

广义估计方程 估计方法

周 勇 著



科学出版社

014008983

021
389

现代数学基础丛书 150

广义估计方程估计方法

周 勇 著



本书得到中国科学院出版基金资助，同时部分获得国家杰出青年基金（编号：70825004）、国家自然科学基金重点资助项目（编号：10731010）和国家自然科学基金委创新研究群体科学基金（编号：10721101）的资助，并获得上海财经大学“211 工程”四期重点学科建设项目和上海市一流学科布局建设项目资助

科学出版社

北 京



北航

C1696008

021
389

内 容 简 介

估计方程方法是统计推断中最为普通但也非常有用的统计推断方法之一,其思想和结果广泛应用于生存分析、生物统计、计量经济及金融计量中.本书共分20章和两个附录,着重讨论参数模型、时间序列模型、非参数模型、半参数及变系数模型等模型中有关估计方程的统计推断方法,并讨论几种现代统计学中通常遇见的数据类型下估计方程方法.这是目前新的统计推断方法,主要包括最为常见的独立样本、非独立样本(时间序列样本)、纵向数据、缺失数据、缺失数据等下的估计方程方法.本书总结了一批应用估计方程方法进行统计推断的统计模型,同时也概括了可以应用一般估计方程方法处理的复杂数据.书中内容除了数理统计的常用统计推断方法外,也包括最新统计方法有关估计方程估计的研究成果.

本书适合大学数理统计专业、数学专业和计量经济学专业等高年级的学生做选修课程,同时也可作为是数理统计、统计学以及计量经济学等专业研究生专业基础课教程.另外,本书还适合各行各业的应用数理统计科学工作者,主要涉及经济、金融、社会学、心理学、生物医学和工业工程等专业人士.

图书在版编目(CIP)数据

广义估计方程估计方法/周勇著. —北京: 科学出版社, 2013

(现代数学基础丛书; 150)

ISBN 978-7-03-038641-0

I. ①广… II. ①周… III. ①数理统计-研究 IV. ①O21

中国版本图书馆 CIP 数据核字(2013) 第 223510 号

责任编辑: 陈玉琢 / 责任校对: 张怡君

责任印制: 赵德静 / 封面设计: 陈 敬

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>

源海印刷有限责任公司 印刷

科学出版社发行 各地新华书店经销

*

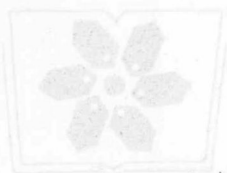
2013 年 8 月第 一 版 开本: 720 × 1000 1/16

2013 年 8 月第一次印刷 印张: 31 1/4

字数: 602 000

定价: 138.00 元

(如有印装质量问题, 我社负责调换)



现代数学基础丛书

《现代数学基础丛书》编委会

主 编：杨 乐

副主编：姜伯驹 李大潜 马志明

编 委：（以姓氏笔画为序）

王启华 王诗宓 冯克勤 朱熹平

严加安 张伟平 张继平 陈木法

陈志明 陈叔平 洪家兴 袁亚湘

葛力明 程崇庆

《现代数学基础丛书》序

对于数学研究与培养青年数学人才而言,书籍与期刊起着特殊重要的作用.许多成就卓越的数学家在青年时代都曾钻研或参考过一些优秀书籍,从中汲取营养,获得教益.

20世纪70年代后期,我国的数学研究与数学书刊的出版由于文化大革命的浩劫已经破坏与中断了10余年,而在这期间国际上数学研究却在迅猛地发展着.1978年以后,我国青年学子重新获得了学习、钻研与深造的机会.当时他们的参考书籍大多还是50年代甚至更早期的著述.据此,科学出版社陆续推出了多套数学丛书,其中《纯粹数学与应用数学专著》丛书与《现代数学基础丛书》更为突出,前者出版约40卷,后者则逾80卷.它们质量甚高,影响颇大,对我国数学研究、交流与人才培养发挥了显著效用.

《现代数学基础丛书》的宗旨是面向大学数学专业的高年级学生、研究生以及青年学者,针对一些重要的数学领域与研究方向,作较系统的介绍.既注意该领域的基础知识,又反映其新发展,力求深入浅出,简明扼要,注重创新.

