估计方程及结构方程模型的统计推断

张艳青　唐年胜　赵　慧◎著

科学出版社

估计方程及结构方程模型的统计推断

张艳青　唐年胜　赵　慧　著

科学出版社

北　京

内 容 简 介

本书阐述了估计方程及结构方程模型最新的研究成果，全书分为5章，分别介绍了估计方程及结构方程模型的基本理论和方法、带有不可忽略缺失数据的估计方程的Bayes局部影响分析方法、基于估计方程及经验似然的Bayes变量选择方法、结构方程模型中潜在变量选择的探讨，以及基于估计方程的分位数结构方程模型下的Bayes经验似然推断问题等. 此外，本书还介绍了这些理论和方法在医学、心理学和社会学等领域的若干具体应用.

本书可作为统计学、医学、心理学、社会学等专业研究生的参考书，也可供相关专业的研究生、教师和统计工作者参考.

图书在版编目(CIP)数据

估计方程及结构方程模型的统计推断/张艳青，唐年胜，赵慧著. —北京：科学出版社，2017

ISBN 978-7-03-053206-0

Ⅰ. ①估… Ⅱ. ①张… ②唐… ③赵… Ⅲ. ①统计模型–统计推断–研究 Ⅳ. ①O212

中国版本图书馆CIP数据核字（2017）第125694号

责任编辑：张振华 / 责任校对：王万红

责任印制：吕春珉 / 封面设计：东方人华平面设计部

科学出版社 出版

北京东黄城根北街16号

邮政编码: 100717

http://www.sciencep.com

北京京华虎彩印刷有限公司 印刷

科学出版社发行　各地新华书店经销

*

2017年12月第 一 版　开本：B5(720×1000)

2017年12月第一次印刷　印张：7 1/2

字数：140 000

定价：48.00元

（如有印装质量问题，我社负责调换〈京华虎彩〉）

销售部电话 010-62136230　编辑部电话 010-62135120-2005

前　　言

本书是根据当前统计学热点问题和学术成果，并结合作者的研究成果撰写而成的，既考虑了内容的科学性和应用性，又体现了学术思想. 本书在写作上，注重阐述方法论、理论证明、模拟计算和实例分析；在结构上，每一章介绍一个问题，读者可以独立阅读任何一章的内容. 本书涉及的概率论和数理统计的一些基础知识，假定读者已经熟悉，所以没有一一介绍，初学者可以避开阅读理论证明部分. 本书展示最新的研究成果，阐述估计方程及结构方程模型的统计理论和方法，包括估计方程用于处理缺失数据、参数的Bayes经验似然估计，经验似然下的Bayes局部影响分析，经验似然下的Bayes变量选择，处理Bayes经验似然的蒙特卡罗算法，结构方程模型中潜在变量的选择，分位数结构方程模型中的参数的Bayes经验似然估计等. 本书的内容不仅为从事该领域的科研人员提供了相关的研究资料，也为实际应用者提供了一些数据分析方法和计算模拟方法，同时也为有兴趣了解此领域的读者提供了参考.

本书依托于云南省科技领军人才培养项目，由张艳青、唐年胜、赵慧共同撰写.

由于作者水平有限，加之撰写时间仓促，书中不足之处在所难免，恳请广大读者批评指正.

作　者

2017年5月

目　　录

第 1 章　绪　　论

1.1　估计方程推断方法概述

Durbin（1960）最早在统计学和计量经济学中使用现代词语“估计方程”，他观察到，通过求解估计方程的根而获得的估计，可以保留（渐近）无偏性，并猜想估计方程估计具有与最小二乘估计类似的某些最优性，其中最基本的性质就是估计的无偏性和相合性. 我们知道，极大似然估计具有很多很好的统计性质，在应用极大似然方法时，一般假设感兴趣的参数来自某一个确定的参数模型. 当假设的模型不是真正的模型时，基于极大似然方法的统计推断可能产生很大的偏差，甚至不相合，从而导致错误地使用极大似然估计. 另外，当模型发生较大的偏离时，也会使极大似然估计产生很大的偏差，甚至无效，这就是所谓的模型误判问题. 针对模型误判问题，人们扩展了极大似然方法，提出了拟似然估计、伪似然估计、广义矩估计和估计方程估计等. 其中，估计方程估计是一种应用广泛的估计方法，许多参数估计方法，如著名的极大似然法、最小二乘法及矩估计都是它的特殊情形. 估计方程最大的优点是不依赖于任何分布，在错误指定模型的情形下也可以得到可信赖的结果，具有稳健性.

