有限因变量模型中的参数估计

PARAMETER ESTIMATION IN LIMITED DEPENDENT VARIABLE MODELS

马建军 张玉春 赵晓丽 ◎ 著

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

有限因变量模型中的参数估计

马建军　张玉春　赵晓丽　著

電子工業出版社
Publishing House of Electronics Industry
北京 • BEIJING

内 容 简 介

有限因变量模型是具有潜在变量的一大类模型的统称。根据模型中响应变量的不同可以将其分为多项选择模型、多元秩-序模型、多重二元响应模型等。本书基于逆回归方法构造了多重二元响应模型、多元秩-序模型、广义多项选择模型中回归系数的估计，证明了估计是渐近正态的，并且利用 δ-方法推导了估计的渐近分布，在这个基础上构造了模型的假设检验；基于极大似然方法构造了多重二元响应Probit模型参数的渐近有效估计和多元秩-序Logit模型回归系数的估计。本书介绍的参数估计方法有效避免了维数问题。

本书适合概率论与数理统计及相关专业的科研人员阅读。

图书在版编目（CIP）数据

有限因变量模型中的参数估计/马建军，张玉春，赵晓丽著. —北京：电子工业出版社，2018.8
ISBN 978-7-121-35095-5

Ⅰ. ①有… Ⅱ. ①马… ②张… ③赵… Ⅲ. ①统计模型－参数估计 Ⅳ. ①C81

中国版本图书馆 CIP 数据核字（2018）第 216977 号

策划编辑：刘小琳
责任编辑：刘小琳　　特约编辑：许波建
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编：100036
开　　本：720×1000　1/16　印张：7.5　字数：152 千字
版　　次：2018 年 8 月第 1 版
印　　次：2018 年 8 月第 1 次印刷
定　　价：39.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：（010）88254538，liuxl@phei.com.cn。

前言

Preface

有限因变量模型（Limited Dependent Variable Models，LDV 模型）是具有潜在变量的一大类模型的统称。根据模型中响应变量的不同类型可以将其分为多项选择模型、多元秩-序模型、多重二元响应模型等。有限因变量模型是典型的高维数据模型，且模型中具有潜在变量，维数问题给有限因变量模型中的参数估计带来了很大困难。

为了解决高维数据模型中的参数估计问题，D. L. McFadden 将蒙特卡罗模拟方法应用于有限因变量模型的参数估计，相继提出了模拟矩方法、模拟极大似然方法等模拟估计方法。V. A. Hajivassiliou 在统计学手册中对有限因变量模型的模拟估计方法进行了系统的阐述。模拟估计方法计算量大、操作复杂，难以被非专业人员应用。

李克昭提出的充分降维理论方法，如切片逆回归方法（Sliced Inverse Regression）、基本海森方向（Principal Hessian Direction），可以建立降维模型以估计中心降维空间的基向量，从而在高维数据的回归问题中进行大规模数据降维。切片逆回归方法在进行数据降维时，有计算简单、估计效率高等优点，甚至有人提出，切片逆回归方法会像最小二乘法一样被广泛应用。

本书引入了数据降维的方法来构造有限因变量模型中回归系数的估计。根据有限因变量模型的特点，我们给出了一种新的逆回归方法求有限因变量模型中回归系数的估计，这种方法有效地避免了维数问题，并且有以下几个优点：

（1）构造的估计量是具有显式表达式的估计量。

（2）证明估计量具有渐近正态性。

（3）推导出回归系数估计的渐近分布。

（4）基于回归系数估计的渐近分布构造了模型相关的假设检验。

除了应用逆回归方法之外，本书中也应用极大似然方法构造了多重二元响应 Probit 模型中参数的渐近有效估计和多元秩-序 Logit 模型中回归系数的估计。在这两个模型中所构造的估计也有效地避免了维数问题。

本书的内容安排如下，第 1 章详细介绍了有限因变量模型的定义、研究方法及发展现状，第 2 章构造了多重二元响应模型回归系数的估计，第 3 章构造了多元秩-序模型回归系数的估计，第 4 章构造了广义多项选择模型回归系数的估计，第 5 章构造了多重二元响应 Probit 模型参数的渐近有效估计，第 6 章构造了多元秩-序 Logit 模型回归系数的极大似然估计。每章均研究了估计的渐近分布及假设检验问题。

