单位代码	10602
学　　号	2011010857
分 类 号	O212.7
密　　级	公开

广西師範大學
GUANGXI NORMAL UNIVERSITY

硕士学位论文

单指标模型的高维惩罚经验似然

Penalized High Dimensional Empirical Likelihood for Single-index models

学　　院：数学与统计学院

专　　业：概率论与数理统计

研究方向：数理统计

年　　级：2011级

研 究 生：李吉妮

指导教师：张军舰　教授

完成日期：2014年4月

单指标模型的高维惩罚经验似然

专业名称：概率论与数理统计

申 请 人：李吉妮

指导老师：张军舰

论文答辩委员会

主席：

委员：

单指标模型的高维惩罚经验似然

硕士研究生: 李吉妮　导 师: 张军舰

专 业: 概率论与数理统计　研究方向: 数理统计　年 级: 2011

摘 要

单指标模型是广义回归中一种特殊的半参数模型, 是处理多元非参数回归问题的有效工具, 应用非常广泛. 近二十年, 高维数据的变量选择问题已成为统计及其相关领域中研究的热点之一. 在处理高维数据时, 单指标模型的降维特性不仅有效地避免了“维数灾难(curse of dimensionality)”问题, 还抓住了高维数据的稀疏特性. 有关运用单指标模型讨论变量选择方法的文章层出不穷, 但大部分都是针对参数维数 p 是固定时的情况. 然而在很多高维的变量选择问题中, 参数维数 p 一般都会随着样本容量 n 的增大而同时增大. 因此, 在本文中, 我们对单指标模型提出了一种稳健的变量选择方法: 基于 SCAD(Smoothly Clipped Absolute Deviation) 惩罚函数及经验似然的惩罚经验似然. 在一定正则条件下, 我们发现参数维数 p 随样本量 n 同时增大的惩罚经验似然估计仍具有 Oracle 性质, 即如果已知真实模型是稀疏的模型, 则以概率趋向于 1, 惩罚经验似然确定模型的非零参数估计具有稀疏性.

本文在前人已有研究成果的基础上对单指标模型进行综合分析, 主要考察高维单指标模型的参数估计及检验问题. 我们主要结合 Fan&Peng(2004) 的惩罚似然思想和 Hjort(2009),Chen (2009) 的经验似然思想, 针对单指标模型提出了高维惩罚经验似然方法. 理论证明和模拟结果显示, 在处理单指标模型下的变量选择及检验问题中, 惩罚经验似然方法比传统的单一的经验似然方法更加简单有效. 从实用的角度来看, 利用惩罚经验似然方法可以有效节约成本, 实用性更强, 具有较高的推广价值.

本论文特色主要体现在以下几点:

1. 对已有的方法进行重新组合, 取长补短, 提高了估计精度, 拓宽了应用范围;

2. 采用惩罚经验似然方法避免了各自独立使用惩罚似然或经验似然的一些不足. 惩罚似然需要正确的分布假定, 而惩罚经验似然仅需要满足一些矩条件限制即可. 众所周知, 矩条件比分布假定更稳健; 另外, 惩罚经验似然方法下的统计推断问题不用估计参数的方差, 从而使统计推断更加容易.因此,惩罚经验似然的统计推断更具有实际意义.

3. 本文结论可以丰富和完善惩罚经验似然的理论, 为实际应用工作者提供简便可行的工具.

关键词: 单指标模型, 经验似然, 高维数据分析,SCAD

Penalized High Dimensional Empirical Likelihood for Single-index models

Postgraduate: Li Jini Supervisor: Junjian Zhang

Specialty: Probability theory & Mathematical Statistics Research Fields: Mathematical Statistics Grade:

Abstract

Single index model is a special kind of generalized semi-parametric model and is an effective tool for dealing with the problem of multivariate nonparametric regression, and is widely used.Nearly two decades, variable selection of high-dimensional data has become one of the hot areas of statistics and related research.In dealing with high-dimensional data, the droping Witt of single-index model not only effectively avoid the "curse of dimensionality" problems, but also to seize the properties of sparse high-dimensional data.The use of a single index model discussed article about variable selection method after another, but most are for parameter p is the dimension of the fixed time.However, in many high-dimensional variable selection, the parameter dimension p generally increases as the sample size n, while increases.Therefore, in this paper, we propose a single index model robust variable selection methods: Based on the SCAD(Smoothly Clipped Absolute Deviation)penalty function and empirical likelihood of punishment empirical likelihood.Under certain regularity conditions, we found that the parameter p dimension increases with the sample size n increases punishing experience likelihood estimation Oracle still has a nature, that is, if the true sparse model, then the probability tends to 1, the experience seems to punish non-zero parameters of the model to determine the estimated contingent with sparsity.

Based on the results of previous research has been on the basis of a comprehensive analysis of the single index model, mainly on high-dimensional single-index model parameter estimation and testing problems.We mainly with Fan & Peng (2004) the likelihood of punishment ideas and Hjort (2009), Chen (2009) empirical likelihood thought, for a single high-dimensional index model proposed punishment empirical likelihood method.Theoretical proof and simulation results show that the variables in the selection of a single index model processing and inspection issues, empirical likelihood method of punishment than the traditional single empirical likelihood method is more simple and effective.From a practical standpoint, the use of punishment empirical likelihood

method can be cost-effective, more practical, with high promotional value.

This thesis is mainly reflected in the following characteristics:

1.of regrouping existing methods, learn from each other and improve the estimation accuracy and broaden the range of applications;

2.punitive empirical likelihood method avoids the use of separate punishment likelihood or empirical likelihood of some shortcomings. Punishment likelihood requires proper distribution is assumed, and the likelihood of punishment experienced only need to satisfy some moment conditions.As we all know, the distribution of moment conditions more robust than assumed; addition, the experience seems to penalize natural methods of statistical inference problem not under estimate the variance parameters, so that the statistical inference therefore easier to punish the experience more meaningful likelihood of statistical inference.

This paper conclusions can enrich and improve the likelihood of punishment empirical theory for the practical application of workers to provide simple and feasible tool.

Key Words: Single index model; empirical likelihood; high-dimensional data analysis ;SCAD

目 录

中文摘要 II

ABSTRACT III

第一章 绪 论 1

§1.1 研究背景及研究意义 1

§1.2 文献综述 2

§1.3 论文的主要内容和思路 3

§1.4 论文的书写框架 4

§1.5 本文的主要成果和创新点 4

第二章 单指标模型的高维惩罚经验似然 5

§2.1 经验似然方法 5

§2.2 惩罚函数及其推广和门限准则 6

§2.3 惩罚经验似然的表示 8

§2.4 单指标模型的惩罚经验似然 9

§2.5 主要结果 13

§2.6 主要结果的证明 14

第三章 模拟研究 26

§3.1 模拟算法 26

§3.2 模拟条件 27

§3.3 模拟结果 27

第四章 总结和进一步的问题 28

参考文献 29

致 谢 32

声 明 33

第一章 绪 论

§1.1 研究背景及研究意义

随着科学技术的发展, 在实际应用过程中, 人们经常会碰到各种类型的复杂数据，其中比较典型的是高维数据. 高维数据是指观测数据的维数 p 比较大甚至比观测数据的个数 n 还要大. 计量经济学、生物信息学、生物遗传学及航空航天方面等领域都涉及了高维数据的应用. 详见 Hastie,Tibshirani 和 Friedman(2001)[1]. 由于高维数据的普遍存在性, 使得对高维数据的研究有着非常重要的意义. 近几年, 高维数据分析成为统计学范围内的研究专题, 并且为此专题开展了几次国际会议. 例如,2007 年在云南大学召开的高维数据分析国际会议,2008 年由国际数理统计研究学院与浙江大学联合主办的概率统计国际学术会议. 会议中一致认为高维数据分析将成为数理统计发展的前沿.

在高维数据分析过程中, 首先, 碰到最大的问题就是维数的膨胀, 也就是通常所说的“维数灾难 (curse of dimensionality) 问题”. 研究表明, 当维数 p 越来越大时, 数据通常会得越来越稀疏, 与空间的维数相比, 样本量总是显得非常之少. 因此, 分析和处理高维数据的复杂度、成本及所需的空间样本数是呈现指数级数增长的. 其次, 碰到的问题是欧氏距离. 欧氏距离在 2∼10 维的低维空间中是非常有意义的, 可以用来度量数据之间的相似度. 但是在高维空间, 由于高维数据存在稀疏性, 随着维数的增加, 数据对象之间距离的对比性将不复存在, 如果将低维空间中的度量函数应用于高维空间, 这时会大大降低度量函数的有效性. 因此, 在处理高维数据时, 许多处理低维数据的经典或传统方法如主成份分析、回归分析及聚类分析中的层次与划分方法等存在难以解决的问题. 如数据不符合正态分布或对数据没有很多先验信息等. 再次, 我们分析高维数据的基本出发点和关键是从数据描述中找到感兴趣的变量或者在众多的影响中寻找能揭示事物本质规律的变量. 这种方法我们称之为变量选择或者变量降维. 变量选择的目的在于选择出正确有用的变量来解释模型并为之作预测. 这样不但能确保所选模型的稳定性, 而且也避免了变量选择后统计推断出现偏差. 综合高维数据中存在的问题与任务. 在本文中, 我们将采取变量选择与非参数方法去解决上述问题.

另一方面, 单指标模型的独特之处就是为能降低变量的维数, 且又能抓住高维数据的稀疏特征, 有效地避免了“维数灾难”问题, 提高了统计推断的精度. 它是广义回归模型中的一种特殊的统计模型, 该模型是在八十年代中后期发展起来的, 在经济、生物等统计学领域具有广泛的应用背景. 正因如此, 国内外很多学者从不同角度及不同假设条件对此模型进行一系列的研究. 然而在高维数据上对单指标模型的研究还很少. 因此, 在本文中, 我们融合单指标模型的优势及惩罚经验似然在 p 随 n 同时增大的有效性, 对单指标模型在高维数据时

的惩罚经验似然进行研究, 这是非常有实际意义的.

本章第一节给出有关单指标模型的已有方法及研究思路, 并相应指出这些方法的优劣势; 第二节将对单指标模型理论, 经验似然方法及惩罚思想等方面做一般性的介绍, 第三节给出我们的主要研究内容及创新之处.

§1.2 文献综述

为了讨论方便及叙述更为清晰, 本节分类地进行文献综述, 先给出单指标模型的综述,接着书写高维惩罚经验似然的发展进程, 最后给出文献综述的评价.

(1) 单指标模型

单指标模型首先由 Friedman 等 (1981)[2] 在研究投影追踪回归问题时, 为了降低数据的维数而提出来的, 它是一种重要而又广泛的半参数模型, 是处理多元非参数回归问题的有效工具. 单指标模型首先考虑 p 维解释变量 X 的线性组合, 把所有的解释变量投影到一个线性空间上, 然后在这个一维的线性空间上拟合一个一元函数 [3]. 因此, 有效避免了多元非参数回归中所谓的“维数灾难”问题. 单指标模型中未知参数的估计问题被广泛研究.Ichimura(1993)[4] 和 Horowitz(1998)[5] 研究了单指标模型的可识别性,Manski(1988)[6] 讨论了二元响应模型的可识别性.Stoker(1986)[7]、Ichimura(1993)[4] 和 Härdle(1993)[8] 采用参数回归模型中最小二乘估计的思想提出了半参数最小二乘估计方法, 由于该方法需要解决非常复杂的非线性最优化问题, 因此导致计算相当的复杂;Härdle&Stoker(1989)[9],Powell 等 (1989)[10]、Tsybakov[11] 利用平均导数法获得指标模型的参数估计, 但是使用平均导数法有一个问题是随机分母的出现使分母在一些点趋于 0. 为此, 研究者采取了一些方法比如对函数截尾或对函数的支撑加以限制, 从而避免了随机分母可能为 0 的情形, 如 Rilstone(1991)[12] 利用截尾函数回避“小分母问题”, 从而建立参数向量的渐近正态性;Xue 和 Zhu(2006)[13] 采用经验似然方法构造了单指标模型中未知参数的置信域, 并得出该经验似然比统计量渐近于加权的 χ^2 分布, 为了减少计算量及提高估计精度, 从而提出了调整的经验似然比统计量, 并证明了该统计量是渐近于标准 χ^2 分布;Härdle(1997)[14],Xia(2009)[15],Escanciano&Song(2010)[16],Klein&Shen(2010)[17] 等研究了单指标模型的检验问题.

最近, 利用单指标模型对复杂数据进行统计分析取得了一些成果.Bai(2009)[18] 用惩罚样条与二次推断函数方法研究了纵向数据下线性模型单指模型, 并证明了其中的渐近性质.Lin&Kulasekera(2009)[19] 则在纵向数据下对单指标模型的同等问题进行了研究.Lopez(2009)[20] 研究了右删失响应下的单指标模型, 对 M- 估计方法进行了推广并证明了提出的估计量的相合性和渐近正态性.Lu(2010)[21] 对随机删失样本的单指标模型提出了两个估计方法, 构造了指标系数估计量并证明了它们的 $\sqrt{n}$ 相合性和渐近正态

性.Wang(2010)[22] 在缺失响应下对单指标模型中兴趣参数的估计问题进行了研究, 并对所提出的大样本性质进行了证明.

(2) 高维经验似然

经验似然方法是 Owen 于 1988 年在完全数据样本下系统提出及发展的一种构造未知参数的置信域的非参数统计方法, 它与已有的统计方法相比, 优点更加突出. 如构造的置信域具有域保持性、变换不变性、置信域的形状完全由数据自行决定而不用预先给定、Bartlett 纠偏性及无需构造枢轴量等等 [23].

对于高维数据问题的经验似然函数的存在性是一项比较突出的问题.Tsao(2004)[24] 指出, 当 $p > \frac{n}{2}$ 时, 基于样本量 n 的 p 维总体均值的经验似然函数定义在一个正概率集上是失效的. 其主要原因是遭遇“凸包”问题, 因此, 标准的经验似然方法不能应用于 $p > \frac{n}{2}$ 的高维问题. 为了解决这种高维问题, 从已有的研究文献, 可以归结为如下两种解决方法. 在添加伪数据方面,Chen、Variyath、Abraham(2008)[25] 通过添加一个额外的伪数据, 给出了调整经验似然.Emerson&Owen(2009)[27] 在调整经验似然的基础上又做了更进一步的研究, 其主要是添加了两个伪数据, 以便将均值参数的假设值覆盖在扩张数据集的凸包里, 从而使得调整经验似然有了很好的定义. 在运用惩罚思想方面,Bartolucci(2007)[27] 将凸包一起放到带有约束条件的经验似然中, 通过马氏距离惩罚不带约束的经验似然, 来重新定义参数的似然, 但是对于 $p > n$ 的情况,Bartolucci 并不能对其有很好地定义, 主要是因为公式中使用到样本协差阵的逆.Fan 和 Li(2001)[28],Fan 和 Peng(2004)[29] 提出了惩罚似然思想, 并将其应用到高维问题中, 得到了很多令人满意的结果, 但是惩罚似然方法需要正确的分布假定, 然而实际中得到的数据一般是不知道其分布情况的. 为此, Otsu(2007)[30],Tang 和 Leng(2010)[31] 利用非参数方法在标准经验似然中通过添加惩罚函数, 研究表明, 当 p 增大到 n 的分数幂时, 其方法是有效的.Hjort、Mckeague、Van Keilogom(2009)[32] 提出当数据维数 p 随着样本容量 n 增大而增大时, 经验似然方法仍然有效.

