

中国科学技术大学数学教学丛书

数理统计

(第二版)

韦来生 编著



科学出版社

中国科学技术大学数学教学丛书

数理统计

(第二版)

韦来生 编著

科学出版社

北京

内 容 简 介

本书是数理统计学专业的基础课教材。内容包括绪论、抽样分布及若干预备知识、点估计、区间估计、参数假设检验、非参数假设检验、Bayes 方法和统计决策理论等 7 章，各章都配备了习题。

本书可作为综合性大学、理工科院校和师范院校概率论与数理统计(简称概统)专业本科生的“数理统计”课的教材或参考书。适当删除书中标“*”的章节，可作为上述相关院校数学系非概率统计专业本科生的“数理统计”教材或参考书。具备微积分、矩阵代数及概率论基本知识的读者皆可使用本书。本书也可作为相关院校研究生、青年教师以及从事统计工作的工程技术人员的参考书。

图书在版编目(CIP)数据

数理统计/韦来生编著。—2 版。—北京：科学出版社，2015.12

(中国科学技术大学数学教学丛书)

ISBN 978-7-03-046573-3

I. ①数… II. ①韦… III. ①数理统计-高等学校-教材 IV. ①O212

中国版本图书馆 CIP 数据核字(2015) 第 288627 号

责任编辑：姚莉丽 / 责任校对：邹慧卿

责任印制：霍 兵 / 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>

大厂书文印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2008 年 7 月 第一 版 开本：720 × 1000 1/16

2015 年 12 月 第二 版 印张：23

2015 年 12 月 第八 次印刷 字数：463 000

定价：46.00 元

(如有印装质量问题，我社负责调换)

《中国科学技术大学数学教学丛书》编委会

主 编 程 艺

顾 问 (按汉语拼音排序)

陈希孺 方兆本 冯克勤 龚 昇 李翊神

石钟慈 史济怀

编 委 陈发来 陈 卿 陈祖墀 侯定丕 胡 森
蒋继发 李尚志 林 鹏 刘儒勋 刘太顺
缪柏其 苏 淳 吴耀华 徐俊明 叶向东
章 璞 赵林城

第二版前言

《数理统计》第一版出版至今已有 7 年, 根据作者在中国科学技术大学教学实践中的体会和学生提出的问题与建议, 并借鉴国内外同类教材一些成功的经验, 这次再版对第一版的部分内容进行了增补、删减和修改, 以便更好地适应数理统计教学的需要.

第二版与第一版相比主要有如下变化:

首先, 增加了一小部分重要内容. 例如, 在 2.5 节增加了引理 2.5.1, 它在区间估计和假设检验的大样本问题中发挥了重要的作用. 在第 3 章末增加了 3.6 节“概率密度函数及其导数的核估计”, 作为点估计的非参数方法提供给读者阅读、了解. 第二版中还将 5.3 节和 5.5 节进行了对调. 在介绍了似然比检验后, 再讲解“N-P 引理”学生可能更容易接受.

其次, 第二版删减了一小部分可放入研究生阶段学习的内容. 例如, 在 3.5 节中删去“C-R 不等式等号成立的条件”这个小节的主要内容, 只保留了对主要结论的简介. 在 3.5 节中还删去了“渐近效率”的定义及例子. 在第 5 章末删去了 5.6 节“序贯概率比检验”的全部内容. 在 6.2 节和 6.3 节删去了有关“置换检验”的内容.

第二版对一些重要的例子作了补充和修改, 使得结果更完善. 例如, 在例 3.3.5、例 3.4.8 和例 3.4.10 等例题中, 对解题方法作了修改和补充, 使得解题方法多样化、更合理. 在 4.2 节的一些求置信区间的例子中增加了求置信上、下限的内容; 在 6.2 节符号检验与符号秩和检验的例子中增加了如何利用检验的 p 值对检验问题作出结论的内容.

第二版对各章习题作了适当的增补和删减, 对部分习题的先后次序作了调整, 使用起来更方便.

这里仍需说明的是: 凡是标“*”号的章节(包括小节)可作为老师选讲的内容或供给读者阅读的材料. 凡是标“*”的习题表示有相当难度, 可供学生选做. 即为可做也可不做的习题.

非统计专业的学生使用本教材, 第 7 章的内容可以不讲, 可用“线性回归模型”的内容代替. 作者在给中国科学技术大学数学类本科生讲授本课程时采用过这种方式, 效果不错. 有关“线性回归模型”一章的内容可从作者个人主页 <http://staff.ustc.edu.cn/~lwei/books.htm> 下载.

在第二版准备过程中, 科学出版社为本书第二版的出版给予了大力支持, 在此

表示真诚的感谢。

由于本人水平有限，本书一定会有不少缺点和疏漏，恳请同行专家及广大读者批评指正。

韦来生

2015年8月于合肥

第一版前言

作者在 20 世纪 80 年代初给中国科学技术大学数学系 81 级数理统计专业讲授“数理统计”课，当时没有合适的教材，就自编了讲稿，学生记笔记。在给数学系 83 级、84 级讲授“数理统计”课时对讲稿进行了充实和完善，并编印了一本习题集。1988 年，陈希孺院士等编写出版了《数理统计学教程》（以下简称《教程》）。1990 年后，《教程》作为中国科学技术大学概率论与数理统计专业“数理统计”课的教材。《教程》的特点是统计理论严谨，对统计思想和统计问题的背景等阐述清楚明了。作者在教学实践中充分发扬了《教程》的特色，结合过去的讲稿对教学内容作了适当的增补和调整，教学效果良好。本书稿就是在这一基础上完成的。

全书共分 7 章。前 2 章是预备知识，分别介绍数理统计的若干基本概念和抽样分布。特别要强调的是，第 2 章抽样分布是后面几章的基础。后 5 章介绍数理统计的方法和理论，其中第 3 章和第 4 章分别介绍点估计和区间估计；第 5 章和第 6 章介绍参数假设检验和非参数假设检验；最后一章，即第 7 章介绍 Bayes 方法和统计决策理论，这是近半个多世纪迅速发展起来的数理统计的一个重要分支。在第 3 章参数估计和第 5 章参数假设检验问题中，对有关统计推断方法的最优化理论作了较系统的介绍。在每一章的介绍中注重对问题的背景和统计思想、方法的阐述，并附有大量例题和习题。

这本教材的主要内容在中国科学技术大学概率论与数理统计专业讲授过多次，大约可在 72 小时内讲授全书各章的主要内容。适当删除书中标“*”的章、节、段的内容后，仍成系统，可组成 54 学时左右的课程。因此，本书可作为概率论与数理统计专业基础课的教材，也可作为数学系非概率论与数理统计专业本科生的“数理统计”课教材。

本书编写过程中主要参考了陈希孺院士等编写的《数理统计学教程》，同时还参考了华东师范大学、北京大学等兄弟院校的数理统计教材，在此表示衷心的感谢。

中国科学技术大学赵林城教授仔细地审阅了书稿，提出了一些非常宝贵修改意见。作者在修改时充分考虑了他的意见，在此向他表示深深的谢意。

中国科学技术大学统计与金融系 04 级统计班的几位同学在班长冯文宁同学的组织下，帮助完成书稿 1-6 章的中文 Tex 的打字和编译，对他们的辛勤工作表示真诚的感谢。作者的同事张伟平博士帮助完成全书的制图和一些章节的排版工作，并对书稿的修改提出了一些有益的建议，在此表示诚挚的感谢。科学出版社为本书的出版给予了大力支持，在此一并致谢。

由于作者水平有限,本书一定会有不少缺点和疏漏,恳请国内同行及广大读者批评指正。

作 者

2008年2月于中国科学技术大学

常用符号及缩写

\mathcal{X}	样本空间
Θ	参数空间
$N(a, \sigma^2)$	均值为 a , 方差为 σ^2 的正态 (normal) 分布
$\Phi(\cdot)$	标准正态分布函数
$b(1, p)$	成功概率为 p 的两点分布 (也称 Bernoulli 分布)
$b(n, p)$	参数为 n, p 的二项 (binomial) 分布
$M(N, p_1, \dots, p_r)$	参数为 N, p_1, \dots, p_r 的多项 (multinomial) 分布
$P(\lambda)$	参数为 λ 的泊松 (Poisson) 分布
$U(a, b)$	区间 $[a, b]$ 上的均匀 (uniform) 分布
$Be(a, b)$	参数为 a, b 的贝塔 (Beta) 分布
$\Gamma(\gamma, \lambda)$	形状参数为 γ , 刻度参数为 λ 的伽马 (Gamma) 分布
$\Gamma^{-1}(\alpha, \beta)$	参数为 α, β 的逆伽马 (Inverse Gamma) 分布
$Exp(\lambda)$	参数为 λ 的指数 (exponential) 分布
$C(\mu, \lambda)$	位置参数为 μ , 刻度参数为 λ 的 Cauchy 分布
u_α	标准正态分布的上侧 α 分位数
$\chi_n^2, \chi_n^2(\alpha)$	自由度为 n 的卡方 (Chi-square) 分布及其上侧 α 分位数
$t_n, t_n(\alpha)$	自由度为 n 的 t 分布及其上侧 α 分位数
$F_{m,n}, F_{m,n}(\alpha)$	自由度分别为 m, n 的 F 分布及其上侧 α 分位数
R_n	n 维欧几里得空间
\mathbf{X}	由若干个随机变量作为分量构成的随机向量
x	随机向量 \mathbf{X} 的观测值
$E(Y), D(Y)$	随机变量 Y 的均值和方差
r.v.	随机变量
i.i.d.	相互独立相同分布
$f(x, \theta)$	r.v. X 的概率函数, 连续型 r.v. 为密度函数, 离散型 r.v. 为概率分布
$I_A(x), I_A$	示性函数, 表示当 $x \in A$ (或 A 发生) 函数值为 1, 否则为 0
MLE	极大似然估计
UMVUE	一致最小方差无偏估计
UMPT	一致最优 (功效) 检验

目 录

第二版前言

第一版前言

常用符号及缩写

第 1 章 绪论	1
1.1 什么是数理统计学	1
1.2 数理统计的若干基本概念	8
1.3 统计量	17
习题 1	22
第 2 章 抽样分布及若干预备知识	24
2.1 引言	24
2.2 正态总体样本均值和样本方差的分布	24
2.3 次序统计量的分布	28
2.4 χ^2 分布, t 分布和 F 分布	32
2.5 统计量的极限分布	43
*2.6 指数族	45
2.7 充分统计量	50
*2.8 完全统计量	58
习题 2	64
第 3 章 点估计	70
3.1 引言	70
3.2 矩估计	75
3.3 极大似然估计	84
*3.4 一致最小方差无偏估计	96
3.5 Cramer-Rao 不等式	108
*3.6 概率密度函数的核估计	116
习题 3	121
第 4 章 区间估计	127
4.1 区间估计的基本概念	127
4.2 极轴变量法 —— 正态总体参数的置信区间	131