本书汇聚了作者对有限因变量模型相关理论的研究成果。本书的出版还要感谢辽宁省教育厅科学研究一般项目（项目编号：LG201623）的大力支持。

由于作者水平所限，书中难免有错误和不足的地方，欢迎读者批评指正。

马建军

2018 年 6 月

目录

Contents

第1章 绪 论

有限因变量模型的英文全称为 Limited Dependent Variable Models，简记为 LDV 模型。LDV 模型在经济计量学中占有重要的地位，并且在很多领域有着广泛的应用，如市场调查、心理学测试、环境研究和选举学分析等。

1.1 LDV 模型的定义

1.1.1 模型的定义

定义 1.1.1 设 $p\times1$ 向量 $\boldsymbol{Y}^*$ 与 $q\times p$ 矩阵 $\boldsymbol{X}$ 有如下关系：

$$\boldsymbol{Y}^* = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1.1}$$

式中，$\boldsymbol{\varepsilon}\in\mathbf{R}^p$ 为误差向量，通常 $\boldsymbol{X}$ 与 $\boldsymbol{\varepsilon}$ 独立，$\boldsymbol{\beta}\in\mathbf{R}^q$，是未知的回归系数。在一些实际问题中，$\boldsymbol{Y}^*$ 是一个潜在变量（Latent Variable），它是不可观测的，只能观测到一个与它有关的变量 $\boldsymbol{Y}$。$\boldsymbol{Y}$ 和 $\boldsymbol{Y}^*$ 之间的关系由一个映射 $\tau(\cdot)$ 确定，即

$$\boldsymbol{Y} = \tau(\boldsymbol{Y}^*) \tag{1.2}$$

由线性模型式（1.1）和映射式（1.2）所决定的模型称为有限因变量模型（LDV 模型）。

不同的映射 τ 决定了不同的 LDV 模型，它包含很多子模型，在本书中主要研究以下三种 LDV 模型。

1．多项选择模型

定义 1.1.2 记 $\boldsymbol{Y}^* = (y_1^*, y_2^*, \cdots, y_p^*)^{\mathrm{T}}$，若映射 τ 满足：$Y\in\{1,2,\cdots,p\}$，并且 Y 是 $\boldsymbol{Y}^*$ 中最大分量的下标，即

$$Y = \arg\max_h \{y_1^*, y_2^*, \cdots, y_p^*\} \tag{1.3}$$

这时我们称由映射式（1.3）和线性模型式（1.1）所确定的模型为多项选择模型（Multi-Nominal Choice Model，MNC 模型）。

在经济计量学中，通常称 $\boldsymbol{Y}^* = (y_1^*, y_2^*, \cdots, y_p^*)^{\mathrm{T}}$ 中的各分量为各个选择的效用函数，而 Y 是一个指示标志，表明哪一个选择的效用最高。

2. 多元秩-序模型

定义 1.1.3 如果按照大小对 $\boldsymbol{Y}^*$ 中的各分量进行排序，即

$$y_{o_1}^* \geqslant y_{o_2}^* \geqslant \cdots \geqslant y_{o_p}^*$$

上述排列是向量 $\boldsymbol{Y}^*$ 中各分量的一个序，其中 $(o_1, o_2, \cdots, o_p)$ 是 $(1, 2, \cdots, p)$ 的一个置换。若定义

$$\boldsymbol{Y} = \tau(\boldsymbol{Y}^*) = (o_1, o_2, \cdots, o_p)$$

称 $\boldsymbol{Y}$ 为 $\boldsymbol{Y}^*$ 的一个序，称由此映射定义的 LDV 模型为多元秩-序模型（Multi-Variate Rank-Ordered Model，MRO 模型）。或者，下面等价的定义：令 r_j 为向量 $\boldsymbol{Y}^*$ 中小于等于第 j 个分量的分量个数，这时称 r_j 为第 j 分量的秩。同样令

$$\boldsymbol{Y} = \tau(\boldsymbol{Y}^*) = (r_1, r_2, \cdots, r_p) \tag{1.4}$$

由线性模型式（1.1）和映射式（1.4）所确定的 LDV 模型也称作多元秩-序模型。

在一些应用中，建立多元秩-序模型可能更为合理，因为在面对多个选择的时候，消费者可能会在多个选择中做出一个次序选择。相对于多项选择模型而言，多元秩-序模型包含更多的信息。

