□ 应用统计学丛书

Large Dimensional Statistical Analysis

大维统计分析

白志东　郑术蓉　姜丹丹

高等教育出版社
HIGHER EDUCATION PRESS

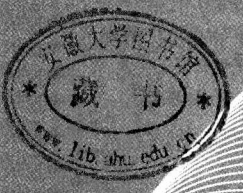
□ 应用统计学丛书

Large Dimensional Statistical Analysis

大维统计分析

白志东　郑术蓉　姜丹丹

DAWEI TONGJI FENXI



高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

图书在版编目(CIP)数据

大维统计分析/白志东，郑术蓉，姜丹丹编著. --
北京：高等教育出版社，2012.5
ISBN 978-7-04-034830-9

Ⅰ.①大… Ⅱ.①白… ②郑… ③姜… Ⅲ.①统计
分析 Ⅳ.①C812

中国版本图书馆 CIP 数据核字(2012)第 063747 号

策划编辑 王丽萍　　责任编辑 李华英　　封面设计 王凌波　　版式设计 于 婕
责任校对 金 辉　　责任印制 朱学忠

出版发行 高等教育出版社
社　　址 北京市西城区德外大街 4 号
邮政编码 100120
印　　刷 保定市中画美凯印刷有限公司
开　　本 787mm × 1092mm 1/16
印　　张 34
字　　数 700 千字
购书热线 010 - 58581118

咨询电话 400 - 810 - 0598
网　　址 http://www.hep.edu.cn
http://www.hep.com.cn
网上订购 http://www.landraco.com
http://www.landraco.com.cn
版　　次 2012 年 5 月第 1 版
印　　次 2012 年 5 月第 1 次印刷
定　　价 79.00 元

物 料 号 34830 - 00

前　言

在一元统计分析中, 对每个样本个体只观测一个指标值。如果个体包含多重特性时, 一元统计很难做出比较精确的推断。例如研究小孩身体发育, 如果仅观测样本中每个小孩的身高, 则仅根据身高信息, 很难对小孩的身体发育情况得到精确的结论。因此在实际问题中通常要观测每个样本个体的多个指标值, 进行多元统计分析。如研究小孩的身体发育, 我们需要观测样本中每个小孩的身高、体重、腰围、胸围等多个指标值。多个指标蕴含的信息量大, 有助于做出精确的统计推断。因此, 我们需要将一元统计中发展起来的统计推断理论推广到多元场合。

另外, 需要指出的是, 在 20 世纪 40 年代, 即多元分析理论建立和发展的时期, 因为实际问题中收集的样本量很小, 与一元统计相似, 多元分析也是在多元正态分布假定下发展起来的。这是因为在正态分布假定下很多理论都可以得到精确的结果, 如统计量的精确分布等。为保持这一历史事实, 也为了小样本统计的需要, 我们对于多元分析的基本理论, 也在多元正态的假定下来介绍。但是, 随着科学技术的发展, 实际问题中能收集到大量的样本, 例如, 在无线电电子学中, 1 秒钟可以收集 500 个样本; 在股票市场中, 1 秒钟也能收集数百个数值。以前因为样本量小不能解决的问题就可以在大样本下得到渐近的结果。特别是近数十年来, 随着计算机技术的飞速发展和广泛应用, 人们得以搜集、储存和处理大维巨量数据。例如, 金融市场上的资料的维数可以高达数千或数万; 手机通讯资料的维数可达百万乃至千万计; 生物学中关于 DNA 的资料的维数更是大到难以想象。高速计算技术的发展使得人们可以对大维巨量数据进行计算和统计推断。它给我们带来了巨大的好处, 同时, 也带来了巨大的挑战：经典的极限理论是在维数固定, 而样本容量趋于无穷的假定下得到的。首先, 许多经典的多元统计方法应用于大维资料时, 它们或者根本不可以应用, 如点估计的偏差很大, 第一类错误可能会趋于 1 等, 这在点估计和假设检验中是根本不可能被接受的; 或者这些经典方法即使可以应用, 其功效也会很低。其次, 当维数不是很大时, 人们早就创立了各种降维方法, 例如变量选择、主成分分析、因子分析等。

但是这些方法根本不能适用于维数非常大的情况。作为本书的特点, 除基本理论部分外, 对于各种多元统计方法, 我们尽可能引进一些传统统计分析方法的修正, 使之更适用于多元统计分析。由于许多修正方法依赖于大维随机矩阵的谱分析理论, 所以我们在附录中简要介绍该理论中的一些主要结果。

本书面向统计学专业及与统计学相关专业的本科生、硕士生、博士生和科研人员。对于本科生来说, 我们建议讲授如下章节: 2.1 ~ 2.4, 3.1 ~ 3.2, 5.1 ~ 5.3, 6.1 ~ 6.5, 11.1 ~ 11.2, 12.1 ~ 12.3, 14.1 ~ 14.7。对于硕士生建议增加章节: 2.5, 3.3 ~ 3.5, 4.1 ~ 4.4, 5.4 ~ 5.6, 6.6 ~ 6.9, 7.1 ~ 7.9, 8.1 ~ 8.10, 9.1 ~ 9.6, 10.1 ~ 10.8, 11.3 ~ 11.4, 12.4 ~ 12.8, 13 章和 14 章。博士生除讲授上述内容外, 建议自学所有章节, 特别注意各章中的大维统计分析部分。

在此真诚地感谢东北师范大学张宝学教授、胡果荣博士、宋海燕博士、胡江博士在本书完成过程中所做的工作。

作 者

2011 年 9 月 7 日

目 录