从上述的综述中我们不难发现, 国外的研究早于国内研究, 且相对来说较为成熟. 此外, 国内外研究结果存在一个共同点. 即对单指标模型的研究没有同时涉及到高维和非参的情形, 使得研究许多实际中的数据的方法难以兼顾, 或多或少存在这样那样的不足. 不同的解决方法有各自的优点和缺点, 然而它们却可以相互补充. 为此, 借鉴单指标模型的优点, 从高维情形去分析单指标模型的统计推断, 这样才能够较为准确的估计模型参数.

§1.3 论文的主要内容和思路

本论文主要研究单指标模型下, 数据维数在不断增加时的经验似然方法, 其中包括讨论了惩罚经验似然估计的相合性,Oracle 性质以及惩罚经验似然的渐近卡方性质.

本论文的研究思路: 首先借助 Fan&Li(2001)[33] 提出的基于 SCAD 惩罚函数的惩

罚似然方法, 将变量选择从最基本的线性模型推广到更广、更复杂的模型; 然后借助 Owen(1988)[24] 提出的经验似然方法对于数据具体分布未知情况下仍可进行统计推断, 加上 Qin&Lawless(1994)[35] 可将经验似然和估计方程进行有机结合; 最后借助 Fan&Peng(2004)[34] 能够将惩罚似然推广到参数维数 p 随着样本量 n 同时增大以及 Hjort 等 [32],Chen 等 [35] 关于经验似然在参数维数 p 随着样本量 n 同时增大仍有效的思想; 对上述三者思想的有机结合, 我们针对单指标模型提出了参数维数 p 同时随着样本量 n 增大而增大的惩罚经验似然. 模拟得结果表明, 我们提出的惩罚经验似然方法是可行的.

§1.4 论文的书写框架

本论文的写作框架如下: 全论文共分四部分: 第一章主要介绍了课题研究的背景及前人工作的成果; 并给出本论文的研究思路、内容及创新点; 第二章构造了未知参数 θ 的惩罚经验似然估计值, 并发现参数维数 p 随样本量 n 增大而增大时的惩罚经验似然估计仍具有 Oracle 性质及惩罚经验似然的渐近正态性; 第三章给出了模拟结果, 并证实了惩罚经验似然的作用; 第四章是对全论文进行总结并给出进一步研究的问题.

§1.5 本文的主要成果和创新点

本论文的主要成果和创新点可用以下几点进行归纳:

1. 本论文提出的参数维数 p 随样本量 n 同时增大的 SCAD 惩罚经验似然方法对单指标模型进行变量选择, 大大地推广了惩罚经验似然的应用范围, 并且证明了在大样本下惩罚经验似然仍具有相合性和 Oracle 性及惩罚经验似然的渐近卡方性. 这些成果是前人所没有讨论的, 是完全新的成果.

2. 本论文结论可以丰富和完善惩罚经验似然的理论, 为实际应用工作者提供简便可行的工具.

第二章　单指标模型的高维惩罚经验似然

§2.1 经验似然方法

为了讨论方便, 下面对经验似然方法及基本思想进行简单介绍.

设 $X_1, X_2, \cdots, X_n$ 是独立同分布于分布函数为 F(未知) 的 d 维随机向量,$(x_1, x_2, \cdots, x_n)$ 为其对应的观测值,θ 是 p 维待估计参数向量,$\theta = (\theta_1, \theta_2, \cdots, \theta_p)^{\mathrm{T}}$. 并且知道某一已知的 r 维随机向量函数的均值满足

$$E_F[g(x, \theta)] = 0.$$

其中 $g(x, \theta) = (g_1(x, \theta), g_2(x, \theta), \cdots, g_r(x, \theta))^{\mathrm{T}}, (r \geq p)$.

在上述信息情况下, 我们希望对待估参数 θ 进行统计推断. 因为分布 F 未知, 因此, 我们将借助非参数似然来对它进行讨论. 定义非参数似然函数为

$$L(F) = \prod_{i=1} \mathrm{d}F(x_i) = \prod_{=1} p_i,$$

其中 $p_i = dF(x_i) = F(\{x_i\}) = Pr(Y_i = y_i)$. 我们都知道, 如果没有 $g(x, \theta)$ 这个附加的信息, 分布函数 F 的非参数极大似然估计为经验分布函数 $F_n = n^{-1} \sum_{i=1} I[X_i \leq x]$, 即在 $p_i = 1/n (i = 1, 2, \cdots, n)$ 时, 上式达到最大值. 它实际上是

$$L(\theta) = \sup \left\{ \prod_{i=1}^{n} p_i | p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}.$$

的解所对应的 p_i. 即

$$L(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}) = \prod_{i=1}^{n} \frac{1}{n} = n^{-n}.$$

因此在附加信息 $E[g(x, \theta)] = 0$ 下, 很自然地想到应将此信息作为约束条件, 也就是定义如下似然函数为

$$L(\theta) = \sup \left\{ \prod_{i=1}^{n} p_i | p_i \geq 0, \sum_{i=1}^{n} p_i g(x_i, \theta) = 0 \right\},$$

这就是由 Qin & Lawless[33] 于 1994 定义的半参数模型中的关于待估参数 θ 的经验似然函数. 其中当 $g(x, \theta) = x - \theta$ 时, 上式便是 Owen 于 1988 年所定义的经验似然函数. 因此, 可定义参数 θ 的对数经验似然函数为

$$l(\theta) = -2 \log L(\theta).$$

求解带有约束条件的函数的最大值问题, 一般采用 Lagrange 乘子法, 容易得到

$$\hat{p}_i = \frac{1}{n\{1 + \lambda^{\mathrm{T}} g(x_i, \theta)\}}.$$

其中 λ 满足

$$\sum_{i=1}^{n} \frac{g(x_i;\theta)}{1 + \lambda^{\mathrm{T}} g(x_i;\theta)} = 0.$$

因此, 估计量 θ 的对数经验似然可写为

$$l(\theta) = 2\sum_{i=1}^{n} \log(1 + \lambda^{\mathrm{T}} g(x_i;\theta)) + 2n\log n.$$

Qin&Lawless(1994)[33] 证明了在适当的正则条件下, 对固定的真值 θ_0 有

1. 以概率 1, 对充分大的 $n, l(\theta)$ 在 $||\theta - \theta_0|| \leq n^{-(1/3)}$ 内部某点 $\hat{\theta}_n$ 处达到最小点;

2.$\sqrt{n}(\hat{\theta}_n - \theta_0)$ 依分布收敛于正态分布;

3. $l(\theta)$ 依分布收敛于 $\chi^2(r)$(自由度为 r 的中心卡方分布); 当假设 $H_0 : \theta = \theta_0$ 成立时, 有 $l(\theta_0) - l(\hat{\theta}_n)$ 收敛于 $\chi^2(p), l(\hat{\theta}_n)$ 依分布收敛于 $\chi^2(r-p)$.

在参数统计推断中, 人们利用参数似然比进行假设检验和置信区间估计. 类似地, 在分布函数 F 未知的情况下, 我们也可以用非参数似然比进行统计推断, 即

$$R(F) = \frac{L(F)}{L_n(F_n)} = \prod_{i=1}^{n} p_i / \frac{1}{n} = \prod_{i=1}^{n} np_i,$$

由似然函数的定义及性质可知, 上述式子的改变并不会影响参数的统计推断. 因此, 我们也可以通过经验似然比对参数进行统计推断 [39].

§2.2 惩罚函数及其推广和门限准则

近年来, 随着社会的不断发展, 统计所面临的问题也越来越多, 相应地, 为解决这些问题而提出的新方法和思想也层出不穷. 在这些新方法和思想中, 比较突出的是惩罚思想的应用. 所谓惩罚, 顾名思义, 是指在某一反应出现后, 能够抑制那一反应的事物（包括方法和手段）. 它实际上是对某一现象保持均衡的一种手段. 统计学中有许多方法可以用来解决同一问题, 在实际数据处理时, 每一种方法都有自己的优点和缺点, 单靠一种方法往往很难给出满意的结果, 此时就需要对这些方法的负面影响进行惩罚, 从而得到相对比较满意的结论.

下面简单给出几种常见的惩罚函数的定义及与之对应的门限准则:

1. 硬门限 (hard thresholding) 惩罚函数 [36]:

$$p_\tau(|\theta|) = \tau^2 - (|\theta| - \tau)^2 I(|\theta| \leq \tau),$$

其中 θ 为估计参数,τ 为调整参数, 下文中出现的符号代表的意思也一样. 产生相应的硬门限

准则 (Hard-thresholding rule):

$$\hat{\theta} = zI(|z| \geq \tau).$$

2. 软门限 (L_1,LASSO) 惩罚函数:

$$p_\tau(|\theta|) = \tau|\theta|,$$

产生相应的软门限准则 (Soft-thresholding rule):

$$\hat{\theta} = sgn(z)(|z| - \tau)_+,$$

其中 $sgn(.)$ 为符号函数,$(|z| - \tau)_+$ 表示 $(|z| - \tau)$ 的正部.

3.SCAD(Smoothly Clipped Absolute Deviation Penalty) 惩罚函数:

$$p_\tau(|\theta|) = \begin{cases} \tau|\theta|, & 当0 \leq |\theta| \leq \tau \\ -(|\theta|^2 - 2a\tau|\theta| + \tau^2)/\{2(a-1)\}, & 当\tau \leq |\theta| < a\tau \\ (a+1)\tau^2/2, & 当|\theta| \geq a\tau. \end{cases}$$

其中 $a > 2, \theta > 0$. 或者是它关于 $|\theta|$ 的一阶导数

$$p'_\tau(|\theta|) = \tau\{I(|\theta| \leq \tau) + \frac{(a\tau - \theta)_+}{(a-1)\tau} I(\theta > \tau)\}.$$

产生相应的 SCAD 门限准则 (SCAD-thresholding rule):

$$\hat{\theta} = \begin{cases} sgn(z)(|z| - \tau)_+, & 当|z| \leq 2\tau, \\ \{(a-1)z - sgn(z)a\tau\}(a-2), & 当2\tau \leq |z| \leq a\tau, \\ z, & 当|z| > a\tau. \end{cases}$$

惩罚函数几乎决定了惩罚似然估计值的样本性质, 但是什么样的惩罚函数才算是一个好的惩罚函数呢? Fan 和 Li(2001)[32] 指出, 选择一个好的惩罚函数, 由此函数得出的解必须同时具有下面三方面的性质:

1. 无偏性: 当真实未知参数不是零时, 相应的估计应该是无偏的, 即对大参数没有过大的惩罚, 这样可以避免不必要的模型偏差;

2. 稀疏性: 惩罚似然估计值应该有一个限制, 这个限制自动将那些较小的估计系数设为 0, 进而去掉, 并删除对应的变量, 从而降低模型的复杂度;

3. 连续性: 所得的惩罚最小二乘估计应具有连续性, 从而降低模型在做预测时的不稳定性.

接下来我们比较这三种惩罚函数的二次逼近和相应门限准则的关系, 在下面的图 2.1 与图 2.2 中, 我们取 $\tau = 2, a = 3.7$. 图 2.1 与图 2.2 分别展示了这些不同的惩罚函数与门限准

则产生估计之间的关系.

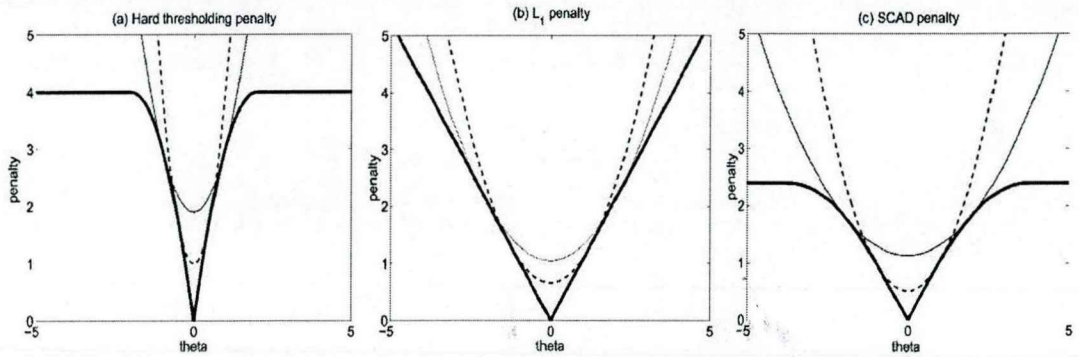

图 2.1　三种惩罚函数 $p_\lambda(\theta)$ 及它们的二次逼近.

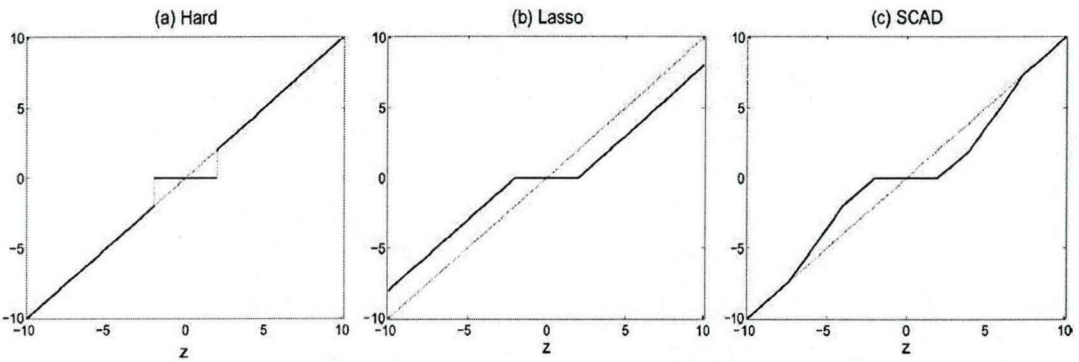

图 2.2　三种惩罚函数 $p_\lambda(\theta)$ 的惩罚最小二乘估计和最小二乘估计之间的关系.

由图可知, 硬门限惩罚函数, 软门限惩罚函数都不能同时满足上面的三条性质. 因为由图 2.2 可知, 硬门限准则是不连续的, 软门限准则是有偏的,SCAD 则结合了两者的优点, 即在连续的情况下, 还能保证参数在很大时是无偏的. 因此它自然满足了上面三条性质. 其实 SCAD 惩罚函数是 Fan 和 Li(2001)[32] 根据这三条性质构造出来的惩罚函数, 该函数被称为绝对偏差的光滑剪切 (简称 SCAD). 详见 Fan and Li(2001)[32].