3. 多重二元响应模型

定义 1.1.4 若定义映射 τ 满足 $\boldsymbol{Y} = \tau(\boldsymbol{Y}^*) = (y_1, y_2, \cdots, y_p)^{\mathrm{T}}$

$$y_j = \begin{cases} 1, & y_j^* > 0 \\ 0, & y_j^* \leqslant 0 \end{cases} \quad j = 1, \cdots, p \tag{1.5}$$

式中，$\boldsymbol{Y} = (y_1, \cdots, y_p)^{\mathrm{T}}$ 是一个由 0 和 1 构成的 p 维列向量，则称线性模型式（1.1）和映射式（1.5）所确定的 LDV 模型为多重二元响应模型（Multi-Binary Response Model）。

多重二元响应模型是一种是非选择模型，对于某一个选择项或者是选择，或者是不选择，也可以描述一些二元现象是否发生，如价格的涨跌等。但多重二元响应模型在各个选择项之间不存在多项选择模型和多元秩-序模型中的大小次序关系。

以上三种模型有以下共同特点：

（1）观测不到与 $\boldsymbol{X}$ 有直接函数关系的 $\boldsymbol{Y}^*$ 的具体值，并且无法利用映射 τ 由观测值 $\boldsymbol{Y}$ 恢复 $\boldsymbol{Y}^*$ 的值[1]。在多项选择模型和多元秩-序模型中，观测到的是 $y_j^*(j = 1, 2, \cdots, p)$ 之间的一种大小关系。在多重二元响应模型中，观测到的是

$y_j^*(j=1,2,\cdots,p)$ 与 0 比较的大小关系。

（2）若不对参数进行一定的限制，那么模型是不可识别的。

对于任意一个常数 $k(k>0)$，用 k 乘以线性方程式（1.1）的两边，有 $k\boldsymbol{Y}^*=\boldsymbol{X}^{\mathrm{T}}(k\boldsymbol{\beta})+k\boldsymbol{\varepsilon}$，对于上述三种 LDV 模型有 $\boldsymbol{Y}=\tau(\boldsymbol{Y}^*)=\tau(k\boldsymbol{Y}^*)$，$\boldsymbol{Y}$ 的值保持不变。当误差项 $\boldsymbol{\varepsilon}$ 的分布假定为多元正态分布 $N(\boldsymbol{0},\boldsymbol{\Omega})$ 时，则被称为一类 Probit 模型，Probit 模型中的参数为 $(\boldsymbol{\beta},\boldsymbol{\Omega})$。对于任意常数 $k>0$，$\boldsymbol{Y}$ 在参数 $(\boldsymbol{\beta},\boldsymbol{\Omega})$ 下的分布与在参数 $(k\boldsymbol{\beta},k^2\boldsymbol{\Omega})$ 下的分布相同。因此模型中的参数是不可识别的。为了能够使模型中的参数可以识别，一般加上一些限制条件，例如，在 Probit 模型中可以令随机误差项的协方差阵 $\boldsymbol{\Omega}$ 中的第一对角线元素为 1，并称之为可识别性限[2]。

在多重二元响应模型中，如果响应变量 $y_j\ (j=1,2,\cdots,p)$ 采用如下定义：

$$\begin{cases} y_j=1, y_j^*>c_j \\ y_j=0, y_j^*\leqslant c_j \end{cases}$$

式中，$c_j\ (j=1,2,\cdots,p)$ 不全为 0，则模型中的参数在不加限制时是可以识别的。

在上述三种模型中，如果对参数没有任何限制，尽管参数是不可识别的，但是回归系数的方向是可以识别的，也就是说，回归系数可以被估计到相差一个正的常数乘积。由于以上三种模型着重于比较各选择项的大小关系，如果得到 $k\boldsymbol{\beta}\ (k>0)$ 的估计，其中 $\boldsymbol{\beta}$ 是回归系数，仍然可以建立模型，只不过是 $\boldsymbol{Y}^*$ 的尺度被放大了。由于 $y_j^*(j=1,2,\cdots,p)$ 之间被放大了相同的尺度，所以并不影响 $\boldsymbol{Y}$ 的结果。所以，如果只对回归系数感兴趣，则可以估计回归系数的方向。