近年来,数学在各门科学、高新技术、经济、管理等方面取得了更加广泛与深入的应用,还形成了一些交叉学科.我们希望这套丛书的内容由基础数学拓展到应用数学、计算数学以及数学交叉学科各个领域.

这套丛书得到了许多数学家长期的大力支持,编辑人员也为其付出了艰辛的劳动.它获得了广大读者的喜爱.我们诚挚地希望大家更加关心与支持它的发展,使它越办越好,为我国数学研究与教育水平的进一步提高做出贡献.

杨 乐

2003年8月

前 言

估计方程方法是统计推断中最为普通但同时也是非常有用的统计推断方法之一,其思想和方法广泛应用于生存分析、生物统计、工程统计、管理统计、计量经济学,以及金融计量学等学科领域中.本书着重讨论参数模型、时间序列模型、非参数模型、半参数及变系数模型等模型中有关估计方程的统计推断方法,并讨论几种现代统计学中经常遇到的数据类型下的估计方程估计方法.这些是目前较新的统计推断方法,主要包括最为常见的独立样本、时间序列样本、纵向数据、删失数据、缺失数据等复杂数据下的估计方程估计方法.一般地,估计方程的估计可以看作是最大似然估计和矩估计方法等古典概率统计推断方法的推广,并涉及 GMM 估计、拟似然估计、伪似然估计和广义最小二乘估计、经验似然估计等.结合各种数据类型和相适应的统计模型,深入地展示这些统计推断的理论和方法以及它们在实际中的应用.本书理论与实践相结合,灵活地运用估计方程方法,使学习者能够在学习本书的过程中,发现统计方法的奇妙和无穷魅力.同时能通过掌握这些方法使学习者能在较短的时间里追上国际前沿,以及熟练地应用这些方法,甚至在统计学的方法及理论上进行创新研究.

本书主要涉及概率统计、生存分析、生物医学统计、计量经济学和金融风险管理等常用的统计模型及其方法,本书总结了一系列应用估计方程估计方法进行统计推断的统计模型,并概括了可以应用一般估计方程估计方法处理的复杂数据.书中除了常用的传统统计推断方法外,也包括最新的有关估计方程估计的研究成果.

本书适合大学数理统计专业、数学专业和计量经济学等专业高年级本科生选修课程,同时也可作为数理统计、统计学以计量经济学等专业方向的研究生基础课程.另外,本书还适合各行各业应用数理统计科学工作者和经济、金融、社会学、心理学、生物医学和工业工程等专业人士.

全书分为 20 章.第 1 章绪论,对全书的内容进行了概述,并给出一些基本结果,以使读者能很快地了解全书的内容和所涉及的理解和方法.第 2 章和第 3 章是准备知识,主要介绍一些估计方程方法的相关基础知识和常用的数据类型,以及统计大样本的常用极限理论,涉及大数律、中心极限定理,以及一致大数律等.如果读者对理论不感兴趣,则只需知道结论而跳过理论证明即可.第 4 章介绍的 Delta 方法是一种常用工具,在推断参数函数估计的大样本性质时,特别是渐近方差时,非常有用和有效.

第 5 章介绍两种经典的统计方法,即矩估计和极大似然估计方法.这两种方法

在一般的教科书上都有介绍,但并不十分全面.本书旨在通过更为全面的介绍,让读者对统计推断方法有更深入的理解.同时,本书的其他章节都是以极大似然估计方法为参照,应用估计方程的过程与应用极大似然估计方法具有相似性.首先是参数估计,其次构造置信区间或进行假设检验,理论结果也类似.当然不同数据类型和模型下结论各异,但是,掌握了最基础的极大似然估计将有助于理解书中其他复杂的方法.另外本章还介绍了一些复杂数据下的极大似然估计.

第6章的极值目标函数估计方法是估计方程估计方法的一般化,但它有独立的意义,因此,有关极值目标函数估计可自成一章,让读者能够很容易掌握处理估计方程的一般方法.在第9章将提供一些选择最优估计函数的框架和方法.

第7章介绍了经验似然估计方法及其在估计方程中的统计推断问题.经验似然方法是处理过识别估计方程组最有力的两种工具之一.本章也介绍了一些较新的复杂数据下的经验似然方法,感兴趣的读者如果学习完这章,已足够进入统计的研究前沿,并能进行相关的研究工作.