估计方程不会离开统计模型而单独存在，构造估计方程最重要的一步是寻找一个估计函数$\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})$，使

$$E\{\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})\}=0 \tag{1.1}$$

成立. 满足此条件的估计函数称为无偏估计函数. 由估计函数的样本均值可得到估计方程

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta})=0. \tag{1.2}$$

如果方程(1.2)的解存在且唯一，则通过求解方程(1.2)可以得到$\boldsymbol{\theta}$的一个估计，这个估计称为估计方程估计.

在估计方程中，满足条件(1.1)的参数值$\boldsymbol{\theta}_0$通常称为真实参数，满足条件(1.1)的性质称为估计函数的无偏性. 无偏性是一个自然要求，它保证了由估计方程获得的解$\boldsymbol{\theta}$靠近真实参数$\boldsymbol{\theta}_0$. 要求估计函数满足无偏性是非常必要的，首先，在一些正则条件下，由无偏估计函数产生的估计方程可以得到真实参数的相合估计. 在有限样本情况下，由无偏估计方程获得的估计比基于有偏估计方程获得的估计具有更优的性质. 其次，无偏性标准也与一致最小方差无偏估计的古典理论一致. 一致最小方差无偏估计方法是在一族无偏估计中寻找一个具有最小方差的估计. 类似地，估计方程方法是在一个无偏估计函数族中寻找一个可能具有最大方差的估计函数. 一个有更大方差的无偏估计方程一般优于一个有偏估计方程或方差较小的无偏估计方程，因此基于更大方差的无偏估计方程会得到真实参数$\boldsymbol{\theta}_0$的一个更好的估计. 最后，基于无偏估计方程所获得的估计虽不必是真实参数$\boldsymbol{\theta}_0$的无偏估计，但基于无偏估计函数构造的估计方程所获得的估计通常是真实参数的渐近无偏估计.

在估计方程统计推断中，寻找一个无偏估计函数是重要基础，如果能够对产生数据的模型有一定的认识，那么将有助于寻找合适的无偏估计函数. 当假设样本来自一个参数分布族时，可以利用得分函数来构造参数的估计方程，这是产生估计方程估计的最初想法. 但是，这些方法毕竟是有限的，很多参数估计问题并非都可以应用这些方法构造无偏估计函数. 即使能构造出无偏估计函数，也可能因没有充分利用数据所包含的信息，而造成有效信息的损失. 因此，构造无偏估计方程的原则，就是使无偏估计函数简单化及充分利用数据的信息. 但在很多情况下，这样的原则很难同时满足. 最好的解决方法是对数据来源总体假设一个统计模型，当然对于统计模型的假设并不是一个简单的过程，除了已有的经验和知识外，更要结合研究问题本身，以及待分析数据的类型和背景.

在很多情况下，确认了参数的估计方程，便可直接应用矩方法的思想获得参数的估计. 记$\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})$为所构造的$q$维无偏估计函数向量，即满足$E\{\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})\}=0$，其中，$\boldsymbol{\theta}$为$p$维未知参数，$\boldsymbol{X}$为观测变量. 当$q=p$，即方程个数等于参数向量的维数时，应用矩方法的思想，令估计函数的样本矩函数值为0，便可获得参数的估计，这个估计称为矩估计. 当$q>p$时，则存在过度识别，即估计方程的个数大于参数维数，此时要想应用普通的矩方法来估计参数，一种简单的处理方法就是抛弃一

些方程，但这样会损失一些有用的信息.