4.3 极轴变量法——非正态总体参数的置信区间.....	141
*4.4 Fisher 的信仰推断法.....	148
*4.5 容忍区间与容忍限.....	154
习题 4.....	162
第 5 章 参数假设检验	167
5.1 假设检验的若干基本概念.....	167
5.2 正态总体参数的假设检验.....	174
5.3 似然比检验.....	193
*5.4 一致最优检验与无偏检验.....	202
5.5 假设检验与区间估计.....	217
习题 5.....	225
第 6 章 非参数假设检验	232
6.1 引言	232
6.2 符号检验及符号秩和检验	233
6.3 Wilcoxon 两样本秩和检验	241
6.4 拟合优度检验	247
6.5 列联表中的独立性和齐一性检验	258
*6.6 其他的非参数检验方法	265
习题 6	272
*第 7 章 Bayes 方法和统计决策理论	276
7.1 引言和若干基本概念	276
7.2 先验分布的确定	281
7.3 Bayes 统计推断	295
7.4 Bayes 统计决策理论	305
*7.5 Minimax 准则	316
*7.6 同变估计及可容许性	319
习题 7	324
参考文献	328
附录	329
附表 1 标准正态分布表	329
附表 2 t 分布表	330
附表 3 χ^2 分布表	331
附表 4 F 分布表	333
附表 5 泊松分布表	340

附表 6 正态分布容许限 $\bar{X} + \lambda s$ 或 $\bar{X} - \lambda s$ 中系数 $\lambda(n, \beta, \gamma)$ 值表	341
附表 7 正态分布容许区间 $\bar{X} \pm \lambda s$ 中系数 $\lambda(n, \beta, \gamma)$ 值表	342
附表 8 非参数容许限 —— 相应于总体比例 $1 - \beta$ 和置信水平 $1 - \gamma$ 的 样本容量 n	343
附表 9 非参数容许区间 —— 相应于总体比例 $1 - \beta$ 和置信水平 $1 - \gamma$ 的样本容量 n	343
附表 10 符号检验临界值表	343
附表 11 符号秩和检验临界值表	344
附表 12 秩和检验临界值表	345
附表 13 柯尔莫哥洛夫检验临界值 $D_{n,\alpha}$	346
附表 14 柯尔莫哥洛夫检验统计量 D_n 的极限分布	347
附表 15 W 检验 统计量 W 的系数 $a_i(n)$ 的值	348
附表 16 W 检验 统计量 W 的 α 分位数 W_α	350
附表 17 D 检验 统计量 Y 的 α 分位数 Y_α	351
索引	352

第1章 絮 论

1.1 什么是数理统计学

1.1.1 数理统计学的任务和性质

自然界的现像大致可以分为两大类，一类称为确定性现像，另一类称为非确定性现像，亦称为随机现像。确定性现像的例子，如物理学中的自由落体运动，可以用数学方程 $s = gt^2/2$ 刻画其运动规律。这样的例子还有许多，如物理学中的许多定律、化学中的反应规律和其他学科中的一些现像，它们皆可以用数学中的方程式，如微分方程等来精确描述。随机现像的例子，如在农业试验中，在面积相等且相邻的两块土地上种植同一种小麦，生产条件相同，但在收获时小麦产量不完全一样。又如在工业生产中，进行某化工产品得率的试验，使温度、压力和配方等主要因素控制在相同水平下，获得的两批化工产品得率不能保证完全相同。再如战士打靶试验，同一战士在相同条件下每次打靶命中的环数不尽相同。这些都是随机现像，它们在自然界是大量存在的。数理统计和概率论一样，是研究随机现像的统计规律性的数学学科。为了说明它的研究方法与概率论和其他数学学科有什么不同，首先来介绍什么是数理统计学。

统计学的任务是研究怎样有效地收集、整理和分析带有随机性影响的数据，从而对所考虑的问题作出一定结论的方法和理论。它是一门实用性很强的学科，在人类活动的各个领域都有着广泛的应用。统计学的思想和方法是人类文明的一个组成部分。研究统计学方法中理论基础问题的那一部分构成“数理统计学”的内容。一般地，可以认为

数理统计是数学的一个分支，它是研究如何有效地收集和使用带有随机性影响的数据的一门学科。

下面通过例子对此陈述加以说明。

1. 有效地收集数据

收集数据的方法有：全面观察（或普查）、抽样调查和安排试验等方式。

例 1.1.1 人口普查和抽样调查。 我国在 2000 年进行了第 5 次人口普查。如果普查的数据是准确无误的，则无随机性可言，不需用数理统计方法。由于人口普查调查的项目很多，我国有 13 亿人口，普查工作量极大，且缺乏训练有素的工作人员。

员, 所以虽是全面调查, 但数据并不很可靠. 例如, 农村超计划生育瞒报、漏报人口的情况时有发生. 针对普查数据不可靠, 国家统计局在人口普查的同时还派出专业人员对全国人口进行抽样调查, 根据抽样调查的结果, 对人口普查的数字进行适当的修正. 抽样调查在普查不可靠时是一种补充办法. 如何安排抽样调查, 是有效收集数据的一个重要问题, 这构成数理统计学的一个重要分支——抽样调查方法.

例 1.1.2 考察某地区 10000 户农户的经济状况, 从中挑选 100 户作抽样调查. 若该地区分成平原和山区两部分, 平原较富, 占该地区农户的 70%, 而占 30% 的山区农户较穷. 抽样方案规定在抽取的 100 户中, 从平原地区抽 70 户, 山区抽 30 户, 在各自范围内用随机化方法抽取.

在本例中, 有效收集数据是通过合理地设计抽样方案来实现的. 在通过试验收集数据的情形中, 如何做到有效收集数据, 请看下例.

例 1.1.3 某化工产品的得率与温度、压力和原料配方有关. 为提高得率, 通过试验寻找最佳生产条件. 试验因素和水平如下:

因素	水平			
	1	2	3	4
温度	800°C	1000°C	1200°C	1400°C
压力	10	20	30	40
配方	A	B	C	D

3 个因素, 每个因素 4 个水平共要做 $4^3 = 64$ 次试验. 做这么多试验, 人力、物力、财力都不可能. 因此, 如何通过尽可能少的试验获得尽可能多的信息? 例如, 采用正交表安排试验就是一种有效的方法. 如何科学安排试验方案和分析试验结果, 这构成数理统计的另一分支——试验的设计和分析. 在本例中, 有效收集数据是通过科学安排试验的方法来实现的.

在有效收集数据中, 一个重要问题是数据必须具有随机性. 在例 1.1.2 中, 随机性体现在抽样的 100 户农户是从 10000 户农户中按一定的方式“随机抽取”的, 它具有一定的代表性(山区和平原地区农户按比例抽取). 假如只在该地区富裕的那部分农户中挑选, 得到的数据就不具有代表性, 也谈不上有效. 而在例 1.1.3 中, 数据的随机性是由试验误差来体现的. 化工产品的得率除了受温度、压力和配方影响外还受一些无法控制, 甚至仍未被人们认识的因素影响, 如每次试验中受试验材料产地的影响、所使用仪器设备精度的影响和操作者水平的影响等. 这些因素无法或不便加以完全控制, 从而对试验结果产生随机性的影响, 这就带来不确定性.

2. 有效地使用数据

获取数据后, 需要用有效的方法去集中和提取数据中的有关信息, 以对所研究

的问题作出一定的结论, 这在统计上称为“推断”.

为了有效地使用数据进行统计推断, 需要对数据建立一个统计模型 (如何建模见例 1.2.6 和例 1.2.7), 提出统计推断的方法, 并给定某些准则去评判不同统计推断方法的优劣. 例如, 为估计一个物体的重量 a , 把它在天平上称 5 次获得数据 x_1, x_2, \dots, x_5 , 它们都受到随机性因素的影响 (天平的精度反映了影响的大小). 估计 a 的大小采用下列 3 种不同方法: ①用 5 个数的算术平均值 $\bar{x} = (x_1 + \dots + x_5)/5$ 去估计 a ; ②将 x_1, x_2, \dots, x_5 按大小排列为 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(5)}$, 取中间一个值 $x_{(3)}$ 去估计 a ; ③用 $W = (x_{(1)} + x_{(5)})/2$ 去估计 a . 可能认为 \bar{x} 优于 $x_{(3)}$, 而 $x_{(3)}$ 优于 W . 这是不是对的? 为什么? 在什么条件下才对? 事实上, 对这些问题的研究正是数理统计学的任务. 以后可以看到在一定的统计模型和优良性准则下, 上述 3 个估计方法中的任何一个都可能是最优的.

下面举例说明, 针对不同问题采用适当的统计方法也是有效使用数据的一个重要方面.

例 1.1.4 某农村有 100 户农户, 要调查此村农户是否脱贫. 脱贫的标准是每户年均收入超过 1 万元. 经调查此村 90 户农户年收入 5000 元, 10 户农户年收入 10 万元, 问此村农户是否脱贫?

(1) 用算术平均值计算该村农户年均收入如下:

$$\bar{x} = \frac{90 \times 0.5 + 10 \times 10}{100} = 1.45(\text{万元}).$$

按此方法得出结论: 该村农民已脱贫. 但 90% 的农户年均收入只有 5000 元, 事实上并未脱贫.

(2) 用样本中位数计算该村农户年均收入, 即将 100 户的年收入分别记为 x_1, x_2, \dots, x_{100} , 将其按大小排列为 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(100)}$. 样本中位数定义为排在最中间两户的平均值, 即

$$\frac{x_{(50)} + x_{(51)}}{2} = 0.5(\text{万元}).$$

按此方法得出结论: 该村农民尚未脱贫. 这与实际情况相符.

由此可见, 不同的统计方法得出的结论不同. 有效地使用数据, 需要针对不同问题选择合适的统计方法.

3. 数理统计学与各种专门学科的关系

数理统计方法有很广泛的实用性, 它与很多专门学科都有关. 但是应当了解: 数理统计方法所处理的只是在各种专门学科中带普遍性 (共性) 且受随机性影响的数据收集、整理和推断问题, 而不去涉及各种专门学科中的具体问题. 这种带共性

的问题既然从专门领域中提炼出来，就可以用数学的方法去研究，这就是数理统计学的研究任务，因此数理统计是一个数学的分支。

拿例 1.1.3 来说，实地进行这个化工产品得率试验，当然要涉及一系列专门的化工知识。在安排试验时，没有这些专门知识是不行的。但是，有些试验安排上的数学问题却与化工这个专门领域无关。例如，数理统计学告诉我们，各因素取同样水平数参与试验，以后的数据分析比较方便。如果在 $4^3 = 64$ 种搭配中只做一部分，则按某种方式挑选这一部分（如按正交表安排试验），以后的数据分析就容易进行。当下一次碰到其他工业试验时，这些考虑仍有效。又如，称重试验中用多次称重结果的平均数去估计物体的重量，是一个常用的统计方法，不管这个量是物理的、化学的或是生物的，这一点都适用。

由统计方法的这个性质就引申出一个重要特点：统计方法只是从事物外在数量上的表现去推断该事物可能的规律性。统计方法本身不能说明何以会有这个规律性，这是各个专门学科的任务。例如，用统计方法分析一些资料发现，吸烟与某些呼吸系统的疾病有关。这纯粹是从吸烟者和不吸烟者的发病率的对比上得出的结论，它不能解释吸烟何以会增加患这类疾病的危险性，这是医学这个专门学科的任务。

但是，应当认识到，这并不意味着一个数理统计学者可以不过问其他专门领域的知识。相反，如果要将统计方法用于实际问题，必须对所涉及问题的专门知识有一定的了解，这不仅可以帮助选定适当的统计模型和统计方法，而且在正确解释所得结论时，专门知识是必不可少的。例如，数理统计学在遗传基因分析中很有用，但一个对遗传基因学一无所知的统计学家，很难在这个领域有所作为。