第8章是极大似然方法的扩展——伪似然方法.拟似然方法不像极大似然方法需要知道真实的分布函数,从而避免或减小了模型误判的问题.拟似然方法是处理估计方程问题的一种重要方法,其思想具有一般性.

第9章介绍了广义估计方程估计方法及这些估计良好的大样本性质,还讨论了广义估计方程估计的最优方差,即最优估计方程等问题,本章总结了最优估计函数的思想及方法,它是本书的重要基础之一.

第10章是估计方程方法的一般思想,是本书的重点之一,讨论了如何构造无偏估计函数来获得估计方程,并通过几个实用的模型来理解利用广义估计方程估计方法建模的问题,同时也涉及了最优估计方程的一般思想.本章还重点探讨了生存分析中几类重要模型构造估计方程的思想方法,并指出估计方程方法在生存分析中具有广泛的应用.

第11章讨论了广义线性模型的一般理论和估计方法.简单介绍了指数分布族及在指数分布族下广义线性模型的极大似然估计,以及在非指数族分布族下,给出了广义线性模型中参数的拟极大似然估计,并讨论了拟极大似然估计与估计方程的关系.

第12章讨论了纵向数据下的估计方程方法.利用估计方程估计方法可以有效处理纵向数据的相关性,并给出工作方差对估计的影响.

第13章介绍了很广泛的一类非参数估计方程,可以包括半参数模型、变系数模型等,主要通过局部光滑技术有效地得出参数或非参数的估计方程估计.这里给出了非参数估计方程的统计推断的一般方法,并应用于一个重要的实际例子中.

第14章首先介绍非参数下的广义线性模型,并扩展了非参数估计方程的估计方法到半参数估计方程的领域.构造半参数估计方程有效地处理复杂模型的统计推

断问题. 本章较难, 可以跳过.

第 15 章讨论非参数估计方程的方法及其在时间时序列模型中应用. 在前面第 9 章选择最优无偏估计函数的意义下, 讨论时间序列数据下非参数估计方程, 给出了推导出非参数估计方程的系数估计的一般方法.

第 16 章主要讨论删失数据下的估计方程方法, 在此, 我们提供了一种新的估计方程估计方法, 这种方法能有效地处理复杂删失数据的影响. 同时把相关的方法应用到一个医疗费用估计的实际例子, 通过在考虑删失数据特征的情况下构造估计方程, 并分割观察时间区间以便可以获得更多的有用信息, 从而提高估计的效率.

第 17 章讨论两样本带有删失数据的一类估计方程的统计推断问题, 这里涉及半参数模型. 两样本问题在评价治疗影响, 差异分析中具有重要的作用. 构造了带有删失数据下两样本的参数估计方程, 用经验似然方法很容易地给出参数的推断, 并且本章中介绍的方法比通常正态逼近方法更好. 这里也给出了一些实际例子.

第 18 章讨论带有删失数据而目标函数不可微情形下如何寻找估计方程估计问题, 主要涉及分位函数估计和 ROC 曲线估计问题. 利用核光滑的思想对 ROC 曲线关于讨厌参数不连续进行处理, 从而有效地给出两样本下 ROC 曲线的统计推断.

第 19 章讨论缺失数据下一般估计方程估计方法. 与删失数据相似, 缺失数据也是一类非常重要的数据, 在各应用学科中广泛存在. 在本章我们介绍最新的非逆概率权估计方程的估计方法, 即通过回归函数的核光滑估计构造渐近无偏的估计方程, 克服缺失数据的影响, 给出参数的估计.

第 20 章利用估计方程方法对分位数和分位数回归进行了统计推断. 在带有缺失数据的影响下, 构造出一个渐近无偏的分位数估计方程, 和多重插补估计方程, 此构造方法不需要逆概率加权便能够获得分位数和分位数回归在缺失数据下的估计.

最后有两个附录, 分别是生存分析的基础知识和非参数回归技术的补充材料, 这些内容在全书中都有应用.