对于处理这种过度识别条件的参数估计情况，Hansen（1982）提出了一种广义矩估计方法，其基本思想是选择某些标准，如距离度量方法，使估计函数$\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})$的均值函数尽量接近0. 例如，选择一个对称的正定矩阵，将估计函数构造为加权形式的二次型，即选择如下度量方法：

$$E\{\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})\}^{\mathrm{T}}\boldsymbol{W}E\{\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})\}, \tag{1.3}$$

式中，$\boldsymbol{W}$为一个$q\times q$的权矩阵，其选取可以与参数$\boldsymbol{\theta}$无关，也可以与参数$\boldsymbol{\theta}$有关. 显然式(1.3)表示的距离是非负的. 当$E\{\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})\}=0$时，必使得式(1.3)达到最小值；相反，当式(1.3)达到最小值时，一般能使$\|E\{\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})\}\|$达到最小值，即$E\{\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})\}$尽量接近于0，其中，$\|\cdot\|$表示欧式距离. 因此式(1.3)定义的距离是合理的，只需使式(1.3)达到最小值便能给出参数θ的一个合理估计. 用样本均值代替式(1.3)中的期望，就得到广义矩方法估计，有时也称为广义估计方程估计，即

$$\hat{\boldsymbol{\theta}}=\arg\min_{\boldsymbol{\theta}}\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta})\right\}^{\mathrm{T}}\boldsymbol{W}\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta})\right\}.$$

广义矩方法是处理过度识别方程常用的一种估计方法，所得到的广义矩估计一般是相合估计，在一些正则条件下也是渐近正态的. 广义估计方程估计与选择的权矩阵有关，适当选取权矩阵，可以使其有效. 另一种过度识别估计方程的估计方法就是经验似然估计方法，该方法在很多情况下不需要选择权矩阵. 经验似然估计方法充分利用了极大似然方法的思想，但因其与分布无关，故它不像极大似然方法那样容易受到模型误判的影响. 同时，经验似然估计通常与选择了最优权矩阵时的广义矩估计等价，甚至在很多条件下它可以像极大似然估计一样有效.

记$\boldsymbol{X}$具有分布函数F，且F依赖于未知的p维参数$\boldsymbol{\theta}$. 假定有关$\boldsymbol{\theta}$及F的信息可以由满足式(1.1)的q（$q\geqslant p$）个独立的无偏估计函数$\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})$给出. 应用经验似然框架，即极大化如下目标函数：

$$L(F)=\prod_{i=1}^{n}\mathrm{d}F(x_i)=\prod_{i=1}^{n}p_i, \tag{1.4}$$

式中，$p_i = \mathrm{d}F(\boldsymbol{x}_i) = P(\boldsymbol{X} = \boldsymbol{x}_i)$，且满足如下限制条件：

$$p_i \geqslant 0,\quad \sum_{i=1}^n p_i = 1,\quad \sum_{i=1}^n p_i\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta}) = 0.$$

因此，$\boldsymbol{\theta}$的极大经验似然估计可以定义为$\hat{\boldsymbol{\theta}} = \arg\max\limits_{\theta} L(\boldsymbol{\theta})$，其中

$$L(\boldsymbol{\theta}) = \max\left\{\prod_{i=1}^n p_i:\quad p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta}) = 0\right\},$$

$L(\theta)$为经验似然函数. 事实上，在无限制的情况下每个观察点$\boldsymbol{x}_i$（$i = 1,2,\cdots,n$）使式(1.4)达到最大值时的概率为$p_i = n^{-1}$，而原假设H_0：$E\{\boldsymbol{h}(\boldsymbol{X},\boldsymbol{\theta})\} = 0$对无限制假设的经验似然比统计量是

$$R(\boldsymbol{\theta}) = \max\left\{\prod_{i=1}^n np_i:\ p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta}) = 0\right\}.$$

由于n不依赖于任何参数$\boldsymbol{\theta}$，故极大化$R(\boldsymbol{\theta})$就是使$L(\boldsymbol{\theta})$达到最大值的经验似然估计. 对于给定的$\boldsymbol{\theta}$，如果0在点$\boldsymbol{h}(\boldsymbol{x}_1,\boldsymbol{\theta})$，$\cdots$，$\boldsymbol{h}(\boldsymbol{x}_n,\boldsymbol{\theta})$构成的凸包里，上述的极大值存在且唯一. 通过拉格朗日乘子方法可获得p_i的最优估计：

$$\hat{p}_i = \frac{1}{n\{1 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta})\}},\quad i = 1,2,\cdots,n,$$