1.1.2　模型的统计推断问题

以上三种 LDV 模型中观测不到 $\boldsymbol{Y}^*$，而且也不能由 $\boldsymbol{Y}$ 恢复得到 $\boldsymbol{Y}^*$，所以最小二乘法无法应用。在具有可识别限制的 LDV 模型中，目前常用的估计方法是极大似然估计。对于 Probit 模型，记 $P_y(\boldsymbol{X};\boldsymbol{\beta},\boldsymbol{\Omega})=P(\boldsymbol{Y}=\boldsymbol{y}\mid \boldsymbol{X};\boldsymbol{\beta},\boldsymbol{\Omega})$ 是给定 $\boldsymbol{X}$ 时 $\boldsymbol{Y}$ 的条件分布，令 $\{(\boldsymbol{Y}^{(i)},\boldsymbol{X}^{(i)}),i=1,2,\cdots,n\}$ 是来自 Probit 模型的一组样本，则似然函数为

$$\prod_{i=1}^{n}P_{y^{(i)}}(\boldsymbol{X}^{(i)};\boldsymbol{\beta},\boldsymbol{\Omega}) \tag{1.6}$$

当 p 较大时 $P_{y^{(i)}}(\boldsymbol{X}^{(i)};\boldsymbol{\beta},\Omega)$ 是在区域 $\{\boldsymbol{Y}^*:\tau(\boldsymbol{Y}^*)=\boldsymbol{Y}\}$ 上关于正态密度的高维积

分。当 $p \geqslant 4$ 时，一般的数值方法无法准确计算这个高维积分，目前一般利用模拟方法计算这个高维积分。由此，产生了所谓的模拟估计方法，而模拟估计方法的计算量是非常大的。在 Probit 模型中，若 $\boldsymbol{\beta}$ 为 q 维向量，扰动项 $\boldsymbol{\varepsilon}$ 的协方差阵 $\boldsymbol{\Omega}$ 为 p 阶方阵：

$$\boldsymbol{\Omega}=\begin{bmatrix} 1 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ & \cdots & \cdots & \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

则待估参数向量为

$$(\beta_1,\cdots,\beta_q,\sigma_{12},\cdots,\sigma_{1p},\sigma_{22},\cdots,\sigma_{2p},\cdots,\sigma_{pp})^{\mathrm{T}}$$

这是一个 $q+p(p+1)/2-1$ 维向量。如果 $q=4, p=5$，则待估参数就是一个 18 维向量。

模拟估计方法求解时需要迭代计算，每次迭代都需要计算 n 个高维积分，所以计算量非常大。

1.2 LDV 模型中参数估计问题的研究现状

对于上述三种 LDV 模型，只有当扰动项取特殊分布时才可以简化计算。McFadden[3]证明了若 ε_j $(j=1,2,\cdots,p)$ 独立同分布（i.i.d.），并且是极值分布时，模型可转化为 Logit 模型。对于协方差阵具有特殊形式的 LDV 模型，Heckman[4]、Bolduc[5]、Bolduc 和 Kaci[6]及 Hausman 和 Wise[7]进行了分析。

近年来，关于 LDV 模型参数估计方法的研究主要集中在模拟方法。1981 年 Lerman 和 Manski[8]首次将 Monte Calro 积分作为一种数字技术用于 LDV 模型的参数估计，Lerman 和 Manski 所使用的模拟器称为 Crude Monte Carlo（CMC）模拟器。CMC 模拟器关于 $\Pr\{\boldsymbol{D};\boldsymbol{\mu},\boldsymbol{\Omega}\}$ 的模拟仅仅模拟 $\boldsymbol{Y}^*$ 落入区域 $\boldsymbol{D}$ 中的频率，其中

$$\boldsymbol{D}=\{\boldsymbol{Y}^*:\boldsymbol{Y}=\tau(\boldsymbol{Y}^*)\},\ \Pr\{\boldsymbol{D};\boldsymbol{\mu},\boldsymbol{\Omega}\}=\Pr\{\boldsymbol{Y}^*\in\boldsymbol{D}|\ \boldsymbol{X},\boldsymbol{Y};\boldsymbol{\mu},\boldsymbol{\Omega}\}$$