在本书的写作和修改过程中得到了中国科学院数学与系统科学研究院许多同事和上海财经大学部分老师的帮助, 特别要感谢我的导师安鸿志研究员, 以及美国普林斯顿大学范剑青教授多年以来对我的关心和支持, 并感谢他们引导我在博士和博士后期间学习了许多现代统计学的理论和方法以及统计思想. 还要感谢严加安院士和马志明院士一直以来对我科研工作的支持和帮助. 同时, 本书的形成过程也得到了中科院数学与系统科学研究院汪寿阳研究员、美国佛罗里达大学惠灵顿工商管理学院经济系终身教授艾春荣博士等的帮助和有益讨论, 并对本书的写作提供了许多有建设性意见, 上海财经大学“千人计划”入选者黄坚教授、中国科学院孙六全研究员对本书的构架和内容取舍提供了许多非常有益的建议. 在本书的写作过程中还得到许多同行的详细阅读和修改建议, 感谢上海财经大学尤进红博士提供一些

有用的参考文献,还要感谢李元教授、刘力平教授和张忠占教授、王绍立博士、朱利平博士、柏杨博士等对本书进行有益的讨论。

同时感谢中国科学院博士研究生马昀蓓、谢尚宇、刘旭、硕士生栾清淑和上海财经大学硕士生袁媛对本书的校订,并提出有益的建议。感谢中国科学院博士生林存洁、刘鹏和上海财经大学博士生张莉多次对本书的校对,做了辛苦的工作。另外感谢中国科学院博士研究生赵目、田军、白芳芳和硕士生王晓婧、刘沛欣,上海财经大学博士生刘晓倩、张飞鹏、陈柏成、李小莉、孟美侠、刘玉涛、王艺馨、邱志平、杨广仁等,中国科学技术大学研究生姚宏伟和云南大学博士生陈雪蓉等同学为本书部分章节的录入和审校做了很多工作,以及在“广义估计方程理论与方法”讨论班上的所有同学,他们对本书在讨论班上讨论时提出了修改意见,在此一并表示感谢。

本书仍有许多不足,由于本人的能力有限,相信存在一些地方的表达可能并不很合适,另外,由于本人的笨拙,本书在写作过程肯定还有对其他人相当重要的研究成果未能充分展示出来。如果本书有什么错误和不妥之处,本人一律承担,并希望读者批评指正。

本书的完成也得到中国科学院出版基金资助和国家杰出青年基金、国家自然科学基金、国家自然科学基金委创新研究群体项目支持,以及上海财经大学“211工程”四期重点学科建设项目和上海市一流学科布局建设项目支持。

最后感谢我的夫人和孩子周于皓长期以来对我科研工作的支持和理解,特别是在本书写作过程的相当长时间里他们在背后默默的支持让我有足够的时间来完成本书的撰写。

周 勇

于中国科学院数学与系统
科学研究所和上海财经大学

2013 年 4 月

目 录

《现代数学基础丛书》序

前言

第 1 章 绪论	1
1.1 估计方程估计方法概述	1
1.2 统计模型与估计方程	4
1.3 带有辅助信息的估计方程估计	8
1.4 估计方程估计的渐近性质概述	11
1.5 广义估计方程估计相合性	15
第 2 章 数据类型	17
2.1 简单数据	17
2.2 时间序列数据	18
2.3 删失数据	18
2.4 缺失数据	21
2.5 纵向数据 (面板数据)	25
第 3 章 准备知识	27
3.1 随机变量序列收敛性	27
3.2 大数律与中心极限定理	31
3.2.1 弱大数律和强大数律	31
3.2.2 重对数律	38
3.2.3 中心极限定理	39
3.2.4 估计的大样本性质	43
3.3 一致大数律及经验过程	46
3.4 一般极限定理	51
3.5 其他一些收敛定理	62
第 4 章 Delta 方法	65
4.1 Delta 方法的思想	65
4.2 向量估计函数 Delta 方法	67
4.3 相关研究及扩展	71