§2.3 惩罚经验似然的表示

惩罚经验似然方法, 顾名思义, 就是在经验似然的基础上融合 Fan 和 Li(2001) 提出的惩罚似然思想. 一般情况下的惩罚经验似然是以经验似然 + 惩罚函数作为基本框架. 为后面讨论方便, 我们简单叙述一下 Otsu(2007)[29] 和 Leng(2010)[30] 所提出的惩罚经验似然方法的思想. 为了描述简洁, 限定均值参数为 μ 的情形, 令

$$L_n^{ST}(\mu) = \sup\left\{\prod_{i=1}^{n}\pi_i : (\pi_1,\pi_2,\cdots,\pi_n)\in\Pi_n, \sum_{i=1}^{n}\pi(X_i-\mu)=0\right\}.$$

表示为 $\mu \in R^p$ 的传统经验似然. 同时令 $p_\tau(\cdot)$ 为惩罚函数, 则惩罚经验似然的形式为

$$L_n^{OTL}(\mu) = L_n^{ST}(\mu)\exp\left(-n\sum_{j=1}^{p} p_\tau(\mu_j)\right).$$

其中 τ 为调整参数, 我们并不要求调整参数对所有的 j 都相同, 对不同的参数有不同的惩罚尺度. 因此, 惩罚经验似然具有非常的灵活性, 适用性更强. 对于更一般的参数的情形, 可通过估计方程去定义, 其中 Otsu(2007)[29]、Tang&Leng(2010)[30] 的研究中针对相应问题构建的经验似然去替换上式中的 $L^{ST}(\mu)$.

§2.4 单指标模型的惩罚经验似然

假设 $\{(X_i, Y_i)_{i=1}^n\}$ 是来自 (X, Y) 的 iid 样本, 单指标模型

$$Y_i = g(\theta^{\mathrm{T}} X_i) + \varepsilon_i. \tag{2.1}$$

其中 $X_i = (X_{i1}, X_{i2}, \cdots, X_{ip})^{\mathrm{T}} \in \mathcal{R}^p,\ i = 1, 2, \cdots, n$, 是独立同分布的 p 维协变量, 且为了简单起见, 本文中假定 $E(X_i) = 0, Var(X_i) = \Sigma = (\sigma_{jk})_{j,k=1}^p$, 对于其它情况可进行中心化处理得到. $g(\cdot)$ 是未知的一元联系函数,未知参数$\theta = (\theta_1, \theta_2, \cdots, \theta_p)^{\mathrm{T}}$ 为 $p \times 1$ 且满足 $||\theta|| = 1$ (其中 $||\cdot||$ 表示欧拉范数),ε_i 是 iid 的随机误差且有 $E(\varepsilon_i|X_i) = 0, Var(\varepsilon_i^2|X_i) = 1$ 几乎处处成立. 约束条件 $||\theta|| = 1$ 是为了模型的可识别 [14]. 在模型 (2.1) 中, 注意到由 $||\theta|| = 1$ 这个约束条件, 可知真实参数 θ 是一个单位球上的边界点. 然而在这些边界点上联系函数 $g(\theta^{\mathrm{T}} X_i)$ 可能不存在导数. 而由 $||\theta|| = 1$ 可知 θ 只有 $p-1$ 个分量是自由的. 因此, 我们将采用 “去一个分量” 的方法. 不妨设 $\theta^{(r)} = (\theta_1, \theta_2, \cdots, \theta_{r-1}, \theta_{r+1}, \cdots, \theta_p)^{\mathrm{T}}$ 是 θ 删去第 r 个分量 θ_r 所形成的 $(p-1)$ 维向量, 这样 θ 的 $p-1$ 个分量值完全可以确定 θ, 因此, 可以通过研究 θ 的这 $p-1$ 个分量的性质来研究 θ, 这样不仅可以降低了参数的维数, 也提高了估计参数的精度. 不失一般性, 我们假定真实参数 θ 具有正的分量 θ_r, 否则, 可以考虑 $\theta_r = -(1 - ||\theta^{(r)}||^2)^{1/2}$. 因此,$\theta$ 表示为

$$\theta = \theta(\theta^{(r)}) = (\theta_1, \cdots, \theta_{r-1}, (1 - ||\theta^{(r)}||^2)^{1/2}, \theta_{r+1}, \cdots, \theta_p)^{\mathrm{T}}.$$

真实参数 $\theta^{(r)}$ 一定满足限制条件 $||\theta^{(r)}|| < 1$. 因此,θ 在真实参数 $\theta^{(r)}$ 的邻域内是无穷阶可微的,Jacobian 矩阵为

$$J_{\theta^{(r)}} = \frac{\partial \theta}{\partial \theta^{(r)}} = (\gamma_1, \gamma_2, \cdots, \gamma_p)^{\mathrm{T}},$$

其中 $\gamma_s (1 \le s \le p, s \ne r)$ 是第 s 个分量为 1 的 $p-1$ 维的单位向量, 且

$$\gamma_r = -(1 - ||\theta^{(r)}||^2)^{-1/2}\theta^{(r)}.$$

定义 $X_i^{(r)}=(X_{i1},X_{i2},\cdots,X_{i,r-1},X_{i,r+1},\cdots,X_{ip})^{\mathrm{T}}$. 因为

$$\theta^{\mathrm{T}}X_i=\theta^{(r)}X_i^{(r)}+(1-||\theta^{(r)}||^2)^{1/2}X_{ir}.$$

因此可知,$g(\theta^{\mathrm{T}}X_i),g'(\theta^{\mathrm{T}}X_i)$ 都是 $\theta^{(r)}$ 的函数. 因此, 我们定义如下的随机辅助向量

$$U_i(\theta^{(r)})=[Y_i-g(\theta^{\mathrm{T}}X_i)]g'(\theta^{\mathrm{T}}X_i)J_{\theta^{(r)}}^{\mathrm{T}}X_i.$$

若 θ 是真实参数, 则有 $E[U_i(\theta^{(r)})]=0.i=1,2,\cdots,n$. 由此, 我们使用这个信息, 来构建单指标模型的经验似然为

$$L(\theta^{(r)})=\sup\left\{\prod_i^n p_i|p_i\geq 0,\sum_i^n p_i=1,\sum_i^n p_iU_i(\theta^{(r)})=0\right\}.$$

$L(\theta^{(r)})$ 中包含有未知的函数 $g(\cdot)$ 和 $g'(\cdot)$, 因此, 不能直接应用它对 θ 进行统计推断. 解决这个问题的一个自然的想法就是将 $L(\theta^{(r)})$ 中的 $g(\cdot)$ 和 $g'(\cdot)$ 分别用它们的估计 $\hat{g}(\cdot),\hat{g}'(\cdot)$ 代替.

对 $g(\cdot),g'(\cdot)$ 估计, 我们采用局部线性最小二乘估计 [37]. 假定 θ 已知, 通过最小化 (2.2) 式, 可以分别得到 $\hat{g}(\cdot),\hat{g}'(\cdot)$. 首先, 我们通过寻找 a,b 使得 (2.2) 式达到最小值.

$$\sum_{i=1}^n(Y_i-a-b(\theta^{\mathrm{T}}X_i-t))^2K_h(\theta^{\mathrm{T}}X_i-t). \tag{2.2}$$

其中 $K_h(\cdot)=h^{-1}K(\cdot/h)$ 是一个核函数.$h=h_n$ 及h_1是一列趋于 0 且为正数的窗宽. 令 $\hat{a}$ 和 $\hat{b}$ 是加权最小二乘问题 (2.2) 的解. 那么

$$\hat{a}=U_{ni}(t;\theta,h)Y_i\diagup\sum_{j=1}^n U_{nj}(t;\theta,h),$$

$$\hat{b}=\widetilde{U}_{ni}(t;\theta,h)Y_i\diagup\sum_{j=1}^n U_{nj}(t;\theta,h).$$

其中

$$U_{ni}(t;\theta,h)=K_h(\theta^{\mathrm{T}}X_i-t)[S_{n,2}(t;\theta,h)-(\theta^{\mathrm{T}}X_i-t)S_{n,1}(t;\theta,h)],$$

$$\widetilde{U}_{ni}(t;\theta,h_1)=K_h(\theta^{\mathrm{T}}X_i-t)[(\theta^{\mathrm{T}}X_i-t)S_{n,0}(t;\theta,h)-S_{n,1}(t;\theta,h)],$$

$$S_{n,l}(t;\theta,h)=\frac{1}{n}\sum_{i=1}^n(\theta^{\mathrm{T}}X_i-t)^lK_h(\theta^{\mathrm{T}}X_i-t),\qquad l=0,1,2.$$

通过简单计算可得

$$\hat{g}(t;\theta)=\sum_{i=1}^n W_{ni}(t;\theta)Y_i,$$

$$\hat{g}'(t;\theta)=\sum_{i=1}^{n}\widetilde{W}_{ni}(t;\theta)Y_i.$$

其中

$$W_{ni}(t;\theta)=U_{ni}(t;\theta,h)\Big/\sum_{j=1}^{n}U_{nj}(t;\theta,h),$$

$$\widetilde{W}_{ni}(t;\theta)=\widetilde{U}_{ni}(t;\theta,h_1)\Big/\sum_{j=1}^{n}U_{nj}(t;\theta,h_1).$$

则估计量的经验似然函数为

$$\hat{L}(\theta^{(r)})=\sup\left\{\prod_{i=1}^{n}p_i|p_i\geq 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\hat{U}_i(\theta^{(r)})=0\right\}. \tag{2.3}$$

其中 $\hat{U}_i(\theta^{(r)})$ 是带有 $g(\theta^{\mathrm{T}}X_i),g'(\theta^{\mathrm{T}}X_i)$ 的 $U_i(\theta^{(r)})$ 分别用 $\hat{g}(\theta^{\mathrm{T}}X_i;\theta),\hat{g}'(\theta^{\mathrm{T}}X_i;\theta),i=1,\cdots,n$ 替换后所得的估计量.

可定义估计量 $\theta^{(r)}$ 的对数经验似然比函数为

$$\hat{l}(\theta^{(r)})=-\log\{\hat{L}(\theta^{(r)})\}-n\log(n). \tag{2.4}$$

由 Lagrange 乘子法有,$p_i=n^{-1}\{1+\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})\}^{-1}$, 其中 λ_θ 满足 $n^{-1}\sum_{i=1}^{n}\frac{\hat{U}_i(\theta^{(r)})}{1+\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})}=0$.
基于 Qin 和 Lawless[33] 关于估计方程的经验似然和 Fan 和 Li 提出的惩罚似然思想, 我们定义估计量 $\hat{\theta}$ 的惩罚经验似然为

$$\log\{\hat{L}(\theta^{(r)})\}+n\log n-n\sum_{i=1}^{p}p_\tau(|\theta_i^{(r)}|). \tag{2.5}$$

其中 $p_\tau(\cdot)$ 是惩罚函数,τ 是调整参数. 由于 Fan 和 Li 提出的 SCAD 惩罚函数是结合其它惩罚函数的优点得到的, 并且满足稀疏性、无偏性和连续性等性质. 因此, 在本文研究中, 我们将采用 SCAD 惩罚函数. 根据 Fan 和 Li 对 a 的讨论结果, 我们在后面的模拟中取 $a=3.7$.

将 p_i 代入到（2.5）可以得到 (2.5) 的最大值等价于

$$\hat{l}(\theta_p^{(r)})=\sum_{i=1}^{n}\log\{1+\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})\}+n\sum_{i\neq r}^{p}p_\tau(|\theta_i|). \tag{2.6}$$

的最小值.

假设对于每个独立同分布的观测值 $X_i\in\mathcal{R}^p$, 我们规定 $X_i=\Gamma Z_i+\mu,\Gamma$ 是 $p\times m$ 的矩阵, 并且 $m\geq p,\Gamma\Gamma^{\mathrm{T}}=\sum,Z_i=(Z_{i1},Z_{i2},\cdots,Z_{im})^{\mathrm{T}}$ 使得满足

$$E(Z_i)=0,Var(Z_i)=I_m,E(Z_{il})^{4k}=m_{4k}\in(0,\infty),$$

$$E(Z_{il_1}^{\alpha_1}Z_{il_2}^{\alpha_2}\cdots Z_{il_q}^{\alpha_q})=E(Z_{il_1}^{\alpha_1})E(Z_{il_2}^{\alpha_2})\cdots E(Z_{il_q}^{\alpha_q}). \tag{2.7}$$

其中, 当 $\sum_{l=1}^{q}\alpha_l \le 4k$, 且 $l_1 \ne \cdots \ne l_q.k \in N^+, I_m$ 是一个单位阵.

令 $\mathcal{L} = \{j : \theta_{0j} \ne 0, j \ne r\}$ 表示真实参数向量 $\theta_0^{(r)}$ 中非零元素所构成的集合, 且 $|\mathcal{L}| = d$, 其中在 $n \to \infty$ 时,d 满足条件 3. 记 $\theta^{(r)} = (\theta_1^{(r)\mathrm{T}}, \theta_2^{(r)\mathrm{T}})^{\mathrm{T}}$, 其中 $\theta_1^{(r)} \in \mathcal{R}^d$, $\theta^{(r)} \in \mathcal{R}^{p-d}$. 因此, 真实参数为 $\theta_0^{(r)} = (\theta_{10}^{(r)\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}$.

假设条件

A1: 观测值 $\{X_i\}_{i=1}^n$ 独立同分布, 对于 $k \ge 3$, 满足模型 (1.6).

A2:Σ 的特征根满足:$C_1 \le \gamma_1(\Sigma) \le \cdots \le \gamma_p(\Sigma) \le C_2$, 其中 $0 < C_1 < C_2 < \infty$.

A3: 当 $n \to \infty, p \to \infty$, 时, 对引理 4 中给定的 δ, 则有 $(p-1)^2/n^{1-1/(4k)} \to 0$, $(p-1)^{1-2\delta}/n^{1/2-2\delta} \to 0$, 并且 $d \le p-1$.

A4: 当 $n \to \infty$ 时, 对于在 A3 中给定的 δ, 调整参数 τ 满足 $\tau \to 0, \tau(n/(p-1))^{1/2-\delta} \to \infty$. 非零元满足 $\min_{j\in\mathcal{L}}|\theta_{oj}|/\tau \to \infty$, 零元满足 $\max_{j\notin\mathcal{L}}|\theta_{0j}|/\tau \to 0$;

A5: 惩罚函数 $p_\tau(\cdot)$ 满足 $\max_{j\in\mathcal{L}} p'_\tau(|\theta_{0j}|) = o\{(n(p-1))^{-1/2}\}$, $\max_{j\in\mathcal{L}} p''_\tau(|\theta_{0j}|) = o\{(p-1)^{-1/2+\delta}n^{-\delta}\}$.

A6:$\{\widetilde{\varepsilon}_i\}_{i=1}^n$ 独立同分布, 并且对于 $k \ge 3$ 满足模型 1.6

A7:$\theta^{\mathrm{T}}X$ 的密度函数在 Λ 上是非零有界的, 并在 Λ 上满足 2 阶的 Lipschitz 条件.

其中 $\Lambda = \{t = \theta^{\mathrm{T}}X : x \in A\}, A$ 是 X 的一个有界支撑集.

A8:$g(t)$ 在 Λ 有三阶的连续导数; $g_{1s}(t)$ 满足 2 阶的 Lipschitz 条件. 其中 $g_{1s}(t)$ 是 $g_1(t) = E(X|\theta^{\mathrm{T}}X = t)$ 的第 s 个分量.