式中，拉格朗日乘子$\boldsymbol{\lambda}$满足方程

$$0 = \sum_{i=1}^n \frac{\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta})}{1 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta})}.$$

从而可得关于$\boldsymbol{\theta}$的对数经验似然比函数为

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ln\{1 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta})\}.$$

极小化$\ell(\boldsymbol{\theta})$可获得$\boldsymbol{\theta}$的极大经验似然估计$\hat{\boldsymbol{\theta}}$，即

$$\hat{\boldsymbol{\theta}} = \arg\min_{\theta}\{\ell(\boldsymbol{\theta})\} = \arg\min_{\theta}\left\{\sum_{i=1}^n \ln\{1 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\theta})\}\right\}.$$

1.2 结构方程模型概述

结构方程模型（structural equation model，SEM）是研究可观测变量与潜在变量之间以及潜在变量与潜在变量之间关系的重要工具. 由于在应用科学研究中经常需要处理潜在变量，因此SEM 在近几年发展非常迅速，并广泛应用于生物医学、教育学、行为学、心理学和社会科学等实践性很强的科学研究中. 关于SEM系统的介绍可参见文献（Song et al.，2012）. 在此，通过简单介绍线性SEM使读者了解SEM.

线性SEM由测量方程和结构方程构成. 记$\boldsymbol{y}_i = (y_{i1},\cdots,y_{ip})^{\mathrm{T}}$为第$i$个个体的$p$维可观测随机变量，$\boldsymbol{\omega}_i = (\omega_{i1},\cdots,\omega_{iq})^{\mathrm{T}}$为第$i$个个体的$q$维潜在随机变量. 可观测变量与潜在变量之间的联系由如下测量方程定义：

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i,\ i = 1,\cdots,n, \tag{1.5}$$

式中，$\boldsymbol{\mu}$为截距项向量；$\boldsymbol{\Lambda}$为$p\times q$因子负荷矩阵；$\boldsymbol{\epsilon}_i$为随机误差项，通常假设$\boldsymbol{\epsilon}_i$服从正态分布$N_p(\boldsymbol{0},\boldsymbol{\Psi}_\epsilon)$，与$\boldsymbol{\omega}_i$独立，且$\boldsymbol{\Psi}_\epsilon = \mathrm{diag}(\tau_1^2,\cdots,\tau_p^2)$为对角矩阵. SEM常用于研究潜在变量之间的相互关系. 测量方程通过具体的因子负荷矩阵$\boldsymbol{\Lambda}$将潜在变量与可观测变量关联起来. 因子负荷矩阵$\boldsymbol{\Lambda}$的结构由可观测变量和潜在变量的先验信息确定，其某些元素可在一定的先验信息下设置为已知的固定值. 因此，测量方程是一种验证性工具. 在一定的研究目的下，可以将潜在变量记为$\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^{\mathrm{T}},\boldsymbol{\xi}_i^{\mathrm{T}})^{\mathrm{T}}$，其中，$\boldsymbol{\eta}_i$设为$q_1\times 1$（$0<q_1<q$）内生潜在变量，$\boldsymbol{\xi}_i$设为$q_2\times 1$（$q_2 = q - q_1$）外源潜在变量. 外源潜在变量$\boldsymbol{\xi}_i$对内生潜在变量$\boldsymbol{\eta}_i$的影响可通过结构方程或称为潜在变量模型

$$\boldsymbol{\eta}_i = \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\zeta}_i,\ i = 1,\cdots,n \tag{1.6}$$

给出. 式中，$\boldsymbol{B}$为$q_1\times q_1$系数矩阵，满足$\boldsymbol{I}-\boldsymbol{B}$是非奇异的且$\boldsymbol{B}$的对角元素为零，代表着内生潜在变量对其他内生潜在变量的影响；$\boldsymbol{\Gamma}$为$q_1\times q_2$系数矩阵，反映了外源潜在变量对内生潜在变量的影响；$\boldsymbol{\zeta}_i$为$q_1\times 1$误差向量. 在一定的应用中，$\boldsymbol{B}$和$\boldsymbol{\Gamma}$中的某些元素可以设置为已知的固定值，以便处理潜在变量之间的复杂关系. 有时，SEM也称为因果模型，其通过回归模型的形式来获得因果关系.

