 1980.

[106] S. M. Hammer, D. A. Katzenstein, M. D. Hughes, H. Gundaker, R. T. Schooley, R. H. Haubrich, W. K. Henry, M. M. Lederman, J. P. Phair, M. Niu, M. S. Hirsch, T. C. Merigan. For the aids clinical trials group study 175 study team, a trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter [J]. *The New England Journal of Medicine,* 1996, 20:1081-1089;

[107] M. Davidian, A. A. Tsiatis, S. Leon. Semiparametric estimation of treatment effect in a pretest-posttest study with missing data [J]. *Statistica Sinica,* 2005, 20:261-301;

[108] Z. Bai, H. Aranadasa. Effect of high dimension: By an example of a two sample problem [J]. *Journal of the American Statistical Association,* 1996, 6:311-329;

[109] S. Chiaretti, X. C. Li, R. Gentleman, A. Vitale, M. Vignetti, F. Mandelli, J. Ritz, R. Foa. Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival [J]. *Blood,* 2004, 103:2771-2778;

[110] R. Gentleman, R. A. Irizarry, V. J. Carey, S. Dudoit, W. Huber. Bioinformatics and Computational Biology Solutions Using R and Bioconductor [M]. *Springer, New York*, 2005.

攻读博士学位期间完成的论文

1. Two-sample high-dimensional empirical likelihood. Communications in Statistics - Theory and Methods. (接受)

2. Penalized empirical likelihood for the additive hazards model with high-dimensional data. Journal of Nonparametric Statistics. (接受)

3. Penalized empirical likelihood for semiparametric model with a diverging number of parameters. Journal of Statistical Planning and Inference. (接受)

4. Empirical likelihood for heteroscedastic partially linear single-index models with high-dimensional data. (投稿)

致 谢

衷心感谢我的导师刘万荣教授! 您严谨的治学态度、崇高的敬业精神、务实的工作作风、高尚的学术风范、深厚广博的学识修养深刻的影响着我, 使我受益终身. 本学位论文能顺利的完成, 倾注了导师大量的心血, 从定题到完稿, 刘老师都十分关心, 并且认真地批阅了全文. 在此, 我向导师表示深深的谢意!

衷心感谢湖南师范大学数学与计算机科学学院的各位老师给我的教导和帮助! 感谢师兄师姐以及师弟师妹对我的帮助和有益讨论!

最后, 我要感谢我的家人在各方面所给予的鼓励和支持!

2017年03月

湖南师范大学学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果．除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果．对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明．本人完全意识到本声明的法律结果由本人承担．

学位论文作者签名：方江林　　　　2017年6月5日

湖南师范大学学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，研究生在校攻读学位期间论文工作的知识产权单位属湖南师范大学．同意学校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，允许论文被查阅和借阅．本人授权湖南师范大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文．

本学位论文属于

1、保密□，在________年解密后适用本授权书．

2、不保密☑

（请在以上相应方框内打"√"）

作者签名：方江林　　　　日期：2017年6月5日

导师签名：刘乃华　　　　日期：2017年6月5日