A9: 核函数 $K(u)$ 是一个有界的概率密度函数, 且满足

$$\int_{-\infty}^{\infty} uK(u)du = 0, \quad \int_{-\infty}^{\infty} u^2K(u)du \ne 0, \quad \int_{-\infty}^{\infty} u^8K(u)du < \infty.$$

A10:$nh^2 \to \infty, nh^4 \to 0; nhh_1^3 \to \infty, \limsup_{n\to\infty} nh_1^5 < \infty$.

A11:$V(\theta^{(r)})$ 和 $V_0(\theta^{(r)})$ 是两个正定矩阵, 其中

$V(\theta^{(r)}) = E\{\widetilde{\varepsilon}^2 g'(\theta^{\mathrm{T}}X)^2 J_{\theta^{(r)}}^{\mathrm{T}}[X - E(X|\theta^{\mathrm{T}}X)][X - E(X|\theta^{\mathrm{T}}X)]^{\mathrm{T}}J_{\theta^{(r)}}\}$,

$V_0(\theta^{(r)}) = E[\widetilde{\varepsilon}^2 g'(\theta^{\mathrm{T}}X)^2 J_{\theta^{(r)}}^{\mathrm{T}} X X^{\mathrm{T}} J_{\theta^{(r)}}]$.

在条件 A1 中模型（2.7）用于表明 X_i 关于相关分量的尾概率行为, 这在高维的统计分析及现实数据中是不可忽略的; 条件 A3 满足时, 允许 p, d 都可以是发散的; 条件 A4 表明, 微弱的影响必须控制惩罚参数 τ, 这是为了保证影响能够恢复.A5 用于控制惩罚在非零元素上的影响. 对于 SCAD 惩罚函数, 因为在 n 充分大, 条件 A4 给定的条件下,$\max_{j\in\mathcal{L}} p'_\tau(|\theta_j|) = 0$.A5 是成立的（详细参见 Fan 和 Li）. 条件 A7 暗示着 $X^{\mathrm{T}}\theta$ 的密度函数是正的, 这个确保了 $\hat{g}, \hat{g}', \hat{g}''$ 非零有界的; 条件 A8 的需要, 是由于我们在局部线性逼近时要用到二阶核, 与 A9 结合一起, 确保了 $\hat{g}, \hat{g}', \hat{g}''$ 有高阶的收敛速度, 并且这个条件也比较容易实现.

§2.5 主要结果

在下面的定理中，我们将会给出惩罚经验似然估计值 $\hat{\theta}^{(r)} = (\hat{\theta}_1^{(r)\mathrm{T}}, \hat{\theta}_2^{(r)\mathrm{T}})$ 的性质.

定理 1 在正则条件 A1-A11 下, 当 $n \to \infty$ 时

(i) (稀疏性) 依概率 1 有,$\hat{\theta}_2^{(r)} = 0$;

(ii) (渐近正态性)$\sqrt{n} W_n I_A^{-1/2} (\theta_1^{(r)} - \theta_{10}^{(r)}) \xrightarrow{\mathcal{D}} N(0, G)$.

其中, 对于固定的 q,$W_n \in \mathcal{R}^{q \times d}, G \in \mathcal{R}^{q \times q}$, 满足 $W_n W_n^{\mathrm{T}} \to G$. 为了进行计数过程, 定义 $I_{p-1} = (H_1^{\mathrm{T}}, H_2^{\mathrm{T}})$, 其中 $H_1 \in \mathcal{R}^{(p-1-d) \times (p-1)}$, 且

$$I_A = H_1 \widetilde{\Sigma}^{-1} H_1^{\mathrm{T}} - H_1 \widetilde{\Sigma}^{-1} H_2^{\mathrm{T}} (H_2 \widetilde{\Sigma}^{-1} H_2^{\mathrm{T}})^{-1} H_2 \widetilde{\Sigma}^{-1} H_1^{\mathrm{T}}.$$

上述定理 1 中的 W_n 表示从一个维数发散的向量 $\theta_1^{(r)}$ 到一个固定维数 q 的投影. 其后, 投影向量的极限分布可以由一个多元的正态分布来描述. 由定理 1 可知, 当单指标模型中的参数个数发散时, 其稀疏性和渐近正态性仍然有效. 由此可见, 惩罚经验似然在模型选择与在真稀疏模型已知下的经验似然估计的效果是一样的. 又因为变量间的相关性被引入到公式中, 因此, 惩罚经验似然在估计非零参数时可提高其有效性.

经验似然方法不仅可以进行参数估计, 而且还可以利用 Wilks 定理在不用估计参数方差的情况下进行假设检验和构造置信区域. 同样, 对于惩罚经验似然方法我们也可以在不用估计选出参数的方差的情况下, 利用 Wilks 定理进行统计推断. 在统计推断中, 当分布未知时, 我们最常用的方法就是似然比检验. 然而对于当参数发散的惩罚经验似然情型, 似然比是否仍然有效？为了回答这个问题, 我们考虑如下的线性假设检验：

$$H_0 : L_n \theta_1^{(r)} = 0 \qquad vs \qquad H_1 : L_n \theta_1^{(r)} \neq 0.$$

其中 L_n 是一个 $q \times d$ 的矩阵, 对于固定的 q, 满足 $L_n L_n^{\mathrm{T}} = I_q$. 这个问题既包含了检验, 同时也包含有一些协变量的意义. 对于这个问题, 我们构建惩罚经验似然比统计量为

$$\widetilde{l}(L_n) = -2 \left\{ \hat{l}_p(\hat{\theta}^{(r)}) - \min_{\theta^{(r)} : L_n \theta_1^{(r)} = 0} \hat{l}_p(\theta^{(r)}) \right\}.$$

接下来, 我们对检验统计量的性质进行总结.

定理 2 在原假设 $H_0 : L_n \theta_1^{(r)} = 0$ 及定理 1 的正则条件下, 当 $n \to \infty$ 时, 有

$$\widetilde{l}(L_n) \xrightarrow{\mathcal{D}} \chi_q^2.$$

基于定理 2, 我们利用 $L_n \theta^{(r)}$ 来构建 $\theta^{(r)}$ 的 $1 - \alpha$ 水平置信域为

$$V_\alpha = \left[\theta^{(r)} : -2\{\hat{l}_p(\hat{\theta}^{(r)} - \min_{\theta^{(r)} : L_n \theta_1^{(r)} = 0} \hat{l}_p(\theta^{(r)})\} \right] \leq \chi_{q,1-\alpha}^2.$$

其中,$\chi^2_{q,1-\alpha}$ 是 χ^2 分布的 $1-\alpha$ 水平分位数. 从而, 当 $n\to\infty$ 时我们有

$$P(L_n\theta_1^{(r)}\in V_\alpha)\to 1-\alpha.$$

定理 2 表明, 一旦我们得到了 $\theta^{(r)}$ 的置信域就可以通过 $\theta_r=(1-||\theta^{(r)}||^2)^{1/2}$ 来构造 θ_r 的置信置信区间. 更重要的是, 这样不仅使 θ 的维数降低了, 而且如此构造域提高了其精度.

对于单指标回归模型, 定理 2 再一次给出了对假设检验和截面数据置信域构建一种方便方法, 对于高维的数据分析, 这个结果是令人欣慰. 结合定理 1 的惩罚经验似然的 Oracle 性质, 这两个定理共同表明了在高维数据分析中的稳健性和有效性.

§2.6 主要结果的证明

引理 1 若对于 $k\geq 1, m^{4k}<\infty$, 有

$$E(||X_i||^{2k})=O(p^k),\qquad Var(||X_i||^{2k})=O(p^{1-2k}).$$

证明 根据 Chen,Peng,Qin(2009)[35] 中的引理 1, 有

$$E(||X_i||^{2k})=O(tr^k(\Sigma)),\qquad Var(||X_i||^{2k})=O(tr^{1-2k}(\Sigma)).$$

由条件 A2 的 Σ 的特征根是有界的正数, 则

$$tr(\Sigma)=||\Sigma||=\sum_{j=1}^{p}\gamma_j(\Sigma)\leq Cp,$$

其中 C 是一正常数; 因此只需把 $tr(\Sigma)$ 改为 p, 即可得到引理 1 的证明.

引理 2 若 $k\geq 1, m^{4k}<\infty$, 则依概率 1 有下式成立

$$\max_{1\leq i\leq n}||X_i||=O_p(p^{1/2})+o_P\{p^{-\frac{2k-1}{4k}}n^{\frac{1}{4k}}\}.$$

证明
$$\begin{aligned}\max_{1\leq i\leq n}||X_i||&=\left\{\max_{1\leq i\leq n}||X_i||^{2k}\right\}^{1/2k}\\&=\left\{\max_{1\leq i\leq n}||X_i||^{2k}-E||X_i||^{2k}+E||X_i||^{2k}\right\}^{1/2k}\\&=\left\{\sqrt{Var(||X_i||^{2k})}\max_{1\leq i\leq n}|\tfrac{||X_i||^{2k}-E||X_i||^{2k}}{\sqrt{Var||X_i||^{2k}}}|+E||X_i||^{2k}\right\}^{1/2k}.\end{aligned}$$

由 Owen(1990)[38] 中引理 3, 当 $n\to\infty$ 时, 依概率 1 有

$$\max_{1\leq i\leq n}|\frac{||X_i||^{2k}-E||X_i||^{2k}}{\sqrt{Var||X_i||^{2k}}}|=o(n^{1/2}).$$

再由引理 1 的结论, 可得引理 2 的证明.

引理 3 [39] 若 $A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ 可逆, 当 $|A|^{-1}\neq 0$ 时, 有

$$A^{-1}=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1}=\begin{pmatrix} A_{11}^{-1}+A_{11}^{-1}A_{12}B^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}B^{-1} \\ -B^{-1}A_{21}A_{11}^{-1} & B^{-1} \end{pmatrix}$$

$$=\begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}+\begin{pmatrix} A_{11}^{-1}A_{12} \\ -I \end{pmatrix}B^{-1}(A_{21}A_{11}^{-1} \quad -I).$$

其中 $B=A_{22}-A_{21}A_{11}^{-1}A_{12}$.

引理 4 若 $a_n=\{(p-1)/n\}^{\frac{1}{2}-\delta}, D_n=\{\theta^{(r)}:||\theta^{(r)}-\theta_0^{(r)}||\leq ca_n\}$. 其中 $\delta,c>0$ 且 δ 满足$(p-1)^{1-\delta}/n^{\frac{1}{2}-\delta}\to 0$, 则当 $\theta^{(r)}\in D_n$ 时,$||\lambda_\theta||=O_p(a_n)$.

证明 记 $\lambda_\theta=\rho\beta$, 其中 $\rho\geq 0,\beta\in\mathcal{R}^{p-1},||\beta||=1,\rho=||\lambda_\theta||$. 首先, 我们来证明

$$\rho\left\{\beta^{\mathrm{T}}\hat{T}_n(\theta^{(r)})\beta-\max_{1\leq i\leq n}||\hat{U}_i(\theta^{(r)})||n^{-1}|\sum_{i=1}^{n}\beta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})|\right\}\leq n^{-1}|\sum_{i=1}^{n}\beta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})|. \tag{2.8}$$

其中 $\hat{T}_n(\theta^{(r)})=n^{-1}\sum\hat{U}_i(\theta^{(r)})\hat{U}_i^{\mathrm{T}}(\theta^{(r)})$.

由前面知,$\lambda_\theta\in\mathcal{R}^{p-1}$, 且满足

$$Q_{1n}(\theta^{(r)},\lambda_\theta)=\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{U}_i(\theta^{(r)})}{1+\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})}=0.$$

而对于 $\theta^{(r)}\in D_n$ 时, 与 Owen(1990)[38] 一样, 反复应用 Taylor 展开, 有

$$\begin{aligned}
0&=||Q_{1n}(\theta^{(r)},\lambda_\theta)||=||Q_{1n}(\rho\beta)||\\
&\geq|\beta^{\mathrm{T}}Q_{1n}(\rho\beta)|\\
&=n^{-1}|\beta^{\mathrm{T}}\{\sum_{i=1}^{n}\hat{U}_i(\theta^{(r)})-\rho\sum_{i=1}^{n}\frac{\hat{U}_i(\theta^{(r)})\hat{U}_i^{\mathrm{T}}(\theta^{(r)})\beta}{1+\rho\beta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})}\}|\\
&\geq\frac{\rho}{n}\beta^{\mathrm{T}}\sum_{i=1}^{n}\frac{\hat{U}_i(\theta^{(r)})\hat{U}_i^{\mathrm{T}}(\theta^{(r)})}{1+\rho\beta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})}\beta-n^{-1}|\sum_{i=1}^{n}\beta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})|\\
&\geq\rho\beta^{\mathrm{T}}\hat{T}_n(\theta^{(r)})\beta\{1+\rho\max_{1\leq i\leq n}||\hat{U}_i(\theta^{(r)})||\}^{-1}-n^{-1}|\sum_{i=1}\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})|\\
&\approx\rho\beta^{\mathrm{T}}\hat{T}_n(\theta^{(r)})\beta-\max_{1\leq i\leq n}||\hat{U}_i(\theta^{(r)})||\rho n^{-1}\sum_{i=1}\beta^{\mathrm{T}}\hat{T}_n(\theta^{(r)})\beta-n^{-1}|\sum_{i=1}\beta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})|\\
&\geq\rho\beta^{\mathrm{T}}\hat{T}_n(\theta^{(r)})\beta-\max_{1\leq i\leq n}||\hat{U}_i(\theta^{(r)})||\rho n^{-1}|\sum_{i=1}\beta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})|-n^{-1}|\sum_{i=1}\beta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})|.
\end{aligned}$$

因此有

$$\rho\left\{\beta^{\mathrm{T}}\hat{T}_n(\theta^{(r)})\beta-\max_{1\leq i\leq n}||\hat{U}_i(\theta^{(r)})||n^{-1}|\sum_{i=1}^{n}\beta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})|\right\}\leq n^{-1}|\sum_{i=1}^{n}\beta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})|.$$

注 下面的推导都是在真值 $\theta^{(r)}=\theta_0^{(r)}$ 下进行的.