在第4章中，我们将提出一个新的SEM，不再考虑验证性的SEM，而考虑探索性的SEM[即第4章中的方程(4.1) 及方程(4.3)]. 我们采用变量选择方法来同时进行参数估计和潜在变量选择，从而识别出潜在变量模型的结构. 第5章中，我们提出一个分位数SEM（记为QSEM），其测量方程的误差项除了假设均值为零外不指定具体分布形式，而其结构方程为分位数回归模型，从而可以研究不同分位数下外源潜在变量对内生潜在变量的影响，在一定程度上解决了传统SEM所遇到的问题. 在所建立的QSEM下，我们基于参数的先验信息提出Bayes（贝叶斯）经验似然（Bayesian empirical likelihood，BEL，Lazow，2003）方法对参数进行估计.

1.3 缺失数据

缺失数据广泛存在于各种领域中，包括公共卫生、医学、经济学及社会科学等. 在过去的几十年中，国内外许多学者都对缺失数据问题进行了卓有成效的研究. 一般地，我们考虑样本容量为n的不完全数据集$\{(\boldsymbol{X}_i, \boldsymbol{Y}_i, \delta_i)$：$i = 1, \cdots, n\}$，其中$\boldsymbol{X}_i$为$d_x$维全观测向量，$\boldsymbol{Y}_i$为$d_y$维可能缺失的观测向量，示性函数$\delta_i = 1$表示$\boldsymbol{Y}_i$可观测，$\delta_i = 0$表示$\boldsymbol{Y}_i$缺失. 记$\mathrm{Z}_{\mathrm{obs}}$表示数据$Z = \{\boldsymbol{X}_i, \boldsymbol{Y}_i\}_{i=1}^n$ 中总能被观测到的数据集，记$\mathrm{Z}_{\mathrm{mis}}$表示数据Z中缺失的数据集. 记条件反应概率为$\Pr(\delta_i = 1|\boldsymbol{X}_i, \boldsymbol{Y}_i)$. 文献（Little et al.，2002）对缺失数据进行了系统性的讨论. 根据此讨论，因为数据的缺失可能与观测数据有关，也可能与缺失数据有关，所以缺失数据的缺失机制可以分为以下3种类型.

(1) 当缺失数据的缺失与数据集中的观测数据和缺失数据都无关系时，此类缺失机制为完全随机缺失（missing completely at random，MCAR），此时，

$$\Pr(\delta = 1|\boldsymbol{X}, \boldsymbol{Y}) = \Pr(\delta = 1).$$

(2) 当缺失数据的缺失仅与数据集中的观测数据有关系而与缺失数据无关系时，此类缺失机制为随机缺失（missing at random，MAR），此时，

$$\Pr(\delta = 1|\ \boldsymbol{X}, \boldsymbol{Y}) = \Pr(\delta = 1|\mathrm{Z}_{\mathrm{obs}}).$$

(3) 当缺失数据的缺失与数据集中的缺失数据本身有关系时，此类缺失机制为非随机缺失（missing not at random，MNAR），此时，

$$\Pr(\delta = 1|\boldsymbol{X}, \boldsymbol{Y}) = \Pr(\delta = 1|\mathrm{Z}_{\mathrm{obs}}, \mathrm{Z}_{\mathrm{mis}}).$$

其中，前两类缺失数据机制与缺失数据本身无关，通常称此类缺失数据为可忽略缺失数据（ignorable missing data）；而第3类缺失数据机制与缺失数据本身有关，这类缺失数据通常称为不可忽略缺失数据（nonignorable missing data）.

对于缺失数据的处理，人们已经提出了很多方法，如EM（expectation-maximization）算法（Dempster et al.，1977）、逆概率加权方法（Horvitz et al.，1952）、增广逆概率加权方法（Robins et al.，1994）及插补法（Rubin，1987；Cheng，1994）. 其中，逆概率加权方法在因果推断中也称为倾向得分调整方法（propensity score adjustment，PSA，Rosenbaum et al.，1983）. PSA方法通过对带有缺失数据的估计方程逆概率加权后进行分析，能得到渐近无偏的估计（Riddles et al.，2016；Wang et al.，2014；Chang et al.，2008；Kott，2006；Tsiatis，2006）. 关于缺失数据问题的全面回顾可参见文献（Little et al.，2002）.