误差项 $\boldsymbol{\varepsilon}$ 服从正态分布 $N(\boldsymbol{\mu},\boldsymbol{\Omega})$。

CMC 模拟器的优点是计算速度较快，对于具有处理向量能力的计算软件是比较理想的。但是，CMC 模拟器关于参数不是连续的，这会导致估计的计算和渐近分布的理论存在严重的缺陷。Hendy[9]定义了 Simulation-Variance-Reduction

方法，改进了 CMC 模拟器的计算精度。

Geweke、Hajivassiliou 和 McFadden[10]及 Keane[11]提出了 GHK 模拟器，GHK 模拟器利用了重要抽样方法[12]，使得在应用中可以减小抽样的方差，并且关于参数是连续的。关于 GHK 模拟器的性质在 Borsch-Supan 和 Hajivassiliou[13]的文献中进行了讨论。Hajivassiliou 和 McFadden 考虑了利用 Gibbs 抽样技术从截断分布中抽取样本[14]，这种技术在实际应用中可以很好地收敛到真实分布，并且模拟器关于参数是连续的、可导的。

Newey 和 McFadden 提出了广义模拟矩方法[15]（Generalized Method of Simulated Moments，GMSM）。Hajivassiliou、McFadden 和 Ruudcite 讨论了多元正态的矩形概率的模拟[16]。Hajivassiliou 和 McFadden 提出了模拟得分函数的方法（MSS）[17]，并建立了模拟器的相合性和渐近正态性。Natarajan、McCulloch 和 Kiefer[18]对于多项 Probit 模型的参数估计提出了 Monte Carlo EM 方法。

利用以上模拟方法的前提条件是必须存在极大似然估计。刘金燕、徐兴忠[19]给出了多周期 Probit 模型的极大似然估计存在性的充分必要条件。孙立敏、徐兴忠[20]给出多元秩-序 Probit 模型的极大似然估计存在性的充分必要条件。赵江涛、徐兴忠给出了多项 Probit 模型的极大似然估计的充分必要条件[21]。

Nobile、McCulloch[22,23]、Ross、McCulloch[24,25]、Polson 和 Rossi[26]考虑在多项 Probit 模型的参数估计中使用先验信息，对多项 Probit 模型进行了 Bayes 分析。在可识别限制下，他们给出了 $\boldsymbol{\beta}$ 和 $\boldsymbol{\Omega}$ 的先验分布，其中 $\boldsymbol{\beta}$ 的先验分布是正态分布，而 $\boldsymbol{\Omega}$ 的先验分布服从逆 Wishart 分布。综合参数的先验可以得到一个后验分布，然后在后验分布中抽取样本，从而可以利用先验信息对参数进行 Bayes 估计。

关于 LDV 模型中参数估计方法的其他文献参看参考文献[27-33]。

模拟方法有以下特点：

（1）模拟极大似然估计需要较复杂的计算程序和大量的计算，同时还存在计算的收敛性和计算误差的问题。尽管目前有功能强大的计算机，但方法运用并不方便。

（2）在可识别条件限制下，用模拟方法得到的回归系数的估计仍然是一个方向的估计，而且其渐近分布和有关的假设检验、区间估计问题也没有得到完全的解决。

1.3 逆回归方法

本书的第 2～4 章将应用逆回归方法来估计回归系数，所以这里对逆回归方法进行简要介绍。

1.3.1 逆回归的总体性质

Duan 和 Li[34]提出了分片逆回归方法（Sliced Inverse Regression，SIR）。在一般回归模型中，可以用下面的条件数学期望来估计回归函数

$$\eta(\boldsymbol{x}) = E(y|\boldsymbol{x}) \tag{1.7}$$

称之为正向回归函数。当 $\boldsymbol{x}$ 的维数很高时，由于维数问题（Curse of Dimensionality）的原因，正向回归不能得到令人满意的结果。为了避开维数引起的问题可以考虑利用逆回归函数