一方面, 注意到 X_i 满足因子模型.ε_i 与 X_i 的相互独立性, 并满足条件 A1,A6,A7, 方便起见,

令

$$X_i = J_{\theta^{(r)}}^{\mathrm{T}} X_i,$$

$$\widetilde{\varepsilon}_i = Y_i - \hat{g}(\theta^{\mathrm{T}} X_i).$$

则 $\hat{U}_i(\theta_0^{(r)}) = \hat{g}'(\theta_0^{\mathrm{T}} X_i)\widetilde{X}_i\widetilde{\varepsilon}_i$. 由引理 2 可得

$$||n^{-1}\sum^{\cdot\cdot}\hat{U}_i(\theta_0^{(r)})|| = O_p\{(\frac{p-1}{n})^{1/2}\}.$$

因此有

$$|n^{-1}\sum_{i=1}^{\cdot\cdot}\theta^{\mathrm{T}}\hat{U}_i(\theta_0^{(r)})| \le ||n^{-1}\sum_{i=1}^{\cdot\cdot}\hat{U}_i(\theta_0^{(r)})|| = O_p\{(\frac{p-1}{n})^{1/2}\}. \tag{2.9}$$

根据薛留根的单指标模型的经验似然 [13] 知

$$\hat{T}_n(\theta_0^{(r)}) = n^{-1}\sum \hat{U}_i(\theta_0^{(r)})\hat{U}_i(\theta_0^{(r)})^{\mathrm{T}} \xrightarrow{\mathrm{P}} V_0(\theta_0^{(r)}).$$

其中 $V_0(\theta_0^{(r)}) = E[\varepsilon_i^2\hat{g}'(\theta_0^{\mathrm{T}} X_i)^2\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}]$, 显然有

$$\theta^{\mathrm{T}}\hat{T}_n(\theta_0^{(r)})\theta = O_p(1). \tag{2.10}$$

$$\max_{1\le i\le n}||\hat{U}_i(\theta^{(r)})|| = \max_i\{|\hat{g}'(\theta^{\mathrm{T}} X_i)^2|\}\max_i(\bar{X}_i\bar{X}_i^{\mathrm{T}}\widetilde{\varepsilon}_i^2)^{1/2}$$

$$= \max\{|\hat{g}'(\theta^{\mathrm{T}} X_i)^2|\}\max_i\{(\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}\widetilde{\varepsilon}_i^2)^k - m^{2k}E(\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}})^k + m^{2k}E(\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}})^k\}^{1/(2k)}$$

$$= \max_i\{|\hat{g}'(\theta^{\mathrm{T}} X_i)^2|\}\sqrt{Var\{(\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}\varepsilon_i^2)^k\}}\max_i\{\frac{(\widetilde{X}_i^{\mathrm{T}}\widetilde{X}_i^{\mathrm{T}}\varepsilon_i^2)^k - m^{2k}E(\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}})^k}{\sqrt{Var\{(\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}\varepsilon_i^2)^k\}}} + m^{2k}E(\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}})^k\}^{1/(2k)}$$

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

因此

$$\max_{1\le i\le n}||\hat{U}_i(\theta_0^{(r)})||n^{-1}|\sum_{i=1}\hat{U}_i(\theta_0^{(r)})| = O\{(p-1)n^{-1/2}\} + o_p\{(p-1)^{-\frac{1}{4k}}n^{\frac{1}{4k}-\frac{1}{2}}\} = o_p(1). \tag{2.11}$$

显然由 (2.9)-(2.11) 式代入 (2.8) 式有

$$||\lambda_{\theta_0}|| = \rho = O_p(\sqrt{(p-1)/n})$$

另一方面, 对于 $\theta^{(r)} \in D_n$（除 θ_0 以外）我们容易得出

$$n^{-1}|\sum_{i=1}\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})| = n^{-1}|\sum_{i=1}\theta^{\mathrm{T}}\hat{g}'(\theta^{\mathrm{T}} X_i)\widetilde{X}_i\widetilde{\varepsilon}_i| = n^{-1}|\sum_{i=1}\theta^{\mathrm{T}}\{\hat{U}_i(\theta_0^{(r)}) - K_1(\theta^{(r)} - \theta_0^{(r)})\} + o_p(1)|. \tag{2.12}$$

其中 $K_1 = \hat{g}'(\theta^{\mathrm{T}} X_i)^2\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}} - \widetilde{\varepsilon}_i\hat{g}''(\theta^{\mathrm{T}} X_i)\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}} + \widetilde{\varepsilon}_i\hat{g}'(\theta^{\mathrm{T}} X_i)\frac{(1-||\theta^{(r)}||^2)I_{p-1} - \theta^{(r)}\theta^{(r)\mathrm{T}}}{(1-||\theta^{(r)}||^2)^{3/2}}X_{ir}$.

又因为 $E[X_i]=0$. 则由引理 2 有

$$\max_i ||\widetilde{X}_i|| = O_p\{(p-1)^{1/2}\} + o_p\{(p-1)^{-\frac{2k-1}{4k}} n^{\frac{1}{4k}}\}.$$

而且, 注意到对于 $\theta^{(r)} \in D_n$,$(p-1)^{1-\theta}/n^{\frac{1}{2}-\theta} \to 0$ 时, 有

$$\max_i ||\hat{g}'(\theta_0^{\mathrm{T}} X_i)^2 \widetilde{X}_i \widetilde{X}_i^{\mathrm{T}} (\theta^{(r)} - \theta_0^{(r)})|| \le c a_n \max_i ||\widetilde{X}_i||^2 = o_p(1).$$

$$\begin{aligned}
\hat{T}_n(\theta^{(r)}) - \hat{T}_n(\theta_0^{(r)}) = & -\frac{2}{n}\ddot{\sum_{i=1}}\{\widetilde{\varepsilon}_i \hat{g}'(\theta^{\mathrm{T}} X_i)^3 \widetilde{X}_i \widetilde{X}_i^{\mathrm{T}} \widetilde{X}_i^{\mathrm{T}} + \widetilde{\varepsilon}_i^2 \hat{g}'(\theta^{\mathrm{T}} X_i) \hat{g}''(\theta^{\mathrm{T}} X_i) \widetilde{X}_i \widetilde{X}_i^{\mathrm{T}} \widetilde{X}_i^{\mathrm{T}} \\
& -\widetilde{\varepsilon}_i^2 \hat{g}'(\theta^{\mathrm{T}} X_i)^2 \tfrac{(1-||\theta^{(r)}||^2) I_{p-1} + \theta^{(r)} \theta^{(r)\mathrm{T}}}{(1-||\theta^{(r)}||^2)^{3/2}} X_{ir} \widetilde{X}_i\}(\theta^{(r)} - \theta_0^{(r)}) + o_p(1) \\
= & -\frac{2}{n}\ddot{\sum}\{\hat{g}'(\theta^{\mathrm{T}} X_i)^3 \widetilde{X}_i \widetilde{X}_i^{\mathrm{T}} - \widetilde{\varepsilon}_i \hat{g}'(\theta^{\mathrm{T}} X_i) \hat{g}''(\theta^{\mathrm{T}} X_i) \widetilde{X}_i \widetilde{X}_i^{\mathrm{T}} \\
& -\widetilde{\varepsilon}_i \hat{g}'(\theta^{\mathrm{T}} X_i)^2 \tfrac{(1-||\theta^{(r)}||^2) I_{p-1} + \theta^{(r)} \theta^{(r)\mathrm{T}}}{(1-||\theta^{(r)}||^2)^{3/2}} X_{ir}\}\{\widetilde{X}_i^{\mathrm{T}}(\theta^{(r)} - \theta_0^{(r)})\}\widetilde{\varepsilon}_i + o_p(1).
\end{aligned}$$

由条件 A7 得 $\hat{g}, \hat{g}', \hat{g}''$ 是连续、可导、有界的函数. 可得到

$$\beta^{\mathrm{T}} \hat{T}_n(\theta^{(r)}) \beta = \beta^{\mathrm{T}} \hat{T}_n(\theta_0^{(r)}) \beta + o_p(1) = O_p(1). \tag{2.13}$$

因此, 由条件 A4 得到

$$\max_i ||\hat{U}_i(\theta^{(r)})|| n^{-1} |\sum \beta^{\mathrm{T}} \hat{U}_i(\theta^{(r)})| = o_p(1). \tag{2.14}$$

联合 (2.10)-(2.14), 引理 4 得证.

因此, 由上面的结论, 可得出如下结论:

$$\max_i |\lambda_\theta^{\mathrm{T}} \hat{U}_i(\theta^{(r)})| = o_p(1), \tag{2.15}$$

$$\max_i |\lambda_\theta^{\mathrm{T}} X_i| = o_p(1). \tag{2.16}$$

引理 5 当 $n \to \infty$,$l_p(\theta^{(r)})$ 在 D_n 中以概率 1 有最小值.

证明 对于 $\theta^{(r)} \in D_n$, 由经验似然的定义,Taylor 展开容易得到

$$0 = Q_{1n}(\theta^{(r)}, \lambda_\theta) = n^{-1} \sum_{i=1}^n \frac{\hat{U}_i(\theta^{(r)})}{1 + \lambda_\theta^{\mathrm{T}} \hat{U}_i(\theta^{(r)})} = n^{-1} \sum_{i=1}^n \hat{U}_i(\theta^{(r)}) - \hat{T}_n(\theta^{(r)}) \lambda_\theta + r_n. \tag{2.17}$$

其中的余项

$$r_n = n^{-1} \sum \hat{U}_i(\theta^{(r)}) \{\lambda_\theta^{\mathrm{T}} \hat{U}_i(\theta^{(r)})\}^2 (1+\xi_i)^{-3}.$$

其中 $|\xi_i| \le |\lambda_\theta^{\mathrm{T}} \hat{U}_i(\theta^{(r)})|$

由 (2.17) 得到 λ_θ 的表达式:

$$\lambda_\theta = \hat{T}_n^{-1}(\theta^{(r)}) \bar{U}(\theta^{(r)}) + \hat{T}_n^{-1}(\theta^{(r)}) r_n. \tag{2.18}$$

其中 $\bar{U}(\theta^{(r)}) = \frac{1}{n}\sum \hat{U}_i(\theta^{(r)})$.

又因为

$$2\hat{l}(\theta^{(r)}) = 2\sum \log\{1+\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})\}$$

$$= 2\sum_{i=1}\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)}) - \sum_{=1}\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})\hat{U}_i^{\mathrm{T}}(\theta^{(r)})\lambda_\theta^{\mathrm{T}} + \tfrac{2}{3}\sum_{=1}\{\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})\}^3(1+\xi_i)^{-4}.$$

将 λ_θ 代入 $\hat{l}(\theta^{(r)})$, 整理得到

$$2\hat{l}(\theta^{(r)}) = n\bar{U}(\theta^{(r)})\hat{T}_n^{-1}(\theta^{(r)})\bar{U}(\theta^{(r)}) - nr_n^{\mathrm{T}}\hat{T}^{-1}(\theta^{(r)})r_n + \frac{2}{3}\sum_{i=1}^{\cdot\cdot}\{\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})\}^3(1+\xi_i)^{-4}. \quad (2.19)$$

注意到, 当 $\theta^{(r)} \in \partial D_n$. 其中 ∂D_n 表示 D_n 的边界, 对于 $||\mathbf{u}|| = 1$, 则有

$$\theta^{(r)} = \theta_0^{(r)} + ca_n\mathbf{u}.$$

对 $2\hat{l}(\theta^{(r)})$ 式在 $\theta_0^{(r)}$ 处进行 Taylor 展开, 我们有

$$2\hat{l}(\theta^{(r)}) - 2\hat{l}(\theta_0^{(r)}) = 2\frac{\partial\hat{l}(\theta_0^{(r)})}{\partial\theta^{(r)\mathrm{T}}}ca_n\mathbf{u} + c^2a_n^2\mathbf{u}^{\mathrm{T}}\frac{\partial^2\hat{l}(\theta_0^{(r)})}{\partial\theta^{(r)}\partial\theta^{(r)\mathrm{T}}}\mathbf{u} + O(||\theta^{(r)} - \theta_0^{(r)}||^2).$$

且当 $\theta_0^{(r)}$ 是真值时,$E(\widetilde{\varepsilon}_i) = 0, E(\widetilde{X}_i) = 0$, 由 Qin 和 Lawless(1995)[40] 中引理 2 的证明有

$$\frac{\partial\hat{l}(\theta_0^{(r)})}{\partial\theta^{(r)\mathrm{T}}}ca_n\mathbf{u} = ca_n\mathbf{u}(\sum_{i=1}^{n}\frac{\hat{U}_i(\theta_o^{(r)})}{1+\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})}\frac{\partial\lambda_\theta^{\mathrm{T}}}{\partial\theta^{(r)\mathrm{T}}} + \sum_{i=1}^{n}\frac{\lambda_\theta^{\mathrm{T}}}{1+\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)\mathrm{T}})}\frac{\partial\hat{U}_i(\theta^{(r)})}{\partial\theta^{(r)\mathrm{T}}}) = cna_n\mathbf{u}\lambda_\theta^{\mathrm{T}}S_n^{\mathrm{T}}(\theta_0^{(r)})$$

$$= O_p(na_n^2).$$

其中 $S_n^{\mathrm{T}}(\theta_0^{(r)}) = n^{-1}\sum_{i=1}^{\cdot\cdot}\frac{1}{1+\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)\mathrm{T}})}\frac{\partial\hat{U}_i(\theta^{(r)})}{\partial\theta^{(r)\mathrm{T}}}$.

$$c^2a_n^2\mathbf{u}^{\mathrm{T}}\frac{\partial^2\hat{l}(\theta_0^{(r)})}{\partial\theta^{(r)}\partial\theta^{(r)\mathrm{T}}}\mathbf{u} = c^2a_n^2\mathbf{u}^{\mathrm{T}}\frac{\partial n\lambda_\theta^{\mathrm{T}}S_n^{\mathrm{T}}(\theta_0^{(r)})}{\partial\theta^{(r)}}\mathbf{u}$$

$$= c^2na_n^2\mathbf{u}^{\mathrm{T}}\left\{\frac{\partial\lambda_\theta^{\mathrm{T}}}{\partial\theta^{(r)}}S_n^{\mathrm{T}}(\theta_0^{(r)}) + \lambda_\theta^{\mathrm{T}}\frac{\partial S_n^{\mathrm{T}}(\theta_0^{(r)})}{\partial\theta^{(r)}}\right\}\mathbf{u}$$

$$= c^2na_n^2\left\{n^{-1}\sum_{i=1}^{n}\hat{g}'(\theta_0^{\mathrm{T}}X_i)^2\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}\mathbf{u}\right\}^{\mathrm{T}}\hat{T}_n^{-1}(\theta_0^{(r)})\{n^{-1}\sum_{i=1}^{n}\hat{g}'(\theta_0^{\mathrm{T}}X_i)^2\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}\mathbf{u}\right\}$$

$$= O_p(na_n^2).$$

因此

$$2\hat{l}(\theta^{(r)}) - 2\hat{l}(\theta_0^{(r)}) = O_p(na_n^2), \quad \text{且} \quad na_n^2 \longrightarrow \infty.$$

这意味着, 对于任意给定的正数 c, 当 $n \to \infty$ 时

$$P\left[\{2\hat{l}(\theta^{(r)}) - 2\hat{l}(\theta_0^{(r)})\} > c\right] \longrightarrow 1.$$

又因为

$$n\sum_{j\neq r}\{p_\tau(|\theta_j|) - p_\tau(|\theta_{0j}|)\} = \sum_{j\neq r}[cna_np_\tau'(|\theta_{0j}|)sign(\theta_{0j})u_j + c^2na_n^2p_\tau''(|\theta_{0j}|)u_j^2\{1+o(1)\}]$$

$$= T_1 + T_2.$$

当 n 足够大时, 由条件 A5 及引理 4 的条件有

$$T_1 \leq \sum_{j\neq r}^{p} |cna_n p'_\tau(|\theta_{0j}|)sign(\theta_{0j})u_j| \leq cna_n\sqrt{p-1}|\max_{j\in\Im} p'_\tau(|\theta_{0j}|)|||\mathbf{u}|| = o_p(1),$$

$$T_2 \leq 2\cdot \max_{j\in\Im} p''_\tau(|\theta_{0j}|)\cdot nc^2a_n^2||\mathbf{u}|| = o_p(1).$$

因此有

$$\begin{aligned}\hat{l}_p(\theta^{(r)}) - \hat{l}_p(\theta_0^{(r)}) &= \hat{l}(\theta^{(r)}) - \hat{l}(\theta_0^{(r)}) + n\sum_{j\neq r}^{p}\{p_\tau(|\theta_j|) - p_\tau(|\theta_{0j}|)\}\\ &\geq \hat{l}(\theta^{(r)}) - \hat{l}(\theta_0^{(r)}) + n\sum_{j\in\mathcal{L}}\{p_\tau(|\theta_j|) - p_\tau(|\theta_{0j}|)\}\\ &\geq \hat{l}(\theta^{(r)}) - \hat{l}(\theta_0^{(r)}).\end{aligned}$$

或者由 SCAD 惩罚函数的无偏性质, 当 n 充分大时, 对于 $j\in\mathcal{L}$, 有 $p_\tau(|\theta_j|) = p_\tau(|\theta_{0j}|)$. 可知最后一个不等式成立.