4. 数理统计方法的归纳性质

数理统计是数学的一个分支，但是它与其他数学学科的推理方法是不一样的。统计方法的本质是归纳式的，而其他数学学科则是演绎式的。统计方法的归纳性质，源于它在作结论时，是根据所观察到的大量的“个别”情况“归纳”起来所得，而不是从一些假设、命题或已知事实出发，按一定的逻辑推理得出来的（后者称为演绎推理）。例如，统计学家通过大量的观察资料发现，吸烟与某种呼吸系统的疾病有关。得出这一结论的根据是：从观察到的大量例子，看到吸烟者中患此种疾病的比列远高于不吸烟者。不可能用逻辑推理的方法证明这一点。试拿统计学与几何学进行比较就可以清楚地看出二者方法的差别所在。在几何学中要证明“等腰三角形两底角相等”，只需从等腰这个前提出发，运用几何公理，一步步地推出这个结论（这一方法属于演绎推理）。而一个习惯于统计方法的人，就可能想出这样的方法：作很多大小和形状不一的等腰三角形，实际测量它的底角查看是否相等，根据所得数据，看看能否作出底角相等的结论，这一方法属于归纳推理。

众所周知，归纳推理是要冒风险的。事实上，归纳推理的不确定性的出现是一

种逻辑的必然。人们不可能作出十分肯定的结论，因为归纳推理所依据的数据具有随机性。然而，不确定性的推理是可行的，所以推理的不确定性程度是可以计算的。统计学的作用之一就是提供归纳推理和计算不确定性程度的方法。不确定性是用概率计算的。以后会见到在参数的区间估计问题中，不但给出区间估计的表达式，而且给出这一区间包含未知参数的可靠程度的大小。

总之，统计推断属于归纳推理方法，归纳推理作出的推断不是 100% 可靠，但它的可靠程度（即结论的正确程度）是可以通过概率来度量的。

1.1.2 数理统计学的应用

人类在科学、生产、管理和各方面活动，大都离不开数据资料的收集、整理和分析的工作。因此统计学的应用领域也极其广泛。

(1) 国家行政机关和职能机构，如国家统计局，经常需要收集相关的数据和资料并加以整理、分析后提供给有关部门作出相应的决策。这里面的统计工作，固然有大量的描述性统计的成分，但统计推断的方法也很有用并且十分必要。例如，在判断某一时期经济运行是否过热，是否需要采取宏观调控措施等重大决策时，对当时经济运行中数据资料进行定量分析是必不可少的。这就离不开统计推断方法。

用数理统计方法进行社会调查，这种工作属于国家职能部门的工作范围。“抽样调查”是常用的方法。统计学的方法在决定调查规模和制定有效的抽样方案时是很有用的，统计推断方法在对调查得来的资料进行正确分析时也有指导意义。例如，经过精心设计和组织的社会抽样调查，其效果有时可达到甚至超过全面调查的水平。在人口学中，确定一个合适的人口发展动态模型需要掌握大量的观察资料，而且要使用包括统计方法在内的一些科学方法。再如，在人寿保险中，对寿命数据的分析、建立精算模型也要用到一些统计方法。

(2) 在工农业生产中，常常要利用试验设计和方差分析的方法寻找最佳生产条件。例如，为提高农业中的单位面积产量，有一些因素对这个指标有影响：种子的品种、施肥量和浇水量等；工业生产中影响某项产品质量指标的因素有原材料产地、配方、温度和压力等因素。为了找到一组较好的生产条件就要进行试验，如何科学地安排试验和分析试验结果，就需要用到统计方法。试验设计的基本思想和方差分析方法就是 R.A. Fisher 等于 1923-1926 年，在进行田间试验中发展起来的，这一方法后来广泛应用于工业生产中。

数理统计方法应用于工业生产的另一个重要方面是产品质量控制、抽样调查和工业产品寿命的可靠性问题。现代工业生产有批量大和高可靠度的特点，需要在连续生产过程中进行工序控制。成批的产品在交付使用前要进行验收，这种验收一般不能进行全面检验，而只能是抽样验收，需要根据统计学的原理制定合适的抽样方案。大型设备或复杂产品（如导弹）包含成千上万个元件。由于元件的数目很大，

元件的寿命服从一定的概率分布, 整个设备(或产品)的寿命与其结构和元件的寿命分布有关, 因此为了估计设备(或产品)的可靠性, 发展了一系列的统计方法。统计质量管理就是由上述提到的这些方法构成的。

(3) 数理统计方法在经济和金融领域也有广泛的应用, 在经济学中定量分析的应用比其他社会科学部门更早更深入。现在有一门称为“计量经济学”的学科, 其内容主要就是将统计方法(及其他数学方法)用于分析经济领域中数量方面的问题。例如, 早在 20 世纪二三十年代, 时间序列的统计分析方法就用于市场预测, 目前金融等领域也广泛地使用时间序列方法。

(4) 统计方法在生物、医学和遗传学中有广泛的应用。一种药品的疗效如何, 要通过精心安排的试验并使用正确的统计分析方法, 才能比较可靠地作出结论。分析某种疾病的发生是否与特定因素有关(一个典型的例子是吸烟与患肺癌的关系), 这些问题常常是从观察和分析大量资料的基础上得到启示, 再提高到理论上的研究。这方面的应用还有流行病数据的统计分析、遗传基因数据的统计分析等。

(5) 数理统计方法在气象预报、水文、地震、地质等领域有广泛应用。在这些领域中, 人们对事物规律性的认识不充分, 使用统计方法有助于获得一些对潜在规律性的认识, 用以指导人们的行动。

(6) 数理统计方法在科学的研究中也具有重要作用。自然科学研究的根本任务是揭示自然界的规律性, 科学实验是重要手段, 而随机因素对实验结果的影响无所不在。一个好的统计方法有助于提取实验和观察数据中根本性的信息, 因而有助于提出较正确的理论或假说。有了一定的理论和假说后, 统计方法可以指导研究工作者如何进一步安排试验或观察, 以使所得数据更有助于判定定理或假说是否正确。数理统计学也提供了理论上有效的方法去评估观察或试验数据与理论的符合程度如何。一个著名的例子是遗传学中的 Mendal 定律。这个根据观察资料提出的定律, 经历了严格的统计检验。数量遗传学中的基本定律: Harday-Weinberg 平衡定律也具有这种性质。由此可见, 科学研究需要数理统计方法。另一方面, 应用上的需要又是统计方法发展的动力。例如, 现代统计学的奠基人、英国著名学者 R.A. Fisher 和 K. Pearson 在 20 世纪初期从事统计学的研究, 就是出于生物学、遗传学和农业科学方面的需求。

1.1.3 统计学发展简史

数理统计学是一门较年轻的学科, 它主要的发展是从 20 世纪初开始, 大概可分为两个阶段。前一阶段大致上到第二次世界大战结束时为止。在这一早期发展阶段中, 起主导作用的是以 R.A. Fisher 和 K. Pearson 为首的英国学派, 特别是 Fisher, 他在本学科的发展中起了独特的作用。其他一些著名的学者, 如 W.S. Gos-

set (student), J. Neyman, E.S. Pearson (K. Pearson 的儿子), A. Wald 以及我国的许宝𫘧教授等都作出了根本性的贡献. 他们的工作奠定了许多数理统计学分支的基础, 提出了一系列具有重要应用价值的统计方法和一系列基本概念和重要理论问题. 有一种意见认为瑞典统计学家 H. Cramer 在 1946 年发表的著作 *Mathematical Methods of Statistics* (见文献 [3]) 标志了这门学科达到成熟的地步.

收集和记录种种数据的活动, 在人类历史上来源已久. 翻开我国二十四史, 可以看到上面有很多关于钱粮、人口、地震及洪水等自然灾害的记录. 在西方国家, Statistics (统计学) 一词源出于 State (国家), 意指国家收集的国情材料. 19 世纪中叶以后, 包括政治统计、人口统计、经济统计、犯罪统计、社会统计等多方面内容的“社会统计学”一词在西方开始出现, 与此相应的社会调查也有了较大发展. 人们试图通过社会调查, 收集、整理和分析数据, 以揭示社会现象和问题, 并提出解决具体问题的方法. 这种情况延续了许多年, 研究方法属于描述统计学的范畴. 这是因为, 当时还没有一定的数学工具特别是概率论的发展, 无法建立现代意义上的数理统计学. 也因为这方面的需求还没达到那么迫切, 足以构成一股强大的推动力. 到 19 世纪末和 20 世纪初情况才起了较大的变化. 有人认为 20 世纪初 K. Pearson 关于 χ^2 统计量极限分布的论文可以作为数理统计诞生的一个标志; 也有人认为, 直到 1922 年 Fisher 关于统计学的数学基础那篇著名论文的发表, 数理统计才正式诞生.

综上所述, 可以得到如下粗略的结论: 收集和整理乃至使用试验和观察数据的工作由来已久, 这类活动对于数理统计学的产生, 可算是一个源头. 19 世纪, 特别是 19 世纪后半期发展速度加快, 且有了质的变化. 19 世纪末到 20 世纪初这一阶段, 出现了一系列的重要工作. 无论如何, 至迟到 20 世纪 20 年代, 这门科学已稳稳地站住了脚跟. 20 世纪前 40 年有了迅速而全面的发展, 到 20 世纪 40 年代时, 数理统计已形成为一个成熟的数学分支.

从战后到现在可以说是第二阶段. 在这个时期中, 许多战前开始形成的数理统计分支, 在战后得以向纵深发展, 理论上的深度也比以前大大加强了. 同时还出现了根本性的发展, 如 Wald 的统计判决理论和 Bayes 学派的兴起. 在数理统计的应用方面, 其发展也给人深刻印象. 这不仅是战后工农业生产和科学技术迅速发展所提出的要求, 也是由于电子计算机这一有力工具的出现和飞速发展推动了数理统计学的进步. 战前由于计算工具跟不上, 许多需要大量计算的统计方法很难得以使用. 战后有了高速计算机使这一问题变得很容易, 这就大大推广了统计方法的应用. 目前, 统计方法仍在蓬勃发展中. 在一些统计学发达的国家中, 特别在美国, 这方面的人才数以十万计, 并在大多数大学中建立了统计系. 近 30 年来, 数理统计学在我国的发展也是令人瞩目的, 尤其是 2011 年统计学从数学和经济学中独立出来成为一级学科, 极大地推动了统计学的发展.

1.2 数理统计的若干基本概念

1.2.1 总体和样本

总体又称为母体. 通过下面的例子说明总体、个体和样本的概念.

例 1.2.1 假定一批产品有 10000 件, 其中有正品也有废品. 为估计废品率, 往往从中抽取一部分, 如 100 件进行检查. 此时这批 10000 件产品称为总体 (population), 其中的每件产品称为个体 (individual), 而从中抽取的 100 件产品称为样本 (sample). 样本中个体的数目称为样本大小 (sample size), 也称为样本容量. 而抽取样本的行为称为抽样 (sampling). 从本例可对总体和样本作如下直观的定义:

总体是由与所研究的问题有关的所有个体组成, 而样本是从总体中抽取的一部分个体.

若总体中个体的数目为有限个, 则称为有限总体 (finite population), 否则称为无限总体 (infinite population).

在统计研究中, 人们所关心的不是总体内个体的本身, 而是关心个体上的一项 (或几项) 数量指标, 如日光灯的寿命、零件的尺寸. 在例 1.2.1 中若产品为正品用 0 表示, 若产品为废品用 1 表示, 关心个体取值是 0 还是 1. 因此又可获得总体的如下定义:

总体可以看成由所有个体上的某种数量指标构成的集合, 因此它是数的集合.