下面对PSA参数估计方法进行简单介绍. 记未知参数$\boldsymbol{\theta}$的真值为$\boldsymbol{\theta}_0$，其由方程$E\{\boldsymbol{\psi}(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\theta}_0)\} = 0$ 决定. 在无缺失的情况下，参数$\boldsymbol{\theta}_0$的相合估计能通过解方程

$$\sum_{i=1}^{n} \boldsymbol{\psi}(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\theta}) = 0 \tag{1.7}$$

获得. 当存在缺失时，估计方程(1.7)就无法进行计算. 此时，可考虑通过解完全观测数据下的估计方程$\left[\text{即} \sum_{i=1}^{n} \delta_i \boldsymbol{\psi}(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\theta}) = 0\right]$来获得参数估计. 这样的估计在MCAR机制下是无偏的，但当MCAR机制不成立时，这样的估计将可能是有偏的. 此时，如果反应概率的真实模型是已知的，记为π_{i*}，则可通过解方程$\sum_{i=1}^{n} \delta_i \pi_{i*}^{-1} \boldsymbol{\psi}(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\theta}) = 0$来获得无偏估计. 一般情况下，$\pi_{i*}$是未知的，此时，可考虑采用条件反应概率$\pi_i = \Pr(\delta_i = 1 | \boldsymbol{X}_i, \boldsymbol{Y}_i)$来代替，通过解方程$\sum_{i=1}^{n} \delta_i \pi_i^{-1} \boldsymbol{\psi}(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\theta}) = 0$来获得参数估计. 条件反应概率也称为倾向得分（propensity score，PS，Rosenbaum et al.，1983），研究者常考虑其参数化形式，即假设反应概率为$\pi_i = \pi(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\gamma})$. 如果参数$\boldsymbol{\gamma}$能被相合地估计，记为$\hat{\boldsymbol{\gamma}}$，则$\boldsymbol{\theta}_0$的估计能通过解方程

$$\sum_{i=1}^{n} \delta_i \hat{\pi}_i^{-1} \boldsymbol{\psi}(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\theta}) = 0 \tag{1.8}$$

获得，式中，$\hat{\pi}_i = \pi(\boldsymbol{Y}_i, \boldsymbol{X}_i; \hat{\boldsymbol{\gamma}})$. 由方程(1.8)所获得的参数估计称为倾向得分调整估

计（propensity-score-adjusted，PSA）.

在缺失数据研究中，很多工作都是在MAR机制假设下进行的，在此假设下可以获得非常合理的理论结果，更重要的是该假设在很多实际问题中是能被合理解释的. 但是，这样的假设并不具有一般性，在某些情况下可能是不合理的. 例如，在关于收入情况的调查及犯罪史的调查中，无响应率常与无响应的数值大小有关. 在此情况下，不能简单地把存在缺失的数据从数据集中删除而只保留完全观测数据来进行分析，也不能用处理可忽略缺失的方法来处理这样的数据，那样得出的结论将会与真实情况相差很大（Little et al.，2002）. 在此情况下，MNAR机制假设可能比传统的MAR 假设更加合理.

1.4 Bayes经验似然方法

由Owen（1988）提出的经验似然（empirical likelihood，EL）方法是一种非参数的似然方法. 在对数据的分布形式不作任何假设的情况下，EL方法可以利用无偏估计方程对参数进行估计和推断，并且能有效地处理估计方程中的过度识别问题. 近年来，EL方法得到了进一步的发展. 例如，Qin et al.（1994）发展了基于估计方程的EL方法并给出参数估计和检验方法的渐近合理性；Molanes et al.（2009）考虑了在非光滑估计方程下的EL方法，发展了相应的计算方法和渐近性质. 关于EL方法全面的回顾可参见文献（Chen et al.，2009；Owen，2001）. 近年来，许多研究者都尝试在Bayes框架下发展EL方法，即BEL（Bayesian empirical likelihood）. Lazar（2003）首先探讨了此问题，她基于Monahan et al.（1992）的工作讨论了BEL推断的有效性. Monahan et al.（1992）提出判断在Bayes推断中一个替代似然是否合适的准则，即一个似然函数被视为合适的Bayes似然当且仅当对于绝对连续的先验分布，后验可信集能达到名义水平. 在此准则下，Lazar考虑了BEL方法的合理性. 然而，她的研究主要是基于蒙特卡罗模拟得出，缺乏概率性的解释. Schennach（2005）提出了Bayes指数倾斜经验似然方法并给出了相应的概率解释. Fang et al.（2006）考虑了单变量总体均值的一类经验似然函数并研究了后验可信集在频率论下覆盖概率的高阶渐近性质. 他们发现在任意data-free（数据不拘）先验下，BEL后验可信集不能达到其名义水平. Mukerjee et al.（2008）发现在data-dependent（数据依赖）先验下BEL后验可信集可以