因此, 对于 $\theta^{(r)}\in\partial D_n$, 有

$$P\left[\hat{l}_p(\theta^{(r)}) - \hat{l}_p(\theta_0^{(r)}) > c\right] \longrightarrow 1.$$

引理 5 得证.

定理 1 的证明

由引理 5 可知, 在球 $\{\theta^{(r)} : ||\theta^{(r)} - \theta_0^{(r)}|| \leq ca_n\}$ 内存在 $\hat{l}_p(\theta^{(r)})$ 的最小值. 由于 $\theta^{(r)}\in D_n$. 我们有

$$\begin{aligned}\frac{1}{n}\frac{\partial\hat{l}_p(\theta^{(r)})}{\partial\theta_j} &= n^{-1}\sum_{i=1}^{n}\frac{-\lambda_\theta^{\mathrm{T}}\hat{g}(\theta^{\mathrm{T}}X_i)\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}\gamma_j + \widetilde{\varepsilon}_i\hat{g}''(\theta^{\mathrm{T}}X_i)\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}\gamma_j - \widetilde{\varepsilon}_i\hat{g}'(\theta^{\mathrm{T}}X_i)\frac{(1-||\theta^{(r)}||^2)I_{p-1}+\theta^{(r)}\theta^{(r)\mathrm{T}}}{(1-||\theta^{(r)}||^2)^{3/2}}X_{ir}\gamma_j}{\{1+\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})\}^2}\\ &\quad + p'_\tau(|\theta_j|)sign(\theta_j)\\ &= I_j + II_j.\end{aligned}$$

通过选择充分大的 n, 由 (2.15)、(2.16) 式及条件 A7 有

$$\begin{aligned}\max_j|I_j| &\leq \max_i|\lambda_\theta^{\mathrm{T}}\widetilde{X}_i|\max_j|n^{-1}\sum_{i=1}^{n}\hat{g}'(\theta^{\mathrm{T}}X_i)^2\widetilde{X}_i^{\mathrm{T}}\gamma_j|\\ &\leq \max_i|\lambda_\theta^{\mathrm{T}}\widetilde{X}_i||n^{-1}\sum_{i=1}^{n}\widetilde{X}_i||\hat{g}'(\theta^{\mathrm{T}}X_i)^2|\\ &= o_p(1)O_p\{(\tfrac{p-1}{n})^{1/2}\} = o_p\{(\tfrac{p-1}{n})^{1/2}\}.\end{aligned}$$

在 $p_\tau(\cdot)$ 上由条件 A4 中的 $\tau(n/(p-1))^{1/2-\theta}\to\infty$, 这意味着 $((p-1)/n)^{1/2-\theta}/\tau\to 0$, 即

$$P\{\max_{j\in\mathcal{L}}|I_j| > \tau/2\} \longrightarrow 0.$$

此外, 由引理 4 的条件, 注意到

$$|\theta_j| \leq ca_n, j\notin\mathcal{L} \quad 和 \quad \tau/a_n\to\infty.$$

因此, 在 $j \notin \mathcal{L}$, 且 n 充分大时, 由 $p'_\tau(\cdot)$ 的表达式有

$$p'_\tau(|\theta_j|) = \tau.$$

所以有

$$\frac{1}{n}\frac{\partial \hat{l}_p(\theta^{(r)})}{\partial \theta_j} = \tau\left\{o_p(\sqrt{\frac{p-1}{n}}/\tau) + sign(\theta_j)\right\}.$$

容易得知,θ_j 的符号完全控制着 $\frac{1}{n}\frac{\partial \hat{l}_p(\theta^{(r)})}{\partial \theta_j}$ 的符号, 即对于 $j \notin \mathcal{L}$ 时,θ_j 的符号完全控制着 $\partial \hat{l}_p/\partial \theta_j$ 的渐近性. 当 $n \to \infty$, 对任意的 $j \notin \mathcal{L}$, 以概率 1 有,

$$\frac{1}{n}\frac{\partial \hat{l}_p(\theta^{(r)})}{\partial \theta_j} > 0, \quad \theta_j \in (0, ca_n), \quad \frac{1}{n}\frac{\partial \hat{l}_p(\theta^{(r)})}{\partial \theta_j} < 0, \quad \theta_j \in (-ca_n, 0).$$

因此, 由函数的连续性质知, 依概率 1 有 $\hat{\theta}_2 = 0$. 综上所述, 我们得到了定理 1 的第一部分.

接下来, 我们将对定理的第二部分进行证明. 由第一部分的结论和惩罚经验似然的定义, 基于惩罚经验似然的估计值 $\hat{\theta}^{(r)}$ 是式（2.6）在 $H_2\theta^{(r)} = 0$ 下的最小值, 且 $\hat{\theta}^{(r)}$ 有 $(\hat{\theta}_1^{(r)}, 0)^T$ 这种形式. 由 Qin and Lawless（1995）[40] 及 Lagrange 乘子法, 得到的估计值与下面的新的目标函数 (2.20) 的最小值是等价的.

$$\widetilde{l}(\theta^{(r)}, \lambda, \nu) = n^{-1}\sum_{i=1}^{n}\log\{1 + \lambda_\lambda^{\mathrm{T}}\hat{U}_i(\theta^{(r)})\} + \sum_{j\neq r}^{d} p_\tau(|\theta_j|) + \nu^{\mathrm{T}}H_2\theta^{(r)}. \tag{2.20}$$

其中 $\nu \in R^{p-d}$ 是另一个 Lagrange 乘子向量.

定义

$$\widetilde{Q}_{1n}(\theta^{(r)}, \lambda, \nu) = n^{-1}\sum_{i=1}^{n}\{1 + \lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})\}^{-1}\hat{U}_i(\theta^{(r)}),$$

$$\widetilde{Q}_{2n}(\theta^{(r)}, \lambda, \nu) = -n^{-1}\sum_{i=1}^{n}\frac{\hat{g}'(\theta^{\mathrm{T}}X_i)^2 J_{\theta^{(r)}}^{\mathrm{T}}X_iX_i^{\mathrm{T}}J_{\theta^{(r)}}\lambda}{1 + \lambda^{\mathrm{T}}\hat{U}_i(\theta^{(r)})} + b(\theta^{(r)}) + H_2^{\mathrm{T}}\nu,$$

$$\widetilde{Q}_{3n}(\theta^{(r)}, \lambda, \nu) = H_2\theta^{(r)}.$$

其中

$$b(\theta^{(r)}) = \{p'_\tau(|\theta_1|)sgn(\theta_1), \cdots, p'(|\theta_{r-1}|)sgn(\theta_{r-1}), p'(|\theta_{r+1}|)sgn(\theta_{r+1}), \cdots, p'(|\theta_d|)sgn(\theta_d), 0^{\mathrm{T}}\}^{\mathrm{T}}.$$

不妨考虑 (2.19) 式最小值的解 $(\hat{\theta}^{(r)}, \hat{\lambda}, \hat{\nu})$ 满足

$$\widetilde{Q}_{jn}(\hat{\theta}^{(r)}, \hat{\lambda}, \hat{\nu}) = 0, \qquad j = 1, 2, 3.$$

由引理 4 结论知 $||\hat{\lambda}|| = O_p(a_n)$ 是随机任意小的, 且由引理 5 有 $||\hat{\theta}^{(r)} - \theta_0^{(r)}|| = O_p(a_n)$. 因此, 类似于 Qin and Lawless(1995)[41], 由 $0 = \widetilde{Q}_{2n}(\hat{\theta}^{(r)}, \hat{\lambda}, \hat{\nu})$, 我们有 $||\hat{\nu}|| = O_p(a_n)$. 因此

$\widetilde{Q}_{jn}(\hat{\theta}^{(r)},\hat{\lambda},\hat{\nu})$,

$(j=1,2,3)$ 在 $(\theta_0^{(r)},0,0)$ 处展开有

$$\begin{aligned}0&=\widetilde{Q}_{jn}(\hat{\theta}^{(r)},\hat{\lambda},\hat{\nu})\\&=\widetilde{Q}_{jn}(\theta_0^{(r)},0,0)+\frac{\partial Q_{jn}(\theta_0^{(r)},0,0)}{\partial\theta^{(r)\mathrm{T}}}(\hat{\theta}^{(r)}-\theta_0^{(r)})+\frac{\partial Q_{jn}(\theta_0^{(r)},0,0)}{\partial\lambda^{\mathrm{T}}}(\hat{\lambda}-0)\\&\quad+\frac{\partial Q_{jn}(\theta_0^{(r)},0,0)}{\partial\nu^{\mathrm{T}}}(\hat{\nu}-0)+o_p(\theta_n).\end{aligned}$$

其中 $\theta_n=||\hat{\theta}^{(r)}-\theta_0^{(r)}||+||\hat{\lambda}||+||\hat{\nu}||, j=1,2,3.$ 得到随机展式为

$$\begin{pmatrix}-\widetilde{Q}_{1n}(\theta_0,0,0)\\-b(\theta_0^{(r)})\\0\end{pmatrix}=\begin{pmatrix}-\widetilde{\Sigma}&-\widetilde{\Sigma}&0\\-\widetilde{\Sigma}&b'(\theta^{(r)})&H_2^{\mathrm{T}}\\0&H_2&0\end{pmatrix}_{(\theta_0^{(r)},0,0)}\begin{pmatrix}\hat{\lambda}-0\\\hat{\theta}^{(r)}-\theta_0^{(r)}\\\hat{\nu}-0\end{pmatrix}+R_n.$$

其中

$$-\widetilde{\Sigma}=\hat{g}'(\theta^{\mathrm{T}}X_i)^2J_{\theta^{(r)}}^{\mathrm{T}}\Sigma J_{\theta^{(r)}},$$

$$b'(\theta^{(r)})=diag\{p_\tau''(|\theta_1|),p_\tau''(|\theta_2|),\cdots,p''(|\theta_{r-1}|),p''(|\theta_{r+1}|),\cdots,p''(|\theta_p|)\}^{\mathrm{T}}.$$

余项为

$$R_n=\sum_{i=1}^{5}R_n^{(k)},R_n^{(1)}=(R_{1n}^{(1)\mathrm{T}},R_{2n}^{(1)\mathrm{T}},0)^{\mathrm{T}},R_{1n}^{(1)}\in R^p,R_{2n}^{(1)}\in R^p,$$

而 $R_{jn}^{(1)},j=1,2$ 的第 k 个分量

$$R_{jn,k}^{(1)}=\frac{(\hat{\eta}-\eta_0)^{\mathrm{T}}\{\partial^2\widetilde{Q}_{jn,k}(\eta^*)/\partial\eta\partial\eta^{\mathrm{T}}\}(\hat{\eta}-\eta_0)}{2}.$$

$\eta=(\theta^{(r)},\lambda)^{\mathrm{T}},\eta^*=(\theta^{*(r)},\lambda^*)$, 满足 $||\theta^{*(r)}-\theta_0^{(r)}||\le||\theta^{(r)}-\theta_0^{(r)}||,||\lambda^*||\le||\lambda||$

$$R_n^{(2)}=\{0,b^{\mathrm{T}}(\theta_0^{(r)}),0\}^{\mathrm{T}},R_n^{(3)}=[0,\{b'(\theta_0^{(r)})(\hat{\theta}^{(r)}-\theta_0^{(r)})\}^{\mathrm{T}},0]^{\mathrm{T}},$$

$$R_n^{(4)}=[\{(\hat{T}_n(\theta_0^{(r)})-\widetilde{\Sigma})\hat{\lambda}\}^{\mathrm{T}},0,0]^{\mathrm{T}},R_n^{(5)}=[\{(\hat{T}(\theta_0^{(r)})-\widetilde{\Sigma})\hat{\lambda}\}^{\mathrm{T}},\{(\hat{T}_n(\theta_0^{(r)})-\widetilde{\Sigma})(\hat{\theta}^{(r)}-\theta_0^{(r)})\}^{\mathrm{T}},0]^{\mathrm{T}}.$$

对于 $k=1,\cdots,4.$ 可证得

$$R_n^{(k)}=o_p(n^{-1/2}).$$

因此有

$$o_p(\theta_n)=||R_n||=||\sum_{k=1}^{5}R_n^{(k)}||\le\sum_{k=1}^{5}||R_n^{(k)}||=o_p(n^{-1/2}).$$

当 $||\hat{T}_n(\theta_0^{(r)})-\widetilde{\Sigma}||=O_p\{(\frac{p-1}{n})^{1/2}\},||\hat{\theta}^{(r)}-\theta_0^{(r)}||=O_p(a_n),||\hat{\lambda}||=O_p(a_n)$ 时, 也可得到

$$||R_n^{(5)}||=o_p(n^{-1/2}).$$

由条件 A4、A5, 在 $n \to \infty, \tau \to 0.$, 于是有

$$P(|b(\theta_0)| = 0) \to 0, \qquad P(b'(\theta_0) = \mathbf{0}) \to 0.$$

因此, 随机展式可写为

$$\begin{pmatrix} -\widetilde{Q}_{1n}(\theta_0,0,0) + o_p(\theta_n) \\ o_p(\theta_n)) \\ o_p(\theta_n) \end{pmatrix} = \begin{pmatrix} -\widetilde{\Sigma} & -\widetilde{\Sigma} & 0 \\ -\widetilde{\Sigma} & 0 & H_2^{\mathrm{T}} \\ 0 & H_2 & 0 \end{pmatrix}_{(\theta_0,0,0)} \begin{pmatrix} \hat{\lambda} - 0 \\ \hat{\theta}^{(r)} - \theta_0^{(r)} \\ \hat{\nu} - 0 \end{pmatrix}.$$

定义

$$A_{11} = -\widetilde{\Sigma}, \qquad A_{12} = (-\widetilde{\Sigma}, \quad 0), \qquad A_{21} = A_{12}^{\mathrm{T}}$$

$$A_{22} = \begin{pmatrix} 0 & H_2^{\mathrm{T}} \\ H^2 & 0 \end{pmatrix}, A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad \xi = \begin{pmatrix} \hat{\theta}^{(r)} - \theta_0^{(r)} \\ \hat{\nu} \end{pmatrix}.$$

则有

$$\begin{pmatrix} \hat{\lambda} \\ \xi \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} \left\{ \begin{pmatrix} -\widetilde{Q}_{1n}(\theta_0^{(r)},0,0) \\ \mathbf{0} \end{pmatrix} + R_n \right\}. \tag{2.21}$$

由引理 3 的分块矩阵求逆得到

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}B^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}B^{-1} \\ -B^{-1}A_{21}A_{11}^{-1} & B^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} A_{11}^{-1}A_{12} \\ -I \end{pmatrix} B^{-1} \begin{pmatrix} A_{21}A_{11}^{-1} & -I \end{pmatrix}.$$

其中 $B = A_{22} - A_{21}A_{11}^{-1}A_{12} = \begin{pmatrix} \widetilde{\Sigma} & H_2^{\mathrm{T}} \\ H_2 & 0 \end{pmatrix}$. 同理可得

$$B^{-1} = \begin{pmatrix} \widetilde{\Sigma}^{-1} - \widetilde{\Sigma}^{-1}H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2\widetilde{\Sigma}^{-1} & \widetilde{\Sigma}^{-1}H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1} \\ (H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2\widetilde{\Sigma}^{-1} & -(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1} \end{pmatrix}.$$

由 (2.21) 式可得

$$\xi = B^{-1}A_{21}A_{11}^{-1}\widetilde{Q}_{1n}(\theta_0^{(r)},0,0) + o_p(\theta_n).$$

这暗示着

$$\hat{\theta}^{(r)} - \theta_0^{(r)} = \{\widetilde{\Sigma}^{-1} - \widetilde{\Sigma}^{-1}H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2\widetilde{\Sigma}^{-1}\}\{\widetilde{Q}_{1n}(\theta_0^{(r)},0,0) + A^{-1}R_n\}.$$

并且有 $\bar{U}(\theta_0^{(r)}) = \widetilde{Q}_{1n}(\theta_0^{(r)},0,0), ||A^{-1}R_n|| = o_p(n^{-1/2})$.