由于每个个体的出现是随机的, 所以相应的个体上数量指标的出现也带有随机性. 从而可以把此种数量指标看成随机变量 (random variable, r.v.), 随机变量的分布就是该数量指标在总体中的分布. 以例 1.2.1 来说明, 假定 10000 只产品中废品数为 100 件, 其余的为正品, 废品率为 0.01. 定义随机变量 X 如下:

$$X = \begin{cases} 1, & \text{废品,} \\ 0, & \text{正品,} \end{cases}$$

其概率分布为 0-1 分布, 且 $P(X = 1) = 0.01$. 因此特定个体上的数量指标是 r.v. X 的观察值 (observation). 这样一来, 总体可以用一个随机变量及其分布来描述, 获得如下定义.

定义 1.2.1 一个统计问题所研究的对象的全体称为总体. 在数理统计学中总体可以用一个随机变量及其概率分布来描述.

由于总体的特征由总体分布来刻画, 所以统计学上常把总体和总体分布视为同义语. 由于这个缘故, 常用随机变量的符号或分布的符号来表示总体. 比如研究某批日光灯的寿命时, 人们关心的数量指标是寿命 X , 那么此总体就可以用 r.v. X 来表示, 或用其分布函数 F 来表示. 若 F 有密度, 记为 f , 则此总体也可用密度函数

f 来表示. 有时也根据总体分布的类型来称呼总体的名称, 如正态总体、二项分布总体、0-1 分布总体. 若总体分布函数记为 F , 当有一个从该总体中抽取的相互独立同分布 (i.i.d.) 的大小为 n 的样本 X_1, \dots, X_n , 则常记为

$$X_1, \dots, X_n \text{ i.i.d.} \sim F. \quad (1.2.1)$$

若 F 有密度 f , 可记为

$$X_1, \dots, X_n \text{ i.i.d.} \sim f. \quad (1.2.2)$$

若所考虑的总体用 r.v. X 表示, 其分布函数为 F , 则样本 X_1, \dots, X_n 可视为 r.v. X 的观察值, 亦可记为

$$X_1, \dots, X_n \text{ i.i.d.} \sim X. \quad (1.2.3)$$

式 (1.2.1)- 式 (1.2.3) 所代表的含义是完全相同的.

当个体上的数量指标不止一项时, 用随机向量来表示总体. 例如, 研究某地区小学生的发育状况时, 人们关心的是其身高 X 和体重 Y 这两个数量指标, 此时总体就可以用二维随机向量 (X, Y) 或其联合分布 $F(x, y)$ 表示. 当 F 有密度 f 时, 总体也可用联合密度 $f(x, y)$ 表示.

注 1.2.1 在一些场合下, 总体具有一定的抽象性, 请看例 1.2.2.

例 1.2.2 用秤去称一个物体的重量, 为了得到较准确的结果, 将物体称 5 次, 取 5 次称重的算术平均值作为物体的重量. 如 5 次称重的结果 (单位: g) 为 125.5, 124, 124.3, 126, 125.2. 此问题中总体和样本是什么?

显然, 如上每一组结果都是一个样本. 总体可以设想为用秤去称此物体, 无限地称下去, 把总体理解为“一切可能出现的称量结果的集合”. 这是一个无限总体.

1.2.2 样本空间和样本的两重性

1. 样本空间

样本是由总体中抽取的一部分个体组成的. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体中抽取的样本, 其样本空间 定义如下.

定义 1.2.2 样本 $\mathbf{X} = (X_1, \dots, X_n)$ 可能取值的全体, 构成样本空间 (sample space), 记为 \mathcal{X} .

例如, 在例 1.2.2 中, 样本空间为 $\mathcal{X} = \{(x_1, \dots, x_5) : 0 < x_i < \infty, i = 1, 2, \dots, 5\}$, 也可以写成 $\mathcal{X} = \{(x_1, \dots, x_5) : -\infty < x_i < \infty, i = 1, 2, \dots, 5\}$. 虽然物重不可以取负数, 但这无关紧要, 因为在考虑样本分布时, 可令样本取负值的概率为 0. 再看下例.

例 1.2.3 打靶试验, 每次打三发, 考察中靶的环数. 如样本 $\mathbf{X} = (5, 1, 9)$ 表示三次打靶分别中 5 环, 1 环和 9 环. 此时样本空间为

$$\mathcal{X} = \{(x_1, x_2, x_3) : x_i = 0, 1, 2, \dots, 10, i = 1, 2, 3\}.$$

这个样本空间中样本点数是有限的, 例 1.2.2 的样本空间中的样本点数是无限的.

2. 样本的两重性

当从总体中作具体抽样时, 每次抽样的结果都是些具体的数, 如例 1.2.3 的打靶问题中, 样本 $\mathbf{X} = (X_1, X_2, X_3)$, 其中 $0 \leq X_i \leq 10$ ($i = 1, 2, 3$) 为整数, 它是数字向量. 但若是在相同条件下, 再打三发, 由于种种不可控制的随机因素的影响, 中靶的环数不大可能和上一次完全一样, 这就是样本的随机性. 如果无穷地打下去, 每次打三发, 出现的结果可视为随机向量 (X_1, X_2, X_3) 的具体观察值.

样本的两重性是说, 样本既可看成具体的数, 又可以看成随机变量 (或随机向量). 在实施抽样后, 它是具体的数; 在实施抽样前, 它被看成随机变量 (或随机向量). 因为在实施具体抽样之前无法预料抽样的结果, 只能预料它可能取值的范围, 因此可以把它看成随机变量 (或随机向量). 为区别起见, 今后用大写的英文字母表示随机变量, 用黑体大写英文字母表示随机向量或一组样本, 用小写字母表示它们的具体观察值.

理论工作者更重视样本是随机变量 (或随机向量) 这一点, 而应用工作者虽重视将样本看成具体的数字, 但仍不可忽视样本是随机变量 (或随机向量) 这一背景. 否则, 样本就是一堆杂乱无章毫无规律可言的数字, 无法进行任何统计处理. 样本既然是随机变量 (或随机向量), 就有分布而言, 这样才存在统计推断问题.

3. 简单随机样本

抽样是指从总体中按一定方式抽取样本的行为. 抽样的目的是通过取得的样本对总体分布中某些未知的量作出推断, 为使抽取的样本能很好地反映总体的信息, 须考虑抽样方法. 最常用的一种抽样方法称为“简单随机抽样”, 它满足下列要求.

(1) 代表性. 总体中的每一个个体都有同等机会被抽入样本, 这意味着样本中每个个体与所考察的总体具有相同分布. 因此, 任一样本中的个体都具有代表性.

(2) 独立性. 样本中每一个个体取什么值并不影响其他个体取什么值. 这意味着, 样本中各个个体 X_1, X_2, \dots, X_n 是相互独立的随机变量.

由简单随机抽样获得的样本 (X_1, \dots, X_n) 称为简单随机样本. 其定义如下.

定义 1.2.3 设有一总体 F , X_1, \dots, X_n 为从 F 中抽取的容量为 n 的样本, 若

(1) X_1, \dots, X_n 相互独立,

(2) X_1, \dots, X_n 相同分布, 即同有分布 F ,

则称 X_1, \dots, X_n 为简单随机样本, 有时简称为简单样本或随机样本.

设总体为 F , X_1, \dots, X_n 为从此总体中抽取的简单随机样本, 则 X_1, \dots, X_n 的联合分布函数 $F(x_1, \dots, x_n)$ 可表示为

$$F(x_1) \cdot F(x_2) \cdots F(x_n) = \prod_{i=1}^n F(x_i).$$

若 F 有密度 f , 则其联合密度函数 $f(x_1, \dots, x_n)$ 可表示为

$$f(x_1) \cdot f(x_2) \cdots f(x_n) = \prod_{i=1}^n f(x_i).$$

若样本是多维的, 如从一大群人中抽取 n 个人, 测出每人的身高和体重. 用随机向量 (X, Y) 或用其分布函数 $F(x, y)$ 记总体, $(X_1, Y_1), \dots, (X_n, Y_n)$ 就是从这一总体中抽取的一组简单随机样本, 其联合分布函数为

$$F(x_1, y_1) \cdot F(x_2, y_2) \cdots F(x_n, y_n) = \prod_{i=1}^n F(x_i, y_i).$$

若 $F(x, y)$ 有密度 $f(x, y)$, 则其联合密度函数为

$$f(x_1, y_1) \cdot f(x_2, y_2) \cdots f(x_n, y_n) = \prod_{i=1}^n f(x_i, y_i).$$

显然, 有放回抽样获得的样本是简单随机样本. 当总体中包含的个体数较大或所抽样本在总体中所占比例较小时, 可以把无放回抽样获得的样本当成简单随机样本.

1.2.3 样本分布

样本既然是随机变量, 就有一定的概率分布, 这个概率分布就称为样本分布. 样本分布是样本所受随机性影响的最完整的描述.

要决定样本分布, 就要根据观察值的具体指标的性质 (这往往涉及有关的专业知识), 以及对抽样方式和对试验进行的方式的了解, 此外常常还必须加进一些人为的假定. 下面看一些例子.

例 1.2.4 一大批产品共有 N 件, 其中废品 M 件, N 已知, 而 M 未知. 现在从中抽出 n 个检验其中废品的件数, 用以估计 M 或废品率 $p = M/N$. 抽样方式为: 不放回抽样, 一次抽一个, 依次抽取, 直到抽完 n 个为止. 求样本分布.

先将问题数量化. 设 X_i 表示第 i 次抽出的样本, 令

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 次抽出的为废品,} \\ 0, & \text{第 } i \text{ 次抽出的为合格品.} \end{cases}$$

样本 X_1, \dots, X_n 中的每一个都只能取 0, 1 两个值之一. 给定一组样本 x_1, \dots, x_n , 每个 x_i 为 0 或 1. 所求的样本分布为 $P(X_1 = x_1, \dots, X_n = x_n)$. 若记事件 $A_i = \{X_i = x_i\}$, 利用概率乘法公式

$$P(A_1 \cdots A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1 A_2 \cdots A_{n-1}),$$

不难求出样本分布. 为便于讨论, 先看 $n = 3$. 设 $x_1 = 1, x_2 = 0, x_3 = 1$, 则

$$\begin{aligned} & P(X_1 = 1, X_2 = 0, X_3 = 1) \\ &= P(X_1 = 1)P(X_2 = 0|X_1 = 1)P(X_3 = 1|X_1 = 1, X_2 = 0) \\ &= \frac{M}{N} \cdot \frac{N-M}{N-1} \cdot \frac{M-1}{N-2} = \frac{M}{N} \cdot \frac{M-1}{N-1} \cdot \frac{N-M}{N-2}. \end{aligned}$$

对一般情形, 记 $\sum_{i=1}^n x_i = a$, 利用概率乘法公式易求

$$\begin{aligned} & P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= \begin{cases} \frac{M}{N} \cdot \frac{M-1}{N-1} \cdots \frac{M-a+1}{N-a+1} \cdot \frac{N-M}{N-a} \cdots \frac{N-M-n+a+1}{N-n+1}, \\ \quad x_1, \dots, x_n \text{ 为 } 0 \text{ 或 } 1 \text{ 且 } \sum_{i=1}^n x_i = a, \\ 0, \quad \text{其余情形.} \end{cases} \end{aligned} \tag{1.2.4}$$

由上述计算可见样本 X_1, \dots, X_n 不是相互独立的, 样本分布是利用乘法公式, 通过条件概率计算出来的.

例 1.2.5 仍以例 1.2.4 为例, 抽样方式改为有放回抽样, 即每次抽样后记下结果, 然后将其放回去, 再抽第二个, 直到抽完 n 个为止, 求样本分布.