达到其名义水平. Grendár et al.（2009）基于Bayes大数定理证明了Bayes下使用EL的合理性.

如今，BEL 方法也被用在各类模型中. 例如，在复杂抽样设计中，Rao 和 Wu（2010）提出Bayes伪经验似然并得出后验可信区间的渐近合理性. Chaudhuri 和 Ghosh（2011）对小域估计讨论了BEL方法. 他们考虑了带有随机效应的单参数指数族模型，通过一阶和二阶Bartlett等式构造EL函数，采用分层先验最终得出估计. Yang et al.（2012）在分位数回归模型中讨论BEL方法并得出极大后验估计的相合性及后验分布的渐近正态性. Mengersen et al.（2013）在近似Bayes计算中考虑了EL方法. Vexler et al.（2014）利用BEL方法建立James-Stein估计的非参数形式，并采用Laplace方法建立了所提出的非参后验期望的渐近近似. Chaudhuri et al.（2016）考虑BEL估计，提出一种有效的哈密尔顿蒙特卡罗方法，用于处理具有非凸有界支撑的后验分布情况下传统蒙特卡罗方法收敛慢或不收敛的问题. 这些在BEL下发展的方法不仅继承了传统EL方法的优点，而且便于将可能的多层结构和参数的附加信息结合起来. 然而，BEL方法在很大程度上依赖于先验，Bayes估计和检验的结果可能对先验假设和观测数据很敏感. 因此，在第2章，我们发展一种Bayes局部影响分析方法来对BEL推断进行敏感性分析. 在第3章，我们基于BEL提出Beyes变量选择方法.

接下来对BEL方法进行简单介绍. 设n个来自分布$F_\theta(\boldsymbol{x})$的随机样本为$\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$，其中分布函数$F_\theta(\boldsymbol{x})\in\mathcal{F}_\theta$未知，且依赖于参数$\boldsymbol{\theta}\in\mathcal{R}^p$，$\mathcal{F}_\theta$为一分布族. 假设分布函数$F_\theta(\boldsymbol{x})$ 和参数真值$\boldsymbol{\theta}_0$都未知，仅知道一系列满足条件$E_{F_\theta}\{\boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta})\}=0$的估计函数$\boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta})$ 和参数的先验信息$p_0(\boldsymbol{\theta})$，欲考虑此情况下的估计问题. 首先，利用估计方程和EL方法可给出分布函数$F_\theta(\boldsymbol{x})$的一种非参数估计. 然后，可将这样的似然结合参数先验信息应用于Bayes分析中.

记$F\in\mathcal{F}_\theta$为依赖于$\boldsymbol{\theta}$的分布函数，F的一个非参数似然定义为

$$\mathcal{L}(F)=\prod_{i=1}^{n}\{F(\boldsymbol{x}_i)-F(\boldsymbol{x}_i-)\},$$

其中$F(\boldsymbol{x}_i-)=\Pr(\boldsymbol{x}<\boldsymbol{x}_i)$. 在$\boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta})$约束下极大化$\mathcal{L}(F)$能得到$F_\theta$的EL估计，记为$\hat{F}_\theta\in\mathcal{F}_\theta$. 具体地说，定义$p_i(\boldsymbol{\theta})=F(\boldsymbol{x}_i)-F(\boldsymbol{x}_i-)$，计算

$$\hat{p}(\boldsymbol{\theta}) = \arg\sup\left\{\prod_{i=1}^{n} p_i(\boldsymbol{\theta}) \middle| p_i(\boldsymbol{\theta}) \geqslant 0,\ \ \sum_{i=1}^{n} p_i(\boldsymbol{\theta}) = 1,\ \ \sum_{i=1}^{n} p_i(\boldsymbol{\theta})\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\theta}) = 0\right\},$$