$$\xi(y) = E(\boldsymbol{x}|y) \tag{1.8}$$

因为 y 是一维的，所以逆回归函数可以避开维数问题。逆回归估计的理论建立在下述定理的基础上。

定理 1.3.1（Duan 和 Li[34]定理 2.1） 对于一般回归模型 $y = f(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}, \varepsilon)$，如果回归变量 $\boldsymbol{x}$ 服从非退化的椭圆对称分布，那么逆回归函数式（1.8）具有下面的形式

$$\xi(y) = \boldsymbol{\mu} + \boldsymbol{\Sigma\beta}\kappa(y)$$

式中，$\boldsymbol{\mu} = E(\boldsymbol{x}), \boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{x}), \kappa(y)$ 具有下面的形式

$$\kappa(y) = \frac{E[(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\beta}|y]}{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma\beta}}$$

根据上述定理可以得到

$$\boldsymbol{\beta} \propto \boldsymbol{\Sigma}^{-1}[\xi(y) - \boldsymbol{\mu}] \tag{1.9}$$

式中，比例常数为 $1/\kappa(y)$。如果 $\kappa(y) \neq 0$，那么可以根据式（1.9）确定 $\boldsymbol{\beta}$ 的方向。

Li[35]用逆回归方法对下面的降维模型进行数据降维，即

$$y = f(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_1, \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_2, \cdots, \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_k, \varepsilon) \tag{1.10}$$

式中，$\boldsymbol{x}$ 是 $p\times 1$ 的随机向量，$\boldsymbol{\beta}_j (j = 1, 2, \cdots, k)$ 是未知回归系数，ε 与 $\boldsymbol{x}$ 独立。

当 $k < p$ 时，若能估计出 $\boldsymbol{S}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_k)$ 的一组基，便可以达到降维的目的，其中 $\boldsymbol{S}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_k)$ 表示由列向量 $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_k)$ 张成的线性空间，这个空间称作有效降维空间（EDR 空间）。

在 Li[35]的文献中放宽了对 $\boldsymbol{x}$ 分布的限制条件，即满足下面的线性条件。

A1 线性条件：对任意 $\boldsymbol{b}\in\mathbf{R}^p$，条件期望 $E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_1,\cdots,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_k)$ 关于 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_1,\cdots,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_k$ 是线性的，即存在 $c_0,c_1,\cdots,c_k$，使得 $E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_1,\cdots,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_k)=c_0+c_1\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_1+\cdots+c_k\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_k$ 成立。

由于 $\boldsymbol{\beta}_j$ 是未知的，无法验证上述的线性条件，应用时只能要求对所有的 p 维向量都满足线性条件。当 $\boldsymbol{x}$ 的分布为上述的椭圆正态分布时满足这个条件[36-38]。而 Hall 和 Li[39]证明，当随机向量 $\boldsymbol{x}$ 的维数较大时，可以很高的概率发现随机向量 $\boldsymbol{x}$ 满足线性条件。

定理 1.3.2（Li[35]定理 3.1）　对于降维模型式（1.9），如果回归变量 $\boldsymbol{x}$ 满足条件 A1，则中心化的逆回归曲线为

$$E(\boldsymbol{x}|y)-E(\boldsymbol{x})$$

包含在由 $\boldsymbol{\Sigma\beta}_j(j=1,\cdots,k)$ 的列向量所张成的子空间中，其中 $\boldsymbol{\Sigma}=\mathrm{Cov}(\boldsymbol{x})$。

可以看到，定理 1.3.1 是定理 1.3.2 在 $k=1$ 时的特殊情况。根据定理 1.3.2，若 $E(\boldsymbol{x}|y)-E(\boldsymbol{x})$ 包含在列向量 $(\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_k)$ 所张成的子空间内，则 $\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_k$ 与协方差阵 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的列向量所张成的线性空间正交。这样协方差阵 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的 k 个非 0 特征根所对应的特征向量即 $(\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_k)$。$\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_j(j=1,2,\cdots,k)$ 便是向量 $(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_k)$ 张成的 EDR 空间的一组基。只要我们能得到 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的估计，就可以得到 EDR 空间一组基的估计。