在有放回抽样情形, 每次抽样时, N 个产品中的每一个皆以 $1/N$ 的概率被抽出, 此时 $P(X_i = 1) = M/N$, $P(X_i = 0) = (N-M)/N$, 故有

$$\begin{aligned} & P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= \begin{cases} \left(\frac{M}{N}\right)^a \left(\frac{N-M}{N}\right)^{n-a}, & x_1, \dots, x_n \text{ 为 } 0 \text{ 或 } 1 \text{ 且 } \sum_{i=1}^n x_i = a, \\ 0, & \text{其余情形.} \end{cases} \end{aligned} \tag{1.2.5}$$

可见此例比例 1.2.4 要简单, 因为本例中样本 X_1, \dots, X_n 是独立同分布的, 而例 1.2.4 中 X_1, \dots, X_n 不独立. 当 n/N 很小时, 式 (1.2.4) 和式 (1.2.5) 差别很小. 因而当 n/N 很小时, 可把例 1.2.4 中的无放回抽样当成有放回抽样来处理.

例 1.2.6 为估计一物件的重量 a , 用一架天平将它重复称 n 次, 结果记为 X_1, \dots, X_n , 求样本 X_1, \dots, X_n 的联合分布.

要定出 X_1, \dots, X_n 的分布, 就没有前面两个例子那种简单的算法, 需作一些假定: ① 假定各次称重是独立进行的, 即某次称重误差的大小不受其他次称重结果的影响. 这样 X_1, \dots, X_n 就可以认为是相互独立的随机变量. ② 假定各次称重是在“相同条件”下进行的, 可理解为每次用同一天平, 每次称重由同一人操作, 且周围环境(如温度、湿度等)都相同. 在这个假定下, 可认为 X_1, \dots, X_n 是同分布的. 在上述两个假定下, X_1, \dots, X_n 是 n 个相互独立、相同分布的随机变量, 即为简单随机样本.

为确定 X_1, \dots, X_n 的联合分布, 在以上假定之下求出 X_1 的分布即可. 在此考虑称重误差的特性: 这种误差一般由大量的、彼此独立起作用的随机因素叠加而成, 而每一个因素起的作用都很小. 由概率论中的中心极限定理可知这种误差近似服从正态分布. 再假定天平没有系统误差, 则可进一步假定此误差服从均值为 0 的正态分布. 从而 X_1 (它可视为物重 a 加上称重误差之和) 的概率分布为 $N(a, \sigma^2)$. 因此简单随机样本 X_1, \dots, X_n 的分布为

$$f(x_1, \dots, x_n) = (\sqrt{2\pi}\sigma)^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2\right\}. \quad (1.2.6)$$

本例中求样本分布, 引入两种假定: ① 导出样本 X_1, \dots, X_n i.i.d. 的假定, 它与前面例子中“每一个个体有同等机会被抽出”的假定相当; ② 正态假定, 这一点依据问题的性质、概率论的极限理论和以往经验. 后一点是与前例的不同之处.

例 1.2.7 某工厂生产一种电子元件, 如晶体管. 由于大批量生产受随机因素的干扰, 生产出的晶体管的寿命不同. 从中抽取 n 个做寿命试验, 用以估计其平均寿命, 求样本的分布.

设用一定方式抽取了 n 个元件, 将其在正常条件下使用, 直到 n 个元件全部失效为止, 测得 n 个元件寿命为 X_1, \dots, X_n , 这就是样本.

要确定 X_1, \dots, X_n 的联合分布, 要作一些假定: ① 假定工厂的生产量大, 生产条件在所考察的那段时间是稳定的. 这 n 个产品的抽取, 尽量不在同一天、同一个班次或同一个人生产的产品中抽取, 而在生产时间、空间上有足够“跨度”的产品中抽取. 这一假定表明 X_1, \dots, X_n 可以认为是相互独立相同分布的随机变量; ② 为确定 X_1 的分布, 则需作进一步的假定. 例如, 假定寿命具有无后效性, 即在元件已使用了长为 t 的一段时间尚未失效的条件下, 元件至少还能使用一段时间 s 的概率与 t 无关; 又设当元件在时刻 t 尚未失效, 它在 $[t, t + \Delta t]$ 的时间段内失效的概率有 $\lambda\Delta t + o(\Delta t)$ 的形式 (其中 $o(\Delta t)$ 表示 $\lim_{\Delta t \rightarrow 0} o(\Delta t)/\Delta t = 0$), $\lambda > 0$ 为一常数, 则在概率论中已证明了: 在上述两个假定下,

$$P(X_1 < x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

即有指数分布. 也可写成概率密度的形式: X_1 有概率密度

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (1.2.7)$$

本例中 X_1 的分布是严格按概率论的方法推导出来的, 但推导所依据的假定②在实际问题中只能是近似成立. 因此样本 X_1, \dots, X_n 的联合密度为

$$f(x_1, \dots, x_n; \lambda) = \begin{cases} \lambda^n \exp\{-n\lambda\bar{x}\}, & x_1, \dots, x_n > 0, \\ 0, & \text{其他.} \end{cases}$$

此处 $\bar{x} = \sum_{i=1}^n x_i/n$.

例 1.2.8 若样本是多维的, 以上讨论也适用. 例如, 在一群人中抽取 n 个, 测其身高和体重得 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, 这就是大小为 n 的二维样本. 在一定条件下, 假定它们是 i.i.d. 的, 且 (X_1, Y_1) 的分布为二维正态分布 $N(a, b; \sigma_1^2, \sigma_2^2; \rho)$, 记 $\theta = (a, b, \sigma_1^2, \sigma_2^2, \rho)$, 则样本 $(X_1, Y_1), \dots, (X_n, Y_n)$ 的联合密度为

$$f(x_1, y_1, \dots, x_n, y_n; \theta) = \left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \right)^n \times \prod_{i=1}^n \left\{ \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x_i-a)^2}{\sigma_1^2} - \frac{2\rho(x_i-a)(y_i-b)}{\sigma_1\sigma_2} + \frac{(y_i-b)^2}{\sigma_2^2} \right] \right\} \right\}. \quad (1.2.8)$$

1.2.4 统计模型

所谓一个问题的统计模型(statistical model), 就是指研究该问题时所抽样本的分布, 也常称为概率模型或数学模型.

由于模型只取决于样本的分布, 故常把分布的名称作为模型的名称. 例如, 例 1.2.6 和例 1.2.8 中的模型可称为正态分布模型或简称正态模型, 例 1.2.7 中的模型称为指数分布模型等. 对上述的统计模型的定义, 作以下几点说明.

1. 统计模型的确定

在如上定义下, 模型是对指定的样本而言的. 因此只有明确了样本的产生方法, 并辅以必要的假定, 才能定下模型. 按模型的上述定义, 单说“问题的模型”还不很合适, 因为提出问题并非必须给定样本不可. 例如, 例 1.2.4 和例 1.2.5 中问题很明确, 都是为了估计 M 或估计废品率 $p = M/N$. 如何抽样, 可以有各种不同的方法. 事实上, 这两个例子说明对于同一个问题, 有两种不同的抽样方式, 确定了两个不同的统计模型. 因此把模型和样本紧密联系起来是必要的. 统计分析的依据是样本, 从统计上说, 只有规定了样本的分布, 问题才算真正明确了.

2. 很多性质不一样的问题, 可以归入到同一模型下

例如, 涉及测量误差的问题, 只要例 1.2.6 中叙述的假定误差服从正态分布的理由成立, 则都可以用正态模型 (1.2.6). 只要把这个模型中的统计问题研究清楚了, 就可以解决许多专业部门中的这类问题. 这就是数学的抽象. 可以说数理统计学的任务就是研究种种统计模型中所能提出来的种种统计问题. 许多不同专业部门中性质不一样的问题在一定条件下可归入到同一模型下, 对这一模型种种统计问题的研究反过来又可应用于解决这些专业部门提出的性质各异的同一类问题.

3. 同一模型下可以提出很多不同的统计问题

如例 1.2.6 的 $N(a, \sigma^2)$ 模型中, 有了样本 X_1, \dots, X_n , 并规定分布 (1.2.6) 后就有了一个统计模型. 在这个模型下可提出一些统计推断问题, 如在例 1.2.6 中, 问题是估计物重 a , 为了考察天平的精度, 可以提出估计 σ^2 , 当然还可以对 a 和 σ^2 提出假设检验和区间估计问题等.

1.2.5 统计推断

1. 参数和参数空间

前面说过, 统计模型就是样本分布, 当样本分布不全已知时才存在统计推断问题. 在例 1.2.6 中, 设样本 X_1, \dots, X_n i.i.d. $\sim N(a, \sigma^2)$, 其中 a 和 σ^2 未知. 在例 1.2.7 中, 设样本 X_1, \dots, X_n i.i.d. \sim 指数分布 $\text{Exp}(\lambda)$, 其中 λ 未知. 这些未知的量, 只有通过样本去估计. 统计学上把出现在样本分布中的未知的常数称为参数 (parameter), 如例 1.2.6 中的 a 和 σ 分别称为参数, 这时可称 (a, σ) 为参数向量 (parameter vector). 例 1.2.7 中的参数为 λ , 例 1.2.8 中有 5 个参数, 分别为 a, b, σ_1, σ_2 和 ρ .

在一些问题中参数虽未知, 但根据参数的性质可给出参数取值的范围. 参数取值的范围称为参数空间 (parametric space). 在例 1.2.6 中参数空间为 $\Theta = \{(a, \sigma) : a > 0, \sigma > 0\}$. 在例 1.2.7 中参数空间为 $\Theta = \{\lambda : \lambda > 0\}$. 在例 1.2.8 中参数空间为 $\Theta = \{(a, b, \sigma_1, \sigma_2, \rho) : a > 0, b > 0, \sigma_1 > 0, \sigma_2 > 0, |\rho| < 1\}$.

2. 样本分布族

样本分布既然包含未知参数, 则可能的样本分布就不止一个. 当参数取不同值时就得到不同的分布, 因此这些样本分布就构成一个分布族. 以例 1.2.7 为例, 样本分布族为

$$\mathcal{F} = \{f(x, \lambda) : \lambda > 0\},$$

其中 λ 为参数, 它的每一个可能的值对应于一个具体分布.

因此更确切地说, 统计模型就是样本分布族 (distribution family of the sample). 样本分布族, 连同其参数空间, 从总的方面确定了统计问题的范围. 分布族越小, 问题的确定度越高, 意味着可能作出更精确和更可靠的结论. 拿例 1.2.6 来说, 设问题是估计物重 a . 若反映天平精度的 σ 未知, 则分布族大些; 若 σ 已知 (例如, $\sigma = 1$), 则分布族小些. 可以设想: 有关天平精度的知识对估计 a 是有用的. 这可以说成: 在知道 σ 时, 估计 a 的问题更确定些, 且估计的精度比 σ 未知时要高.

3. 统计推断

从总体中抽取一定大小的样本去推断总体的概率分布的方法称为统计推断 (statistical inference).

数理统计是着手于样本, 着眼于总体, 其任务是用样本去推断总体. 当样本分布完全已知时是不存在任何统计推断问题的.

当样本分布形式已知, 但含有未知的实参数时, 统计推断的任务是确定未知参数的值, 这种情况下的统计推断问题称为参数统计推断问题.

在另一些问题中情况就要复杂一些. 这类问题中样本分布的形式未知, 有关统计推断问题称为非参数统计推断问题.