显然,$\theta^{(r)}$ 中的非零元素可扩展为

$$\hat{\theta}_1^{(r)} - \theta_{10}^{(r)} = \{H_1\widetilde{\Sigma}^{-1} - H_1\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2\widetilde{\Sigma}^{-1}\}\{\bar{U}(\theta_0^{(r)}) + o_p(n^{-1/2})\}.$$

因此,$n^{1/2}\hat{\theta}_1^{(r)} = n^{1/2}H_1\hat{\theta}^{(r)}$ 的渐近协方差为

$$I_A = H_1\widetilde{\Sigma}^{-1}H_1^{\mathrm{T}} - H_1\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2\widetilde{\Sigma}^{-1}H_1^{\mathrm{T}},$$

令

$$Y_{ni} = n^{-1/2}Z_{ni}.$$

其中

$$Z_{ni} = W_n I_A^{-1/2}\{H_1\widetilde{\Sigma}^{-1} - H_1\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2\widetilde{\Sigma}^{-1}\}\hat{U}_i(\theta_0^{(r)}).$$

由条件 A2 及 $W_nW_n^{\mathrm{T}} \to G$, 我们有

$$P(||Y_{ni}|| \geq \varepsilon) \leq n^{-1}\varepsilon^{-2}E||Z_{ni}||^2 = O(1/n),$$

$$\begin{aligned} E||Y_{ni}||^4 = n^{-2}E(Z_{ni}Z_{ni}^{\mathrm{T}})^2 &\leq n^{-2}\lambda_{\max}(W_nW_n^{\mathrm{T}})\lambda_{\max}(I_A)E||(\hat{U}_i(\theta_0^{(r)})\hat{U}_i^{\mathrm{T}}(\theta_0^{(r)})||^2 \\ &= O((p-1)^2/n^2), \end{aligned}$$

$$||n^{1/2}W_nI_A^{-1/2}H_1A^{-1}R_n|| \leq n^{1/2}||A^{-1}R_n|| = o_p(1).$$

因此, 对于任意给定的 ε, 有

$$\begin{aligned} \sum_{i=1}^{n} E||Y_{ni}||^2 I(||Y_{ni}|| > \varepsilon) &= nE||Y_{n1}||^2 I\{||Y_{n1}|| > \varepsilon\} \\ &\leq n\{E||Y_{n1}||^4\}^{1/2}\{P(||Y_{n1}|| > \varepsilon)\}^{1/2} \to 0. \end{aligned}$$

另一方面, 当 $W_nW_n^{\mathrm{T}} \to G$ 时, 我们有

$$\begin{aligned} \sum_{i=1}^{n} Var(Y_{ni}) &= n\,Var(Y_{n1}) \\ &= Var(W_nI_A^{-1/2}\{H_1\widetilde{\Sigma}^{-1} - H_1\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2\widetilde{\Sigma}^{-1}\}\hat{U}_1(\theta_0^{(r)})) \\ &\longrightarrow G. \end{aligned}$$

综合以上结果知,Y_{ni} 满足 Lindeberg-Feller 中心极限定理的条件. 因此, 我们有

$$n^{1/2}W_nI_A^{-1/2}(\hat{\theta}_1^{(r)} - \theta_{10}^{(r)}) \xrightarrow{\mathcal{D}} N(0, G).$$

同时也说明了 Y_{ni} 服从多元的渐近正态分布.

定理 2 的证明

在原假设 $H_0 : L_n\theta_1^{(r)} = 0$ 下, 因为 $L_nL_n^{\mathrm{T}} = I_q$, 则存在 $\widetilde{H}_2$ 使得 $\widetilde{H}_2\theta^{(r)} = 0$ 和 $\widetilde{H}_2\widetilde{H}_2^{\mathrm{T}} =$

I_{p-d+q}. 下面重复定理 1 的证明过程, 我们在原假设条件下, 通过建立 $\theta^{(r)}$ 的估计量, 可得到下面式的最小化.

$$\hat{l}_p(\theta^{(r)},\lambda,\nu)=\sum_{i=1}\log\{1+\lambda_\theta^{\mathrm{T}}\hat{U}_i(\theta^{(r)})\}+n\sum_{j\neq r}p_\tau(|\theta_j|)+\nu^{\mathrm{T}}\widetilde{H}_2\theta^{(r)}. \tag{2.22}$$

定义 $(\theta^{(r)},\check{\lambda},\check{\nu})$ 是使上式达到最小的值. 由定理 1 的第一部分证明知, 以概率 1 有 $\check{\theta}_2^{(r)}=0$. 因此, 在 n 充分大时, 以概率 1 有

$$n\left\{\sum_{j\neq r}^{p}p_\tau(|\hat{\theta}_j^{(r)}|)-\sum_{j\neq r}^{p}p_\tau(|\check{\theta}_j^{(r)}|)\right\}=0.$$

类似于定理 1 有

$$\check{\theta}^{(r)}-\theta_0^{(r)}=\{\widetilde{\Sigma}^{-1}-\widetilde{\Sigma}^{-1}\widetilde{H}_2^{\mathrm{T}}(\widetilde{H}_2\widetilde{\Sigma}^{-1}\widetilde{H}_2^{\mathrm{T}})^{-1}\widetilde{H}_2\widetilde{\Sigma}^{-1}\}\{\bar{U}(\theta_0^{(r)})+o_p(n^{-1/2})\}.$$

因此

$$\begin{aligned}\hat{\theta}^{(r)}-\check{\theta}^{(r)}&=-\widetilde{\Sigma}^{-1}(H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2-\widetilde{H}_2^{\mathrm{T}}(\widetilde{H}_2\widetilde{\Sigma}^{-1}\widetilde{H}_2^{\mathrm{T}})^{-1}\widetilde{H}_2)\widetilde{\Sigma}^{-1}\bar{U}(\theta_0^{(r)})\\&=-\widetilde{\Sigma}^{-1/2}\{\widetilde{\Sigma}^{-1/2}(H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2-\widetilde{H}_2^{\mathrm{T}}(\widetilde{H}_2\widetilde{\Sigma}^{-1}\widetilde{H}_2^{\mathrm{T}})^{-1}\widetilde{H}_2)\widetilde{\Sigma}^{-1/2}\}\widetilde{\Sigma}^{-1/2}\bar{U}(\theta_0^{(r)}).\end{aligned}$$

容易验证 $\widetilde{\Sigma}^{-1/2}(H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2-\widetilde{H}_2^{\mathrm{T}}(\widetilde{H}_2\widetilde{\Sigma}^{-1}\widetilde{H}_2^{\mathrm{T}})^{-1}\widetilde{H}_2)\widetilde{\Sigma}^{-1/2}$ 是一个秩为 q 的幂等阵. $\hat{l}_p(\hat{\theta}^{(r)})-\hat{l}_p(\check{\theta}^{(r)})$ 在 $\hat{\theta}^{(r)}$ 进行 Taylor 展开有

$$\hat{l}_p(\hat{\theta}^{(r)})-\hat{l}_p(\check{\theta}^{(r)})=T_1+T_2+T_3+T_4.$$

其中

$$\begin{aligned}T_1&=\nabla^T\hat{l}_p(\hat{\theta}^{(r)})(\hat{\theta}^{(r)}-\check{\theta}^{(r)}),\\T_2&=-\tfrac{1}{2}(\hat{\theta}^{(r)}-\check{\theta}^{(r)})^{\mathrm{T}}\nabla^2\hat{l}_p(\hat{\theta}^{(r)})(\hat{\theta}^{(r)}-\check{\theta}^{(r)}),\\T_3&=\tfrac{1}{6}\nabla^{\mathrm{T}}\{(\hat{\theta}^{(r)}-\check{\theta}^{(r)})^{\mathrm{T}}\nabla^2\hat{l}_p(\hat{\theta}^{(r)})(\hat{\theta}^{(r)}-\check{\theta}^{(r)})\}(\hat{\theta}^{(r)}-\check{\theta}^{(r)}),\\T_4&=\tfrac{1}{2}(\hat{\theta}^{(r)}-\check{\theta}^{(r)})^{\mathrm{T}}\nabla b(\hat{\theta}^{(r)})\{I+o(I)\}(\hat{\theta}^{(r)}-\check{\theta}^{(r)}).\end{aligned}$$

因为 $\nabla^{\mathrm{T}}\hat{l}_p(\hat{\theta}^{(r)})=0$, 所以 $T_1=0$.

$$||\hat{\theta}^{(r)}-\check{\theta}^{(r)}||=O_p((d/n)^{-1/2}).$$

由条件 A3,A5,A7, 及引理 4, 有

$$T_3=o_p(1),$$

$$T_4\leq n\max_{j\in\mathcal{L}}\{p''(|\theta_{0j}|),\theta_{0j}\neq 0\}||\hat{\theta}^{(r)}-\check{\theta}^{(r)}||^2=o_p(1).$$

由引理 5 的证明, 有

$$\nabla^2\hat{l}(\hat{\theta}^{(r)})=O_p(1).$$

因此由引理 5 的证明及 (2.18) 式, 有

$$\widetilde{l}(L_n) = -(2\widetilde{\hat{l}}_p(\hat{\theta}^{(r)}) - 2\widetilde{\vec{l}}_p(\check{\theta}^{(r)})) = -2T_2 + o_p(1)$$

$$= n\bar{U}^{\mathrm{T}}(\theta_0^{(r)})\widetilde{\Sigma}^{-1/2}\{\widetilde{\Sigma}^{-1/2}(H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2 - \widetilde{H}_2^{\mathrm{T}}(\widetilde{H}_2\widetilde{\Sigma}^{-1}\widetilde{H}_2^{\mathrm{T}})^{-1}\widetilde{H}_2)\widetilde{\Sigma}^{-1/2}\}\widetilde{\Sigma}^{-1/2}\bar{U}(\theta_0^{(r)}) + o_p(1).$$

由幂等阵的性质知,$\widetilde{\Sigma}^{-1/2}(H_2^{\mathrm{T}}(H_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}H_2 - \widetilde{H}_2^{\mathrm{T}}(\widetilde{H}_2\widetilde{\Sigma}^{-1}\widetilde{H}_2^{\mathrm{T}})^{-1}\widetilde{H}_2)\widetilde{\Sigma}^{-1/2}$ 可以写成 $\Xi_n^{\mathrm{T}}\Xi_n$, 其中 Ξ_n 是一个 $q \times (p-1)$ 矩阵, 使得 $\Xi_n\Xi_n^{\mathrm{T}} = I_q$. 并且容易得到

$$n^{1/2}\Xi_n\widetilde{\Sigma}^{-1/2}\bar{U}(\theta_0^{(r)}) \xrightarrow{\mathcal{D}} N(0, I_q).$$

因此有

$$n(\Xi_n\widetilde{\Sigma}^{-1/2}\bar{U}(\theta_0^{(r)}))^{\mathrm{T}}(\Xi_n\widetilde{\Sigma}^{-1/2}\bar{U}(\theta_0^{(r)})) + o_p(1) \xrightarrow{\mathcal{D}} \chi_q^2.$$

得到定理 2 的证明.

第三章　模拟研究

§3.1 模拟算法

对于调整参数 τ 的选取，Fan,Li 和 Peng 在 2009 年时指出当参数的维数发散时, 用 BIC 准则选择的 SCAD 的惩罚函数的调整参数能够以概率趋于 1 保证选出真模型. 因此我们采用 BIC 准则来选择调整参数 τ,

$$BIC(\tau) = 2\hat{l}_p(\theta_\tau^{(r)}) + C_n \cdot df_\tau \cdot \log(n).$$

其中, 我们选择的 $\hat{\tau} = \arg\min_\tau BIC(\tau).\theta_\tau^{(r)}$ 是关于调整参数 τ 的 $\theta^{(r)}$ 的惩罚经验似然估计; d_τ 是在 $\theta_\tau^{(r)}$ 中非零系数的个数. 当 p 固定时,$C_n = 1$, 否则 $C_n = \{\log\log p, 1\}$.

由于在惩罚经验似然中目标函数的非二次性质, 我们必须用一个迭代算法去计算 (2.6) 式的最小值. 在本文中, 我们将非线性优化迭代算法与局部二次逼近（local quadratic approximation：LQA）结合使用来求 (2.6) 式的最小值. 即对 $|\theta_j| \approx |\theta_{j0}|$,SCAD 惩罚可由下面式子逼近：

$$p_\tau(|\theta_j^{(r)(k)}|) \approx p_\tau(|\theta_j^{(r)(k)}|) + 1/2\{p'_\tau(|\theta_{j0}^{(r)(k)}|)/|\theta_{j0}^{(r)(k)}|\}\{\theta_j^{(r)(k)2} - \theta_{j0}^{(r)(k)2}\}.$$

其中,$\theta_j^{(r)(k)}$ 是 $\theta_j^{(r)}$ 的第 k 个估计值. 给定相合的初始值 θ^0, 然后采用在 Owen(2001) 中所讨论的通过非线性优化 Newton-Raphson 算法与 LQA 进行结合, 重复步骤直到收敛为止. 在迭代过程中, 我们采用在 Fan&Li(2001) 中的策略, 如果 (2.6) 式中的最小值 $\hat{\theta}^{(r)}$ 的第 k 步的第 j 个分量 $\hat{\theta}_j^{(r)(k)}$ 在零附近, 则令 $\hat{\theta}_j^{(r)} = 0$, 对于 $k = 0, 1, 2, \cdots, \delta_1 = \delta_2 = 10^{-6}$. 具体迭代步骤如下：

1. $k \leftarrow k + 1$.