关于逆回归方法降维的有关内容参看文献[40-44]。

1.3.2　分片逆回归估计（SIR）

基于来自模型式（1.9）的一组独立同分布样本 $(y_i,\boldsymbol{x}_i)(i=1,2,\cdots,n)$，Li[35]提出了分片逆回归的估计方法（SIR）来估计 EDR 空间的一组基，步骤如下。

（1）将 $\boldsymbol{x}$ 标准化得到

$$\tilde{\boldsymbol{x}}_i=\hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})\,(i=1,2,\cdots,n)$$

式中，$\bar{\boldsymbol{x}}=\dfrac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i$，$\hat{\boldsymbol{\Sigma}}=\dfrac{1}{n}\sum_{i=1}^{p}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})^{\mathrm{T}}$。

（2）将 y 的值域分成 H 个片，即 $I_1,I_2,\cdots,I_H$。令 y_i 落入第 h 个片的比例为 $\hat{p}_h$，有

$$\hat{p}_h=\frac{1}{n}\sum_{i=1}^{n}\delta_h(y_i)$$

式中，$\delta_h(y_i)$ 为示性函数，当 y_i 落入第 h 片时取值为 1，否则为 0。

（3）在每个片内计算 $\tilde{\boldsymbol{x}}_i$ 的样本均值，表示为 $\hat{\boldsymbol{m}}_h(h=1,2,\cdots,H)$，有

$$\hat{\boldsymbol{m}}_h=(\frac{1}{n\hat{p}_h})\sum_{y_i\in I_h}\tilde{\boldsymbol{x}}_i$$

（4）加权协方差阵为

$$\hat{\boldsymbol{V}}=\sum_{h=1}^{H}\hat{p}_h\hat{\boldsymbol{m}}_h\hat{\boldsymbol{m}}_h'$$

式中，$\hat{\boldsymbol{V}}$ 是 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的一个样本估计。

（5）对 $\hat{\boldsymbol{V}}$ 进行特征分解，那么 k 个最大特征值所对应的特征向量便是 $\hat{\boldsymbol{\eta}}_j$ $(j=1,2,\cdots,k)$，则估计为 $\hat{\boldsymbol{\beta}}_j=\hat{\boldsymbol{\Sigma}}^{-1/2}\hat{\boldsymbol{\eta}}_j(j=1,2,\cdots,k)$。

在上述估计方法中，由于应用分片技术对逆回归曲线进行估计，所以称之为分片逆回归方法。逆回归估计关键在于对协方差阵 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的估计，在 Li 提出 SIR 方法之后，有许多统计学家研究了多种估计方法改进对协方差阵 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的估计。Hsing 和 Carrol[45]考虑了 two-case 的情况，对应的协方差阵的估计为 $\hat{\boldsymbol{V}}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_{(2i)}-\boldsymbol{x}_{(2i-1)})(\boldsymbol{x}_{(2i)}-\boldsymbol{x}_{(2i-1)})^{\mathrm{T}}$。Zhu 和 Ng[46]研究了 H 个不同的片中所含的观测个数 c 是相等的情况，协方差阵的估计为

$$\hat{\boldsymbol{V}}=\frac{1}{n(c-1)}\sum_{h=1}^{H}\left\{\sum\sum_{1\leqslant j<l\leqslant c}(\boldsymbol{x}_{(h,l)}-\boldsymbol{x}_{(h,j)})(\boldsymbol{x}_{(h,l)}-\boldsymbol{x}_{(h,j)})^{\mathrm{T}}\right\}$$

实际上，以上几种方法就是将 y 的值域分成若干片，然后在每个片中用起到局部平滑作用的平均值作为 $E(\boldsymbol{x}|y)$ 的估计。一些人考虑利用其他的平滑方法。Zhu 和 Fang[47]考虑了利用核估计来估计逆回归曲线。Fung、He、Liu、Shi[48]考虑了利用 B-spline basis 方法估计逆回归曲线。当 SIR 方法提出以后，人们尝试将 SIR 方法应用于不同的模型，其中 Cook [49]，以及 Cook 和 LEE[50]将 SIR 方法应用于具有二项响应变量的回归模型，但 Cook 文中的响应变量是一维的。关于 SIR 方法应用的其他文献还有[51-63]。