参数统计推断有种种不同的形式: 主要有参数估计和假设检验问题. 例如, 例 1.2.6 中样本分布 (亦即总体分布) $N(a, \sigma^2)$ 中, 当 a 和 σ^2 未知时, 从总体中抽取大小为 n 的样本 X_1, \dots, X_n , 对 a 和 σ^2 的取值作出估计, 或对断言 “ $a \leq 1$ ” 作出接受或拒绝这一假设的结论.

非参数统计推断问题中, 统计推断的主要任务是通过样本对总体的分布作出推断.

由于样本的随机性, 统计推断的结论不可能 100% 的正确, 但可以给出衡量推断正确程度的指标. 例如, 在例 1.2.6 中, 若用 $\bar{X} = \sum_{i=1}^n X_i/n$ 估计 a , 可以算出 \bar{X} 与 a 的偏差大于 c 的概率, 即 $P(|\bar{X} - a| > c)$, 作为用 \bar{X} 推断 a 的正确性的合理指标.

统计推断包括下列三方面内容: ① 提出种种统计推断的方法; ② 计算有关推断方法性能的数量指标, 如前述例子中用 \bar{X} 估计 $N(a, \sigma^2)$ 中的 a , 用 $P(|\bar{X} - a| > c)$ 表示推断性能的数量指标; ③ 在一定的条件和优良性准则下寻找最优的统计推断方法, 或证明某种统计推断方法是最优的.

4. 概率论与数理统计的关系

数理统计学与数学的其他学科有密切的关系. 数理统计学中用到很多近代数学知识, 主要的如数学分析与函数论、矩阵代数、组合数学, 也用到泛函分析、拓扑学和抽象代数的知识, 但与数理统计关系最密切的是概率论. 数理统计进行统计推断的第一步是利用概率论提供的种种模型对数据建模. 在统计推断中, 一旦确定了

模型, 就必须把数据看成来自具有一定概率模型的总体, 概率论中关于这种模型的理论结果, 就可用于统计推断的目的. 统计推断则致力于通过数据去检验选定的模型与实际数据是否符合, 以及确定模型中某些未知的参数. 因此在很大程度上可以说: 概率论是数理统计学的基础, 数理统计学是概率论的一种应用. 但是, 它们是两个并列数学分支学科, 并无从属关系.

1.3 统计量

1.3.1 统计量的定义

数理统计的任务是通过样本去推断总体. 而样本自身是一些杂乱无章的数字, 要对这些数字进行加工整理, 计算出一些有用的量, 这就如同为了织布, 首先要把棉花加工纺成纱, 然后利用纱去织布. 可以这样理解: 这种由样本算出来的量, 把样本中与所要解决的问题有关的信息集中起来了. 把这种量称为统计量, 定义如下.

定义 1.3.1 由样本算出的量称为统计量 (statistic). 或者说, 统计量是样本的函数.

对这一定义作如下几点说明:

(1) 统计量只与样本有关, 不能与未知参数有关. 例如, $X \sim N(a, \sigma^2)$, X_1, \dots, X_n 是从总体 X 中抽取的 i.i.d. 样本, 则 $\sum_{i=1}^n X_i$ 和 $\sum_{i=1}^n X_i^2$ 都是统计量, 当 a 和 σ^2 皆为未知参数时, $\sum_{i=1}^n (X_i - a)$ 和 $\sum_{i=1}^n X_i^2 / \sigma^2$ 都不是统计量.

(2) 由于样本具有两重性, 它既可以看成具体的数, 又可以看成随机变量 (或随机向量); 统计量是样本的函数, 因此统计量也具有两重性. 正因为统计量可视为随机变量 (或随机向量), 才有概率分布可言, 这是利用统计量进行统计推断的依据.

(3) 在什么问题中选用什么统计量, 要看问题的性质. 一般说来, 所提出的统计量应是最好地集中了样本中与所讨论问题有关的信息, 这不是容易做到的. 通常是从直观或一般性准则出发提出统计量, 再考察它是否在某种意义下较好地集中了样本中与所讨论问题有关的信息.

1.3.2 若干常用的统计量

1. 样本均值

设 X_1, \dots, X_n 是从某总体 X 中抽取的样本, 则称

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

为样本均值 (sample mean). 它反映了总体均值的信息.

2. 样本方差

设 X_1, \dots, X_n 是从某总体 X 中抽取的样本, 则称

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

为样本方差 (sample variance). 它反映了总体方差的信息, 而 S 称为样本标准差, 它反映了总体标准差的信息. 一些教科书上也采用

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

作为样本方差的定义. 用 S^2 定义样本方差的好处是 $E(S^2) = \sigma^2 = D(X)$ (参看例 3.1.1), 其中 $n-1$ 称为其自由度. 本教材以后各章皆以 S^2 作为样本方差的定义.

样本均值和样本方差是两个最常用的统计量, 它们具有如下三个性质:

$$(1) \sum_{i=1}^n (X_i - \bar{X}) = 0.$$

(2) 设非零实数 a 和 b 为常数, 作变换 $Y_i = aX_i + b$, $i = 1, 2, \dots, n$, 则 Y_1, \dots, Y_n 的样本均值 $\bar{Y} = a\bar{X} + b$, 其样本方差 $S_Y^2 = a^2 S_X^2$, 其中 S_X^2 和 S_Y^2 分别表示 X_1, \dots, X_n 和 Y_1, \dots, Y_n 的样本方差. 这个性质表明: 改变度量的原点与单位, 上述公式给出了新旧度量系统中样本均值和样本方差之间的关系. 利用这一公式可简化样本方差的计算, 如 X_1, \dots, X_n 是含有若干小数点的数字, 经上述公式变换后 (取适当的 a 和 b) 使样本数字变成简单的整数, 易于计算.

(3) 对于任何常数 c , 有

$$\sum_{i=1}^n (X_i - c)^2 \geq \sum_{i=1}^n (X_i - \bar{X})^2,$$

且等号只在 $c = \bar{X}$ 时成立. 这个性质表明, 在偏差平方和最小的准则下, 用总体均值 a 的 n 次测量值的算术平均值估计 a 是最好的.

3. 样本矩

设 X_1, \dots, X_n 为从总体 F 中抽取的样本, 则称

$$a_{n,k} = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots$$

为样本 k 阶原点矩. 特别 $k = 1$ 时, $a_{n,1} = \bar{X}$, 即样本均值. 称

$$m_{n,k} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad k = 2, 3, \dots$$

为样本 k 阶中心矩. 特别 $k = 2$ 时, $m_{n,2} = (n-1)S^2/n$.

样本的原点矩和中心矩统称为样本矩 (sample moments).

4. 二维随机向量的样本矩

设 $(X_1, Y_1), \dots, (X_n, Y_n)$ 为从二维总体 $F(x, y)$ 中抽取的样本, 则

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i, & S_X^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \\ \bar{Y} &= \frac{1}{n} \sum_{i=1}^n Y_i, & S_Y^2 &= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2, \\ S_{XY} &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})\end{aligned}$$

分别称为 X 和 Y 的样本均值、样本方差及 X 和 Y 的样本协方差 (sample covariance).

5. 次序统计量及其有关统计量

设 X_1, \dots, X_n 为从总体 F 中抽取的样本, 将其按大小排列为 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, 则 $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 称为样本 (X_1, \dots, X_n) 的次序统计量 (order statistics), $(X_{(1)}, \dots, X_{(n)})$ 的任一部分也称为次序统计量.

利用次序统计量可以定义下列统计量:

(1) 样本中位数:

$$m_{\frac{1}{2}} = \begin{cases} X_{((n+1)/2)}, & n \text{ 为奇数} \\ \frac{1}{2}[X_{(n/2)} + X_{(n/2+1)}], & n \text{ 为偶数} \end{cases} \quad (1.3.1)$$

称为样本中位数 (sample median), 它反映总体中位数的信息. 当总体分布关于某点对称时, 对称中心既是总体中位数又是总体均值, 此时 $m_{1/2}$ 也反映总体均值的信息.

(2) 极值: $X_{(1)}$ 和 $X_{(n)}$ 称为样本的极小值和极大值, 它们统称为样本极值 (extremum of sample). 极值统计量在关于灾害问题和材料试验的统计分析中是常用的统计量.

(3) 样本 p 分位数 ($0 < p < 1$) 可定义为 $X_{(m)}$, $m = [(n+1)p]$, 此处 $[a]$ 表示实数 a 的整数部分. 当 $p = 1/2$, n 为奇数时, 此定义与 (1) 中的样本中位数相同. 样本 p 分位数 (sample p -fractile) 反映了总体 p 分位数信息.

(4) 样本极差: $R = X_{(n)} - X_{(1)}$ 称为样本极差 (sample range), 它是反映总体分布散布程度的信息.

6. 样本变异系数

设 X_1, \dots, X_n 为从总体 F 中抽取的样本, 则称

$$\hat{\nu} = \frac{S_n}{\bar{X}} \quad (1.3.2)$$

为样本变异系数 (sample coefficient of variation). 它反映了总体变异系数 (population coefficient of variation) 的信息. 总体变异系数的定义是 $\nu = \sqrt{D(X)}/E(X)$, 它是衡量总体分布散布程度的量, 但这散布程度是以总体均值为单位来度量.

7. 样本偏度

设 X_1, \dots, X_n 为从总体 F 中抽取的样本, 则称

$$\hat{\beta}_1 = \frac{m_{n,3}}{m_{n,2}^{3/2}} = \sqrt{n} \sum_{i=1}^n (X_i - \bar{X})^3 / \left(\sum_{i=1}^n (X_i - \bar{X})^2 \right)^{3/2} \quad (1.3.3)$$

为样本偏度 (sample skewness). 它反映了总体偏度的信息, 总体偏度 (population skewness) 的定义是 $\beta_1 = \mu_3/\mu_2^{3/2}$, 此处 μ_i ($i = 2, 3$) 是总体的 i 阶中心矩. β_1 是反映总体分布的非对称性或“偏倚性”的一种度量. 正态分布 $N(a, \sigma^2)$ 的偏度为零.

8. 样本峰度

设 X_1, \dots, X_n 为从总体 F 中抽取的样本, 则称

$$\hat{\beta}_2 = \frac{m_{n,4}}{m_{n,2}^2} - 3 = n \sum_{i=1}^n (X_i - \bar{X})^4 / \left(\sum_{i=1}^n (X_i - \bar{X})^2 \right)^2 - 3 \quad (1.3.4)$$

为样本峰度 (sample kurtosis). 它反映了总体峰度的信息. 总体峰度 (population kurtosis) 的定义是 $\beta_2 = \mu_4/\mu_2^2 - 3$, 其中 μ_i ($i = 2, 4$) 是总体的 i 阶中心矩. β_2 是反映总体分布的密度函数在众数 (即密度函数的最大值点) 附近“峰”的尖峭程度的一种度量. 正态分布 $N(a, \sigma^2)$ 的峰度为零.

1.3.3 经验分布函数

定义 1.3.2 设 X_1, \dots, X_n 为自总体 $F(x)$ 中抽取的 i.i.d. 样本, 将其按大小排列为 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, 对任意实数 x , 称下列函数:

$$F_n(x) = \begin{cases} 0, & x \leq X_{(1)}, \\ \frac{k}{n}, & X_{(k)} < x \leq X_{(k+1)}, \quad k = 1, 2, \dots, n-1, \\ 1, & X_{(n)} < x \end{cases} \quad (1.3.5)$$

为经验分布函数 (empirical distribution function).