2. $\theta^{(r)(k+1)} = \theta^{(r)(k)} + \{[\widetilde{\Sigma}^{-1} - \widetilde{\Sigma}^{-1}\widetilde{H}_2^{\mathrm{T}}(\widetilde{H}_2\widetilde{\Sigma}^{-1}\widetilde{H}_2^{\mathrm{T}})^{-1}\widetilde{H}_2\widetilde{\Sigma}^{-1}$
   $+\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}}({}_2\widetilde{\Sigma}^{-1}H_2^{\mathrm{T}})^{-1}b'(\cdot)\widetilde{\Sigma}^{-1}][\bar{U}(\theta^{(r)}) - b(\cdot)]\}_{\theta^{(r)(k)}}$.

3. 如果 $||\theta^{(r)(k+1)}|| \leq 1$, 则 $\theta_\tau^{(k+1)} = \sqrt{1 - ||\theta^{(r)(k+1)}||}$.

4. 如果 $|\theta_j^{(k+1)}| < \delta_1$, 则令 $\theta_j^{(k+1)} = 0$. 并且重新定义 $\theta^{(k+1)} = \theta^{(k+1)}/||\theta^{(k+1)}||$.

5. 计算 $\theta^{(k+1)}$ 与 $\theta^{(k)}$ 之间的距离：$diff_\theta = ||\theta^{(k+1)} - \theta^{(k)}||$.

6. 如果 $diff_\theta < \delta_2$, 则跳出, 否则重复步骤 1-5.

§3.2 模拟条件

(1) 真实模型：$Y = 6.25\exp(-(\theta^{\mathrm{T}}X)) + \varepsilon$.

其中参数真值 $\theta_0 = (3, 1.5, 0, 0, 2, 0, \cdots, 0)^{\mathrm{T}}/\sqrt{15.25} \in R^p$,X 是均值为 0, 协方差阵为单位矩阵的 p 维正态随机向量.$\varepsilon_i \sim N(0,1), i = 1, 2, \cdots, n$.

(2) 样本容量分别取 n=10,100,200.$p = [4n^{1/3}] - 3$

(3) 模拟重复次数 1000 次.

§3.3 模拟结果

表 1 EL 与 SCAD-EL 的模拟研究

(n,p)	方法＼估计值	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_5$
	Oracle	0.7682213	0.3841106	0.5121475
(10,6)				
	EL	0.7333689	0.3960902	0.5525239
	SCAD-EL	0.7513949	0.3920009	0.5307928
(100,16)				
	EL	0.7609781	0.3891920	0.5191752
	SCAD-EL	0.7629189	0.3859622	0.5186405
(200,21)				
	EL	0.7639742	0.3879213	0.5170992
	SCAD-EL	0.7669537	0.3839057	0.5142007

在表 1 中, 标示有 $\hat{\theta}_j, j = 1, 2, 5$ 的每一列报告了在模拟中分别使用经验似然 (EL) 和惩罚经验似然 (SCAD-EL) 方法得到非零估计值.Oracle 这一行是我们在做模拟时非零参数的真值.

从表 1 的模拟结果显示, 虽然随着样本量的增加, 经验似然的估计值向参数真值靠近, 但是没有 SCAD 惩罚经验似然靠近真值的速度快. 当参数的维数随着样本量增大而增大时,SCAD 惩罚经验似然具有 Oracle 的性质.

第四章　总结和进一步的问题

本论文在对前人已有研究成果的基础上进行综合研究分析, 主要考察高维单指标模型的参数估计及检验问题. 我们主要结合 Fan&Peng(2004) 的惩罚似然思想和 Hjort(2009),Chen (2009) 的经验似然思想, 针对单指标模型提出了高维惩罚经验似然方法. 理论证明和模拟结果显示, 在处理单指标模型下的变量选择及检验问题中, 惩罚经验似然方法比传统的单一的经验似然方法更加简单有效. 并且通过对单指模型使用惩罚函数产生的模型, 在构造置信域时存在以下两个优点：一、稀疏性使得模型的复杂度降低；二、稀疏性对构造的置信域更加稳定. 从实用的角度来看, 利用惩罚经验似然方法可以有效节约成本, 实用性更强, 具有较高的推广价值.

本论文对单指标模型提出的惩罚经验似然方法是在已有的方法基础上, 重新组合, 取长补短, 简易易懂, 通过理论证明与模拟实验的结果, 表明这种方法在处理高维数据时, 是一种不错的备选方法.

本文可以进一步研究的问题有：

1. 经验似然计算相对复杂.Owen(1988)、Qin 和 Lawless(1994) 的论文中都指出, 经验欧氏似然与经验似然具有相同的渐近性质, 但经验欧氏似然是从欧氏距离出发, 在计算上更加简单. 也就是此处若将经验似然中的似然距离 $\log\prod_{i=1}^{n} np_i$ 改为欧氏距离 $-\frac{1}{2}\sum_{i=1}^{n}(p_i-1/n)^2$, 即采用经验欧氏似然代替经验似然来进行变量选择, 将能进一步提高计算的速度, 降低计算的复杂度.

2. 目前统计面临着许多挑战性问题, 如高维、相依和不完全数据等. 如何对传统方法进行改进以得到解决这些问题的新方法, 也许惩罚思想会起到非常重要的作用. 另外, 基于一些已经存在的方法, 利用惩罚思想也许会得到一些新的更加有效的方法.

3. 单指标模型有降维的作用, 在本文中没有对 $p \gg n$ 的情型进行讨论, 我们下一步会从这方面对其进行探讨.

4. 本文所讨论的单指标模型结论很容易推广到广义回归模型, 限于篇幅和精力, 此处不再进一步讨论.

5. 本文所讨论的思想和方法可以用来处理金融市场中的投资组合问题以及进行医学数据 (如：基因表达数据) 分析, 我们下一步将会选择一些实际数据进行实证分析.

参 考 文 献

[1] T.J.Hastie,R.Tibshirani,J.Friedman.The elements of statistical learning:Data mining inference and prediction[J].Springer.2001.

[2] J.H.Friedman,W.Stuetzle.Projection Pursuit regression[J].American Statistical Association.1981,76:817-823.

[3] 薛留根. 单指标模型的统计推断 [J], 数理统计与管理,2012,1, 第 31 卷第 1 期.

[4] H.Ichimura.Semiparametric least square(SLS)and weighted SLS estimation of single-index models[J]. Econorne,1993,58:71-120.

[5] J.L.Horowitz.Semiparametric methods in econometrics[M].London:Spring-Verlag,1998.

[6] C.F.Manski.Identification of binary response models[J]. American Statistical Association,1988,83:729-738.

[7] T.M.Stoker.Consistont estimation of sealed coeffieients[J].Econometrica,1986,54:1461-1481.

[8] W.Härdle,A.B.Tsybakov.How sensitive are average derivative[J].Econometrics,1993,58:31-38.

[9] W.Härdle,T.M.Stoker.Investing smooth multiple regression by the method of average derivatives[J] American Statistical Association.,1989,84:986-995.

[10] J.L.Powell,J.H.Stock,T.M.Stoker.Semiparametric estimation of index coefficients[J]. Econometrika,1989,57:1403-1430.

[11] A.B.Tsybabov.Rubust reconstruction of functions by the local-approximation methods[J].Problems.Inform.Tans.,1986,22:133-146.

[12] P.Rilstone. Nonparametric hypothesis testing with parametric rates of convergence[J]. International Economic Review. 1991,32:209-227.

[13] L.G.Xue,L.X.Zhu.Empirical likelihood for single-index models[J]. Multivariate Analysis97,2006,1295-1312.

[14] W.Härdle,P.Hall,H.Ichimura.Optimal smoothing in single-index models[J].Annals of Statistics,1993,21:157-178.

[15] Y.C.Xia.Model checking in regression via dimension reduction[J].Biometrika,2009, 96(1):133-148.

[16] J.C.Escanciano,K.Song.Testing single-index restrictions with a focus on average derivatives[J]. Econometrics,2010,156(2):377-391.

[17] R.Klein,C.Shen.Bias corrections in testing and estimating semiparametric single-index models[J].Ecomometric Theory,2010,26(6):1683-1718.

[18] Y.Bai,W.K.Fung,Z.Y.Zhu.Penalized quadrtic inference for single-index models with longitudinal data[J]. Multivariate Analysis,2009,100(1):152-161.

[19] W.Lin,K.B.Kulasekera.Testing the equality of linear single-index models[J]. Multivariate Analysis,2010,101(5):1156-1167.

[20] O.Lopez.Single-index regression models with right-censored responses[J].Statistical Planning and Inference,2009,139(3):1082-1097.

[21] X.W.Lu.Asymptotic distributions of two synthetic data estimators for censored single-index models[J].Multivariate Analysis,2010,101(4):999-1015.

[22] Y.H.Wang,J.S.Chen,S.Y.He,Q.HWang.Estimation of single index model with missing response at random[J].Statistical Planning and Inference,2010,140(7):1671-1690.

[23] 王启华. 经验似然统计推断方法发展综述 [J]. 数学进展.2004,33(2):141-151.

[24] A.B.Owen.Empirical likelihood ratio confidence intervals for single function[J]. Biometrika75,1988.23:7-249.

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

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

[27] Emerson,A.B.Owen.Calibration of the empirical likelihood method for a vector mean[J].Electron Statistic,2009,1-1605.

[28] F.Bartolucci.A penalized version of the empirical likelihood ratio for the population mean[J]. Statistical & Probability Letters.2007,77:104-110.

[29] T.Otsu.Penalized empirical likelihood estimation of semiparametric models[J]. Multivariate. Analysis.2007,98:1923-1954.

[30] C.Y.Tang,C.L.Leng.Penalized high dimensional empirical likelihood[J].Department of Statistics and Applied Probability.

[31] N.Hjort,I.Mckeague,I.V.Keilegon.Extending the scope of empirical likelihood[J]. Annals of Statistics 37,2009,1079-1111.

[32] J.Fan,R.Li.Variable selection via nonconcave penalized likelihood and its oracle properties[J]. American Statistical Association.2001,96:1348-1360.

[33] J.Qin,J.Lawless.Empirical likelihood and general eatimating equations[J]. Annals of Statistics.1994,22,300-325.

[34] J.Fan,H.Peng.Nonconcave penalized likelihood with a diverging number of parametersn[J]. Annals of Statistics 32,2004,928-961.

[35] S.Chen,L.Peng,Y.L.Qin.Effects of Data Dimension on Empirical Likelihood[J]. Biometrika96,2009,711-722.

[36] A.Antoniadis, J.Fan.Wavelets in statistics:A review (with discussion)[J].Italian Statistical Society.1997,6:131-138.

[37] J.Fan,I.Gijbels.Local polynomial modeling and its applications[M].London:Chapman&Hall,2010.

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

[39] 张尧庭, 方开泰. 多元统计分析引论 [M]. 科学出版社,1982,6.

[40] J.Qin,J.Lawless.Estimating equations empirical likelihood and constraints on parameters[J].The Canadian Journal of Statistics,1995.2:145-159.

[41] A.B.Owen.Empirical likelihood[J].Annals of Statistics.London.Chapman&hall.2001.

[42] S.N.Lahiri,S.Mukhopadhyay.A penalized empirical likelihood method in high dimensions[J].Annals of Statistics,2012,5:2511-2540.

[43] J.Q.Fan.Variance estimation using refitted cross-validation in ultrahigh dimensional regression[J].Royal Statistical Society.Series B.Statistical Methodology74,2012,37-65.

[44] L.R.Gao,P.Heng,L.X.Zhu.Nonconcave penalized M-estimation with a diverging number of parameters[J].Statistica Sinica,2011,21:391-419.

[45] 张军舰. 统计中的惩罚思想.2008.(手稿).

[46] A.B.Owen.Empirical likelihood for linear models[J].Annals of Statistics.1991,19,1725-1747.

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

[48] J.Fan,R.Li.Statistical challenges with high dimensionality:feature selection in knowledge discovery[J].Proceedings of the International Congress of Mathematicians.2006,595-622.

致　谢

光阴似箭，日月如梭，转眼之间三年就在我埋头苦读中不知不觉的过去，回想走过的岁月，这三年的读研的美好时光还历历在目，不禁的感慨万千.

在读研期间，我首先要感激的是我的家人，如果没有他们悉心的帮助和支持，我的学业也不会延续到今天，感谢他们对我的理解和支持. 是他们一次又一次的鼓励和关怀，给了我莫大的动力和信心，面对困难不再退缩，勇敢的面对生活的挑战，勇于向更高的层次迈!

特别感谢我的导师张军舰副教授. 张老师是一位和蔼可亲的老师，他具有严谨的治学态度，我的每一点进步都离不开张老师的耐心指点，生活中的每一个挫折都离不开张老师的耐心开导，张老师不仅指导我如何做学问，还教我很多做人道理，师从张老师是使我的一生收益匪浅. 正是张老师的悉心指导和鼓励才使我顺利愉快的度过我的读研生活. 对导师的感激，似乎无法用言语来形容我的感激之情，惟愿师生情谊一生延续. 祝愿老师及家人健康快乐，永远幸福!

在这里我还要特别感谢秦永松教授，本文的主体思想来自秦老师，感谢他的理解、支持和帮助，在此对秦老师表示衷心的感谢.

感谢杨善朝教授、秦永松教授、邓国和教授、雷庆祝老师、梁鑫老师、李英华老师、唐织辉老师、…… 对我的教育和培养，老师们兢兢业业，一丝不苟的从事教学科研工作，使我在学习中丰富了知识，开拓了视野，感谢他们在我的学习和研究上给我的莫大的帮助.

感谢师兄晏振对我论文工作的支持，感谢我的师兄师姐胡小凤、黄锦联、何伦春，李自军；同门罗巧铃、易家英；专业同学吴金蔚、陈泊伶、赵彩霞、何鹏飞等；以及师弟师妹们，感谢你们一直以来对我的支持、鼓励和帮助，同时感谢我的室友王淑娟、张玉芳感谢你们给我的生活增添了无穷的乐趣，同时也提出了很多开有益的建议和意见，让我少走了不少的弯路. 在此，深表谢意!

最后，对所有帮助和支持我的亲人，老师，同学和朋友再次道一声：谢谢，并祝好人一生平安!

论文独创性声明

本人郑重声明：所提交的学位论文是本人在导师的指导下进行的研究工作及取得的成果。除文中已经注明引用的内容外，本论文不含其他个人或其他机构已经发表或撰写过的研究成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人承担本声明的法律责任。

论文作者签名：李吉妮 日期：2014.6.5

论文使用授权声明

本人完全了解广西师范大学有关保留、使用学位论文的规定。本人授权广西师范大学拥有学位论文的部分使用权，即：学校有权保留本人所送交学位论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文；学校有权向国家有关部门或机构送交学位论文的复印件和电子版。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外，允许论文被查阅和借阅，可以公布（包括刊登）论文的全部或部分内容。论文的公布（包括刊登）授权广西师范大学学位办办理。

本学位论文属于：

□保密，在年解密后适用授权。

☑不保密。

论文作者签名：李吉妮 日期：2014.6.5

指导教师签名：[illegible] 日期：2014.6.5

作者联系电话：_____ 电子邮箱：_____