第 5 章 矩估计与极大似然	72
5.1 矩估计	72
5.2 极大似然估计	79
5.3 极大似然估计理论	83
5.4 信息阵及 C-R 不等式	85
5.5 有关极大似然估计的假设检验	95
5.6 删失数据下极大似然估计	101
5.7 截断数据极大似然	103
5.8 缺失数据极大似然估计	104
5.9 不可忽略缺失机制下的极大似然估计	105
5.10 条件似然估计	106
5.11 相关研究及扩展	109
第 6 章 极值目标函数估计	111
6.1 广义估计方程估计	111
6.2 极值目标函数估计	114
6.3 极值函数估计量的存在性与可测性	117
6.4 几类重要的极值函数估计	119
6.5 极值函数估计的相合性与渐近正态性	121
6.6 渐近方差估计	125
6.7 极值函数估计统计推断: 拉格朗日检验及置信区间	126
6.8 主要结果证明	128
6.9 补充材料	129
第 7 章 经验似然及估计方程	130
7.1 经验似然的基本思想及概念	130
7.2 一维均值经验似然	135
7.3 多维均值经验似然	137
7.4 估计方程经验似然推断	144
7.5 有偏抽样经验似然	150
7.6 相关研究及拓展	152
7.7 主要定理的证明	153
第 8 章 伪极大似然	160
8.1 伪极大似然估计及推断	160
8.2 分布误判及伪似然估计	162
8.3 伪似然估计相合性的充要条件	164
8.4 关于伪似然估计的假设检验	172

8.5	小结及讨论	175
8.6	补充材料	175
第 9 章	估计方程估计的渐近理论	176
9.1	广义估计方程估计	176
9.2	广义估计方程估计的存在性	177
9.3	估计方程估计的相合性	179
9.4	估计方程估计的渐近正态性	182
9.5	渐近方差估计	183
9.6	渐近有效性	183
9.7	最优估计函数	187
9.7.1	估计函数与高斯-马尔可夫定理	187
9.7.2	得分函数	192
9.8	最优估计方程的一般框架	193
9.8.1	小样本情形下的最优准则	194
9.9	补充材料	198
第 10 章	估计方程的一般思想	200
10.1	估计函数寻找方法	201
10.2	单估计方程	202
10.3	多元估计方程	204
10.4	辅助信息线性模型	205
10.4.1	广义矩估计	209
10.4.2	经验似然估计	210
10.5	带有辅助信息分布估计	213
10.6	传染模型	214
10.7	非线性回归模型	217
10.7.1	无偏估计函数构造方法	217
10.7.2	GEE 估计方法的定义	218
10.7.3	权矩阵的选择	218
10.7.4	估计的渐近性质	219
10.7.5	GEE 方法的步骤	219
10.8	生存分析中的 Cox 模型	220
10.8.1	变系数 Cox 模型	222
10.9	均值剩余寿命模型	224
10.10	复发数据模型	228
10.11	长度偏差数据模型	229

10.12	相关研究与扩展	231
10.13	附录	233
第 11 章	指数族及广义线性模型	239
11.1	指数族	239
11.1.1	简单指数族	239
11.1.2	带有协变量的指数族	241
11.2	广义线性模型	242
11.3	极大似然估计	244
11.3.1	估计方程	244
11.4	参数推断	248
11.4.1	渐近方差估计	250
11.4.2	假设检验	251
11.4.3	拟合优度检验	251
11.5	拟似然估计	252
11.5.1	拟似然的基本模型	252
11.6	拟似然与估计方程	259
11.7	局限性	260
11.8	相关研究及扩展	262
11.8.1	相关研究	262
11.8.2	进一步的讨论	263
第 12 章	纵向数据估计方程	264
12.1	引言	264
12.2	纵向数据下 GMM 方法	264
12.3	经验似然方法	268
12.3.1	工作独立经验似然	269
12.3.2	块经验似然	270
12.4	纵向数据下的广义线性模型	273
12.5	工作独立估计方程	276
12.6	协方差矩阵参数化	277
12.7	冗余参数估计	278
12.7.1	可交换相关系数矩阵	279
12.7.2	时间序列相关系数矩阵	280
12.8	固定影响和随机影响模型	280
12.8.1	无条件固定影响模型	281
12.8.2	条件固定影响模型	281