易见经验分布函数是单调、非降、左连续函数，具有分布函数的基本性质。它在 $x = X_{(k)}$, $k = 1, 2, \dots, n$ 处有间断，它是在每个间断点跳跃的幅度为 $1/n$ 的阶梯函数。 $F_n(x)$ 可以看成总体分布函数 $F(x) = P(X < x)$ 的一个估计量。若记示性函数

$$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & \text{其他}, \end{cases}$$

则 $F_n(x)$ 可表示为

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x)}(X_i). \quad (1.3.6)$$

由定义可知 $F_n(x)$ 是仅依赖于样本 X_1, X_2, \dots, X_n 的函数，因此它是统计量。它可能取值为 $0, 1/n, 2/n, \dots, (n-1)/n, 1$ 。若记 $Y_i = I_{(-\infty, x)}(X_i)$, $i = 1, 2, \dots, n$ ，则 $P(Y_i = 1) = F(x)$, $P(Y_i = 0) = 1 - F(x)$ ，且 Y_1, Y_2, \dots, Y_n i.i.d. $\sim b(1, F(x))$ ，故 $nF_n(x) = \sum_{i=1}^n Y_i \sim b(n, F(x))$ ，因此对 $k = 0, 1, \dots, n$ 有

$$P\left(F_n(x) = \frac{k}{n}\right) = P\left(\sum_{i=1}^n Y_i = k\right) = \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k}.$$

利用二项分布的性质可知对任一固定的 $x \in (-\infty, \infty)$, $F_n(x)$ 具有大样本性质：

(1) 由中心极限定理，则当 $n \rightarrow \infty$ 时有

$$\frac{\sqrt{n}(F_n(x) - F(x))}{\sqrt{F(x)(1 - F(x))}} \xrightarrow{\mathcal{L}} N(0, 1);$$

此处 $\xrightarrow{\mathcal{L}}$ 表示依分布收敛。

(2) 由 Bernoulli (或辛钦) 大数定律，则在 $n \rightarrow \infty$ 时有

$$F_n(x) \xrightarrow{P} F(x);$$

(3) 由 Borel 强大数定律，则有

$$P\left(\lim_{n \rightarrow \infty} F_n(x) = F(x)\right) = 1;$$

(4) 更进一步，有下列格里汶科定理 (Glivenko-Cantelli Theorem)：

定理 1.3.1 设 $F(x)$ 为 r.v. X 的分布函数， X_1, \dots, X_n 为取自总体 $F(x)$ 的简单随机样本， $F_n(x)$ 为其经验分布函数，记 $D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|$ ，则有

$$P\left(\lim_{n \rightarrow \infty} D_n = 0\right) = 1.$$

此定理的证明见文献 [4] 第 8 页。

注 1.3.1 上述定理中的 D_n 可用来衡量 $F_n(x)$ 和 $F(x)$ 之间在所有的 x 值上的最大差异程度, 格里汶科定理表明: 当 n 足够大时, 对所有的 x 值, 经验分布函数 $F_n(x)$ 与理论分布函数 $F(x)$ 之间只有很小的差别.

习题 1

1. 试举出一个有限总体的例子, 并指出其概率分布.
2. 试举出一个无限总体的例子; 并指出其概率分布.
3. 一个总体有 N 个元素, 其指标分别为 $a_1 > a_2 > \dots > a_N$, 指定自然数 $M < N$, $n < N$, 并设 $m = nM/N$ 为整数. 在 (a_1, \dots, a_M) 中不放回地随机抽出 m 个, 在 (a_{M+1}, \dots, a_N) 中不放回地随机抽出 $n - m$ 个. 写出所得样本的分布.
4. 一物体的重量 a 未知, 有两架天平可用, 其随机误差分别服从正态分布 $N(0, \sigma_1^2)$ 和 $N(0, \sigma_2^2)$, σ_1^2 和 σ_2^2 都未知. 先把物件在第一架天平上称两次得 X_1, X_2 , 再在第二架天平上称两次得 X_3, X_4 , 然后视 $|X_1 - X_2| \leq |X_3 - X_4|$ 或否而在第一架天平或第二架天平上再称 $n - 4$ 次得 X_5, \dots, X_n . 写出 (X_1, \dots, X_n) 的密度.
5. 设总体 X 服从两点分布 $b(1, p)$ (即 $P(X = 1) = p$, $P(X = 0) = 1 - p$), 其中 p 是未知参数, $\mathbf{X} = (X_1, \dots, X_5)$ 为从此总体中抽取的简单样本,
 - (1) 写出样本空间 \mathcal{X} 和 \mathbf{X} 的概率分布;
 - (2) 指出 $X_1 + X_2$, $\min_{1 \leq i \leq 5} X_i$, $X_5 + 2p$, $X_5 - E(X_1)$, $(X_5 - X_1)^2/D(X_1)$ 哪些是统计量, 哪些不是, 为什么?
 - (3) 若样本观察值 (X_1, \dots, X_n) 中有 m 个 1, $n - m$ 个 0, 求此样本的经验分布函数.
6. 设 $a \neq 0$ 和 b 皆为常数, 令 $y_i = ax_i + b$, $i = 1, 2, \dots, n$. 试证明 y_1, \dots, y_n 的样本均值 \bar{y} , 样本方差 S_y^2 和 x_1, \dots, x_n 的样本均值 \bar{x} , 样本方差 S_x^2 之间存在下列关系:

$$\bar{y} = a\bar{x} + b, \quad S_y^2 = a^2 S_x^2.$$

根据这个结果, 利用适当的变换, 求下列一组数据的样本均值和样本方差:

$$480, 550, 500, 590, 510, 560, 490, 600, 580.$$

7. 设容量为 10 的样本观察值为

$$0.5, 0.7, 0.2, 0.7, 0.4, 2.5, 1.5, -0.2, -0.5, 0.1.$$

试绘出经验分布函数的图形.

8. 设 X_1, \dots, X_n 是取自总体分布函数为 $F(x)$ 的样本, $F_n(x)$ 为其经验分布函数. 证明对任意给定的实数 x ,

$$P\left(\lim_{n \rightarrow \infty} F_n(x) = F(x)\right) = 1,$$

$$P\left(\lim_{n \rightarrow \infty} F_n(x+0) = F(x+0)\right) = 1.$$

9. 在正态总体 $N(50, 6^2)$ 中抽取容量为 36 的简单样本, 求样本均值 \bar{X} 落在 50.6 和 51.8 之间的概率.
10. 设 X_1, \dots, X_{100} 是取自总体 $X \sim N(\mu, 1)$ 的样本, 试确定常数 c , 使得对任意的 $\mu > 0$, $P(|\bar{X}| < c)$ 都不超过 0.05.
11. 设 X_1, \dots, X_n 是取自总体 $X \sim N(\mu, \sigma^2)$ 的简单样本, 假如要以 99.7% 的概率保证偏差 $|\bar{X} - \mu| < 0.1$, 试问在 $\sigma^2 = 0.5$ 时, 样本容量 n 应取多大?
12. 利用切比雪夫不等式求一枚均匀硬币需抛掷多少次才能使样本均值 \bar{X} 落在 0.4 和 0.6 之间的概率至少为 0.9 (此处 $X_i = 1$ 表示抛掷硬币出现正面, 否则 $X_i = 0$, $i = 1, \dots, n$)? 若用中心极限定理计算这个问题, 需抛掷的次数又是多少?

第2章 抽样分布及若干预备知识

2.1 引言

样本是随机变量, 有一定的概率分布, 称为样本分布. 统计量是样本的函数, 故它也是随机变量, 也有其概率分布. 这个概率分布原则上可由样本分布导出. 统计量的概率分布称为抽样分布.

研究统计量的性质和评价一个统计推断方法的优良性, 完全取决于其抽样分布的性质. 确定种种统计量的抽样分布是数理统计学的一项基本问题, 也是较难的问题. 近代统计学的奠基人之一, 英国统计学家 R.A. Fisher 曾把抽样分布、参数估计和假设检验看作统计推断的三个中心内容.

当总体 X 的分布类型已知时, 若对任一自然数 n , 都能导出统计量 $T = T(X_1, \dots, X_n)$ 分布的表达式, 这种分布称为 T 的精确抽样分布. 能求出统计量精确分布的情形不多. 已知的精确抽样分布大多是在正态条件下得到的. 有些情形下虽能求出统计量的精确分布, 但其表达式太复杂, 使用上不方便; 在更多的情形下统计量的精确抽样分布很难求出; 作为替代方法, 研究其极限分布. 当样本容量 $n \rightarrow \infty$ 时统计量的分布称为极限分布. 只要样本容量足够大, 且极限分布的形式比较简单, 就可以用统计量的极限分布作为其精确分布的近似.

本章将首先研究与正态总体有关的一些统计量的精确分布, 如正态总体样本均值和样本方差的分布、 χ^2 分布、 t 分布和 F 分布. 还将研究次序统计量的分布, 并对统计量极限分布作简要的介绍. 为了后面几章的需要, 本章还将介绍指数族、充分统计量和完全统计量的定义及有关性质.

2.2 正态总体样本均值和样本方差的分布

为方便讨论正态总体样本均值和样本方差的分布, 先给出正态随机变量的线性变换的分布.

2.2.1 正态变量线性函数的分布

先讨论正态变量的线性组合的分布.

定理 2.2.1 设随机变量 X_1, \dots, X_n 相互独立且 $X_k \sim N(a_k, \sigma_k^2)$, $k = 1, \dots, n$.

令 c_1, c_2, \dots, c_n 为常数, 记 $T = \sum_{k=1}^n c_k X_k$, 则 $T \sim N(\mu, \tau^2)$, 其中 $\mu = \sum_{k=1}^n c_k a_k$, $\tau^2 = \sum_{k=1}^n c_k^2 \sigma_k^2$.

证 因 $X_k \sim N(a_k, \sigma_k^2)$, $k = 1, 2, \dots, n$, 故其特征函数 (c.f.) 为

$$\varphi_k(t) = E(e^{itX_k}) = e^{ia_k t - \frac{1}{2}t^2\sigma_k^2},$$

所以 T 的 c.f. 为

$$\begin{aligned}\varphi(t) &= E(e^{itT}) = E\left(e^{it\sum_{k=1}^n c_k X_k}\right) = \prod_{k=1}^n E[e^{i(c_k t)X_k}] \\ &= \prod_{k=1}^n \left(e^{ic_k a_k t - \frac{1}{2}t^2 c_k^2 \sigma_k^2}\right) = e^{it\left(\sum_{k=1}^n c_k a_k\right) - \frac{1}{2}t^2\left(\sum_{k=1}^n c_k^2 \sigma_k^2\right)}.\end{aligned}$$

可见 $T \sim N(\mu, \tau^2)$, 其中 $\mu = \sum_{k=1}^n c_k a_k$, $\tau^2 = \sum_{k=1}^n c_k^2 \sigma_k^2$. 定理得证.

利用此定理, 容易得到如下推论.

推论 2.2.1 在定理 2.2.1 中, 若 $a_1 = \dots = a_n = a$, $\sigma_1^2 = \dots = \sigma_n^2 = \sigma^2$, 则有

$$T \sim N\left(a \sum_{k=1}^n c_k, \sigma^2 \sum_{k=1}^n c_k^2\right).$$

推论 2.2.2 在推论 2.2.1 中若取 $c_1 = \dots = c_n = 1/n$, 即 X_1, \dots, X_n i.i.d. $\sim N(a, \sigma^2)$, $T = \sum_{k=1}^n X_k/n = \bar{X}$, 则

$$\bar{X} \sim N(a, \sigma^2/n).$$

对独立同分布的正态变量的线性变换, 有如下结论.