从而获得$\boldsymbol{F}_{\theta}$的估计为$\hat{\boldsymbol{F}}_{\theta}(\boldsymbol{x}) = \sum_{i=1}^{n} \hat{p}_i(\boldsymbol{\theta}) I\{\boldsymbol{x}_i \leqslant \boldsymbol{x}\}$，其中$I\{\cdot\}$为示性函数. 对应于$\hat{\boldsymbol{F}}_{\theta}$的EL函数为$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \hat{p}_i(\boldsymbol{\theta})$. 注意$\hat{p}_i(\theta) \geqslant 0$当且仅当估计函数$\boldsymbol{h}(\boldsymbol{x}_1, \boldsymbol{\theta}), \cdots, \boldsymbol{h}(\boldsymbol{x}_n, \boldsymbol{\theta})$的凸组合包含0. 当此充分必要条件不成立时，习惯性地定义$L(\boldsymbol{\theta}) = 0$. $\hat{p}_i(\boldsymbol{\theta})$的计算涉及带约束的极大化问题，通过拉格朗日乘子方法可获得其最优估计，即

$$\hat{p}_i(\boldsymbol{\theta}) = \frac{1}{n\{1 + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\theta})\}}, \tag{1.9}$$

式中，$\boldsymbol{\lambda}(\boldsymbol{\theta})$为拉格朗日乘子且满足方程

$$\sum_{i=1}^{n} \frac{\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\theta})} = 0. \tag{1.10}$$

从而有$L(\theta) = \prod_{i=1}^{n} [n\{1 + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\theta})\}]^{-1}$. 有了$L(\boldsymbol{\theta})$和先验$p_0(\boldsymbol{\theta})$，便能定义后验分布

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{L(\boldsymbol{\theta})p_0(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta})p_0(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}.$$

有了后验分布继而可以进行Bayes推断.

第 2 章　带有不可忽略缺失数据的估计方程的Bayes局部影响分析

2.1　引言

缺失数据广泛存在于各种领域中，包括公共卫生、医学、经济学及社会科学等. 在缺失数据分析中，当缺失数据的缺失概率仅与数据集中的观测数据有关而与缺失数据无关时，缺失机制为随机缺失（Little et al.，2002）. 尽管缺失数据经常被假设为MAR，但是这样的假设在某些情况下是有问题的，如在关于收入的调查或犯罪史的调查中，无响应率常与无响应的数值大小有关. 在此情况下，不可忽略缺失机制假设可能比传统的MAR假设更加合理.

近年来，很多统计方法已被提出，并用于对MNAR 数据进行分析，例如，Lee et al.（2006）提出Bayes方法来对带有MNAR数据的非线性结构方程进行分析；Kim et al.（2011）提出指数倾斜模型和半参数估计方法来获得存在MNAR数据的数据均值估计；Zhao et al.（2013a）针对带有MNAR数据的均值统计推断发展一种经验似然（EL）方法；Zhao et al.（2013b）提出非参数及半参数估计方法和增广逆概率加权插补方法来估计带有MNAR响应变量的分布函数和分位数；Tang et al.（2014）基于具有指数倾斜模型形式的反应概率和广义估计方程（estimating equations，EEs）发展EL方法；Linero et al.（2015）针对具有MNAR机制的纵向数据建立Bayes非参数模型；Jiang et al.（2016）针对带有MNAR缺失协变量的线性回归模型的参数估计提出复合分位数回归方法. 以上所提及的工作主要是仅基于Bayes方法或者仅基于EL方法来发展的.

对于MNAR数据，在Bayes框架下发展EL方法（即Bayes经验似然，Lazar，2003）引起我们很大的兴趣. 如今，针对BEL推断的工作已有很多（Grendár et al.，2009；Chang et al.，2008；Fang et al.，2006）. Yang et al.（2012）发现BEL方法能利用先验信息提高参数估计的有效性和帮助寻找极大经验似然估计