1.3.3 SIR 的大样本性质

SIR 估计被证明是 $\sqrt{n}$ 相合的，但是不同的估计方法的渐近性质是不相同的，其中的一些估计方法被证明是渐近正态的，而且得到了渐近方差阵的表达式。应当注意的是 $\boldsymbol{x}$ 的协方差阵 $\boldsymbol{\Sigma}=\mathrm{Cov}(\boldsymbol{x})$ 已知时与未知时的渐近分布是不同的，这一点在 Fung、He、Liu、Shi[48]的文献中有详细的叙述。

Duan 和 Li[34]证明，在一般回归模型 $y=f(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta},\varepsilon)$ 中，对于任意分片的 SIR，如果回归子 $\boldsymbol{x}$ 的分布是椭圆对称分布，则 $\boldsymbol{\beta}$ 的方向估计 $\hat{\boldsymbol{\beta}}$ 在下面的限制下是渐近正态的：

当 $\boldsymbol{\Sigma}$ 已知时，$\hat{\boldsymbol{\beta}}\boldsymbol{\Sigma}\hat{\boldsymbol{\beta}}=1,\hat{\boldsymbol{\beta}}\boldsymbol{\Sigma}\hat{\boldsymbol{\beta}}>0$；

当 $\boldsymbol{\Sigma}$ 未知时，$\hat{\boldsymbol{\beta}}\boldsymbol{\Sigma}\hat{\boldsymbol{\beta}}=1,\hat{\boldsymbol{\beta}}\boldsymbol{\Sigma}\hat{\boldsymbol{\beta}}>0$。

Hsing 和 Carroll[45]，Zhu 和 Ng[46]，Zhu 和 Fang[47]，以及 Fung、He、Liu、Shi[48]，Saracco[64]分别证明了他们所给出的 SIR 估计是渐近正态的。

Duan 和 Li[34]研究了渐近相对效，将 SIR 方法应用于线性回归模型，然后与最小二乘估计进行比较，发现 SIR 方法在均匀分片的情况下有很好的相对效率。

1.4　准备知识：δ-方法

为了方便读者阅读，本节简单介绍本书中所涉及的极限理论及求极限分布的 δ-方法。

1.4.1　多元 δ-方法

在很多情况下，直接求出一个函数形式统计量的精确方差或分布很困难，因此可以应用近似方法。在求这种函数估计的近似方差或分布时，一种常用的方法就是 δ-方法。

δ-方法的思想就是使用泰勒展开。记 T_n 为某个未知参数向量 $\boldsymbol{\theta}$ 相合估计的统计量，是一个已知函数，由连续映射定理可知，如果序列 T_n 概率收敛于 $\boldsymbol{\theta}$，且 f 在 $\boldsymbol{\theta}$ 是连续的，则 $f(T_n)$ 也依概率收敛于 $f(\boldsymbol{\theta})$。感兴趣的问题是如何利用统计量 T_n 的性质来获得 $f(T_n)$ 的性质。我们可以用函数 $f(T_n)$ 在 $\boldsymbol{\theta}$ 处的泰勒展开式 $f(\boldsymbol{\theta})+f'(\boldsymbol{\theta})(T_n-\boldsymbol{\theta})+\cdots$ 来近似随机向量 $f(T_n)$，从而获得统计量 $f(T_n)$ 的渐近性质，特别是渐近协方差和渐近分布，它可以由 $T_n-\boldsymbol{\theta}$ 的渐近分布来推导 $f(T_n)-f(\boldsymbol{\theta})$ 的渐近分布。

下面利用向量表示来讨论 δ-方法的结论。设 $T_1,\cdots,T_k$ 是随机变量，其均值分别为 $\theta_1,\cdots,\theta_k$。定义估计向量 $\boldsymbol{T}=(T_1,\cdots,T_k)^{\mathrm{T}}$ 和参数微量 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)^{\mathrm{T}}$。假设 $f(\boldsymbol{T})$ 是可微的，定义

$$f_i'(\boldsymbol{\theta})=\frac{\partial}{\partial t_i}g(t)\bigg|_{t_1=\theta_1,t_2=\theta_2,\cdots,t_k=\theta_k}$$