12.8.3 随机影响模型	283
12.9 模拟结果	285
12.9.1 线性模型场合	285
12.9.2 非线性模型场合	285
12.10 定理的证明	286
12.11 相关研究及扩展	290
第 13 章 非参数估计方程	292
13.1 非参数估计方程	293
13.2 局部多项式拟合	294
13.2.1 局部多项式拟合的一般方法	294
13.2.2 核函数选择	296
13.2.3 窗宽选择	296
13.3 非参数估计收敛性	297
13.4 局部估计方程的其他进展	299
13.5 变系数回归模型的估计方程	302
13.6 一个例子: 变系数生产函数	304
13.6.1 模型建立及求解	305
13.6.2 弹性系数时变性的广义似然比检验	306
13.6.3 实证研究: 中国时变弹性系数生产函数	307
13.6.4 进一步的讨论	309
第 14 章 非参和半参局部拟似然估计	310
14.1 非参数局部拟似然估计	310
14.2 半参数局部拟似然估计	312
14.3 半参拟似然估计的渐近性质	317
14.4 补充材料	317
第 15 章 非参数时间序列估计方程方法	325
15.1 随机系数估计方程	325
15.2 时间序列基本模型	328
15.3 GEE 方法在非参数时间序列模型中的几个应用	330
15.3.1 随机系数自回归模型 (RCAR)	330
15.3.2 双重随机时间序列模型	331
15.3.3 门限自回归模型	333
15.3.4 特殊情况	333
15.4 一些扩展	334
15.4.1 广义最小二乘法	335

15.4.2	条件最小二乘法	335
15.4.3	分枝过程	335
第 16 章	删失数据下估计方程	337
16.1	无偏估计函数	337
16.2	医疗费用的估计方法简述	342
16.3	经验似然估计及置信区间	343
16.3.1	剖面经验似然比函数	343
16.3.2	置信区间构造	344
16.4	工作独立经验似然方法	347
16.5	边际似然方法	351
16.5.1	恰好识别情形	351
16.5.2	过度识别情形	352
16.6	真实数据应用	355
16.7	进一步讨论	356
16.8	补充材料	356
16.9	医疗费用研究相关文献及扩展	361
第 17 章	两样本估计方程	363
17.1	两样本估计方程的治疗影响	363
17.2	两样本删失数据	364
17.2.1	正态方法	365
17.2.2	经验似然方法	367
17.3	真实数据应用	370
17.4	相关研究及扩展	370
17.5	补充材料	372
第 18 章	光滑经验似然	378
18.1	引言	378
18.2	基于正态方法	379
18.3	光滑经验似然法	381
18.4	相关研究及扩展	382
18.5	补充材料	383
第 19 章	缺失数据估计方程	394
19.1	缺失数据估计方程	394
19.2	核光滑填入法	396
19.3	参数统计推断	397
19.3.1	GEE 估计与经验似然估计	397

19.3.2	估计的渐近性质	399
19.3.3	辅助信息及有效性改进	400
19.3.4	渐近方差估计	401
19.3.5	调整经验似然估计	402
19.4	数据维数减少原则	402
19.5	真实数据例子	403
19.5.1	杜兴肌营养不良症 (duchenne muscular dystrophy) 数据	403
19.5.2	虫蛀水果数据	404
19.6	相关研究与扩展	405
19.6.1	相关研究	405
19.6.2	本章方法的进一步讨论	406
19.7	定理的证明	407
第 20 章	缺失数据下分位数回归	414
20.1	基于估计方程的缺失数据下的样本分位数回归	414
20.2	缺失数据下的非参核插补法	416
20.2.1	非参核插补法下分位数估计的渐近性质	417
20.2.2	非参核插补法下分位数估计的渐近方差估计	418
20.3	缺失数据下的局部多重插补法	419
20.3.1	局部多重插补法下分位数估计的渐近性质	420
20.3.2	局部多重插补法下分位数估计的渐近方差的估计和窗宽选择	420
20.4	缺失数据下的分位数回归	421
20.4.1	核插补法	421
20.4.2	局部多重插补法	422
20.5	相关研究及扩展	423
20.6	定理的证明	424
20.6.1	缺失数据下样本分位数定理证明	424
20.6.2	缺失数据下线性分位数回归定理证明	431
附录 A	计数过程及其鞅理论	436
A.1	计数过程	436
A.2	鞅理论	437
A.3	风险率函数与生存分布	439
附录 B	非参数回归	441
B.1	非参数回归估计	441
B.2	局部线性估计	441
B.3	局部多项式回归	446

B.3.1	提出估计	446
B.3.2	局部多项式估计的偏差及方差	447
B.3.3	窗宽选择	448
B.3.4	核函数	448
B.3.5	补充	449
参考文献		452
索引		470
《现代数学基础丛书》已出版书目		476