定理 2.2.2 设 X_1, \dots, X_n i.i.d. $\sim N(a, \sigma^2)$, $\mathbf{X} = (X_1, \dots, X_n)', \mathbf{Y} = (Y_1, \dots, Y_n)', \mathbf{A} = (a_{ij})$ 为 $n \times n$ 的常数方阵, 记 $\mathbf{Y} = \mathbf{AX}$, 即

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \quad (2.2.1)$$

则有

(1) Y_1, \dots, Y_n 也是正态随机变量, 且

$$E(Y_i) = a \sum_{k=1}^n a_{ik}, \quad D(Y_i) = \sigma^2 \sum_{k=1}^n a_{ik}^2,$$

$$\text{Cov}(Y_i, Y_j) = \sigma^2 \sum_{k=1}^n a_{ik} a_{jk}.$$

(2) 特别当 $\mathbf{A} = (a_{ij})$ 为 n 阶正交阵时, Y_1, Y_2, \dots, Y_n 也相互独立且 $Y_i \sim N(\mu_i, \sigma^2)$, 此处 $\mu_i = a \sum_{k=1}^n a_{ik}$.

(3) 若再进一步假定 $a = 0$, 则 Y_1, Y_2, \dots, Y_n i.i.d. $\sim N(0, \sigma^2)$. 此事实说明: i.i.d. 服从正态分布 $N(0, \sigma^2)$ 的随机变量经正交变换后仍变为 i.i.d. 服从 $N(0, \sigma^2)$ 的随机变量.

证 (1) 由式 (2.2.1) 可知 $Y_i = \sum_{k=1}^n a_{ik} X_k$, 再由推论 2.2.1 可知

$$Y_i \sim N\left(a \sum_{k=1}^n a_{ik}, \sigma^2 \sum_{i=1}^n a_{ik}^2\right).$$

因此有 $E(Y_i) = a \sum_{k=1}^n a_{ik}$, $D(Y_i) = \sigma^2 \sum_{k=1}^n a_{ik}^2$. 而

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= E[(Y_i - EY_i)(Y_j - EY_j)] \\ &= \sum_{k=1}^n \sum_{l=1}^n a_{ik} a_{jl} E[(X_k - a)(X_l - a)] \\ &= \sum_{k=1}^n \sum_{l=1}^n a_{ik} a_{jl} \cdot \delta_{kl} \sigma^2 = \sigma^2 \sum_{k=1}^n a_{ik} a_{jk}, \end{aligned}$$

此处 $\delta_{kl} = 1$, 当 $k = l$; $\delta_{kl} = 0$, 当 $k \neq l$. 故 (1) 得证.

(2) 当 $\mathbf{A} = (a_{ij})$ 为 n 阶正交阵时, $\sum_{k=1}^n a_{ik}^2 = 1$, 则 $\mu_i = E(Y_i) = a \sum_{k=1}^n a_{ik}$, $D(Y_i) = \sigma^2 \sum_{k=1}^n a_{ik}^2 = \sigma^2$. 又由正交阵 \mathbf{A} 的不同行和列的正交性可知: 当 $i \neq j$ 时, $\sum_{k=1}^n a_{ik} a_{jk} = 0$, 故 $\text{Cov}(Y_i, Y_j) = 0$. 因此 Y_1, \dots, Y_n 相互独立且 $Y_i \sim N(\mu_i, \sigma^2)$. 从而 (2) 得证.

(3) 特别若 $a = 0$, 由 (2) 显见 $\mu_i = 0$, $i = 1, \dots, n$, 故此时 Y_1, \dots, Y_n i.i.d. $\sim N(0, \sigma^2)$, 故 (3) 得证. 定理证毕.

下面讨论正态变量样本均值和样本方差的分布.

2.2.2 正态变量样本均值和样本方差的分布

下述定理给出了正态变量样本均值和样本方差的分布和它们的独立性.

定理 2.2.3 设 X_1, X_2, \dots, X_n i.i.d. $\sim N(a, \sigma^2)$, $\bar{X} = \sum_{i=1}^n X_i/n$ 和 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ 分别为样本均值和样本方差, 则有

$$(1) \bar{X} \sim N(a, \sigma^2/n);$$

$$(2) (n-1)S^2/\sigma^2 \sim \chi_{n-1}^2;$$

(3) \bar{X} 和 S^2 独立.

此处 χ_{n-1}^2 表示自由度为 $n-1$ 的卡方分布, 它是通过 $n-1$ 个相互独立的标准正态随机变量平方和的分布来定义的 (参见定义 2.4.1).

证 (1) 由推论 2.2.2 立得 $\bar{X} \sim N(a, \sigma^2/n)$. 下面证 (2) 和 (3). 设

$$A = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

为一正交阵 (这一正交阵的存在性由 Schmidt 正交化方法保证), 作正交变换 $\mathbf{Y} = A\mathbf{X}$, 其中 \mathbf{Y} 和 \mathbf{X} 如定理 2.2.2 所示, 故有

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i = \sqrt{n}\bar{X},$$

由正交变换保持向量长度不变可知

$$Y_1^2 + \cdots + Y_n^2 = X_1^2 + \cdots + X_n^2.$$

所以

$$\begin{aligned} (n-1)S^2 &= \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 \\ &= \sum_{i=1}^n Y_i^2 - Y_1^2 = \sum_{i=2}^n Y_i^2. \end{aligned} \tag{2.2.2}$$

由定理 2.2.2 (2) 可知 Y_1, \dots, Y_n 相互独立且 $Y_i \sim N(\mu_i, \sigma^2)$, $i = 2, \dots, n$. 再由 A 的行向量正交性得

$$\mu_i = a \sum_{k=1}^n a_{ik} = \sqrt{n}a \cdot \sum_{k=1}^n \frac{1}{\sqrt{n}} \cdot a_{ik} = 0, \tag{2.2.3}$$

即 Y_2, \dots, Y_n i.i.d. $\sim N(0, \sigma^2)$, 故 $Y_2/\sigma, \dots, Y_n/\sigma$ i.i.d. $\sim N(0, 1)$, 由式 (2.2.2) 得

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=2}^n (Y_i/\sigma)^2 \sim \chi_{n-1}^2.$$

故 (2) 得证.

由上述 (2) 的证明中可知 Y_1, Y_2, \dots, Y_n 相互独立, S^2 只和 Y_2, \dots, Y_n 有关, \bar{X} 只和 Y_1 有关, 因此 \bar{X} 和 S^2 独立, 故 (3) 得证. 定理证毕.

2.3 次序统计量的分布

次序统计量的定义在 1.3 节中已给出, 即若 X_1, X_2, \dots, X_n i.i.d. $\sim F$, 将其按大小排列为 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, 则称 $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 为样本 (X_1, X_2, \dots, X_n) 的次序统计量. 它的任何一部分, 如 $X_{(i)}$ 和 $(X_{(i)}, X_{(j)})$ ($1 \leq i < j \leq n$) 等也称为次序统计量. 本节将导出在 F 有密度时单个次序统计量的分布、次序统计量的联合分布和极差的分布. F 有密度的情形 $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ 几乎处处成立.

2.3.1 单个次序统计量的分布

设 X_1, X_2, \dots, X_n i.i.d. $\sim F$, f 为 F 的密度, 求 $X_{(m)}$ 的分布, 此处 $1 \leq m \leq n$. 因为

$$\begin{aligned} F_m(x) &= P(X_{(m)} < x) = P(X_1, X_2, \dots, X_n \text{ 中至少有 } m \text{ 个} < x) \\ &= \sum_{i=m}^n P(X_1, X_2, \dots, X_n \text{ 恰有 } i \text{ 个} < x). \end{aligned}$$

若记 $A_i = \{X_i < x\}$, $i = 1, 2, \dots, n$, 则 $P(A_i) = P(X_i < x) = F(x)$, 故知

$$\{X_1, X_2, \dots, X_n \text{ 中恰有 } i \text{ 个} < x\} = \{\text{事件 } A_1, \dots, A_n \text{ 中恰有 } i \text{ 个发生}\}.$$

这一事件的概率可用二项分布 $b(n, F(x))$ 来表示, 所以

$$P\{X_1, X_2, \dots, X_n \text{ 中恰有 } i \text{ 个} < x\} = \binom{n}{i} (F(x))^i (1 - F(x))^{n-i}.$$

因此

$$F_m(x) = \sum_{i=m}^n \binom{n}{i} (F(x))^i (1 - F(x))^{n-i}.$$

利用恒等式 (见本章习题 6)

$$\sum_{i=m}^n \binom{n}{i} p^i (1-p)^{n-i} = m \binom{n}{m} \int_0^p t^{m-1} (1-t)^{n-m} dt,$$

可知

$$\begin{aligned} F_m(x) &= P(X_{(m)} < x) = \sum_{i=m}^n \binom{n}{i} (F(x))^i (1 - F(x))^{n-i} \\ &= m \binom{n}{m} \int_0^{F(x)} t^{m-1} (1-t)^{n-m} dt, \end{aligned}$$

从而其密度函数为

$$f_m(x) = F'_m(x) = m \binom{n}{m} (F(x))^{m-1} (1 - F(x))^{n-m} f(x). \quad (2.3.1)$$

特别当取 $m = 1$ 得到样本极小值 $X_{(1)}$ 的分布函数和密度函数为

$$\begin{aligned} F_1(x) &= P(X_{(1)} < x) = 1 - (1 - F(x))^n, \\ f_1(x) &= n(1 - F(x))^{n-1} f(x). \end{aligned} \quad (2.3.2)$$

当取 $m = n$ 时得到样本极大值 $X_{(n)}$ 的分布函数和密度函数为

$$F_n(x) = P(X_{(n)} < x) = (F(x))^n, \quad f_n(x) = n(F(x))^{n-1} f(x). \quad (2.3.3)$$

2.3.2 次序统计量的联合分布

1. n 个次序统计量 $(X_{(1)}, \dots, X_{(n)})$ 的联合分布

令 $y_i = x_{(i)}, i = 1, \dots, n$, 则

$$G(y_1, y_2, \dots, y_n) = P(X_{(1)} < y_1, X_{(2)} < y_2, \dots, X_{(n)} < y_n)$$

$$= \begin{cases} n! P(X_{j_1} < y_1, X_{j_2} < y_2, \dots, X_{j_n} < y_n), & y_1 < y_2 < \dots < y_n, \\ 0, & \text{其他}, \end{cases}$$

其中 $X_{j_1} < \dots < X_{j_n}$, (j_1, \dots, j_n) 是 $(1, \dots, n)$ 的任一排列. n 个次序统计量的联合密度为

$$g(y_1, y_2, \dots, y_n) = \begin{cases} n! f(y_1) f(y_2) \cdots f(y_n), & y_1 < y_2 < \dots < y_n, \\ 0, & \text{其他}. \end{cases} \quad (2.3.4)$$

2. 两个次序统计量 $(X_{(i)}, X_{(j)})$ 的联合分布

利用本章习题 7 的结果可知

$$\int_{a < x_1 < \dots < x_n < b} \cdots \int f(x_1) \cdots f(x_n) dx_1 \cdots dx_n = \frac{1}{n!} [F(b) - F(a)]^n, \quad (2.3.5)$$

其中 F 和 f 分别为 r.v. X 的分布函数和密度函数. 由这一事实可证明下列定理.

定理 2.3.1 设 X_1, \dots, X_n i.i.d. $\sim F$, F 有密度 f . 记 $\mathbf{X} = (X_1, \dots, X_n)$, 则 \mathbf{X} 的次序统计量 $(X_{(1)}, \dots, X_{(n)})$ 中的任意两个 $(X_{(i)}, X_{(j)})$ ($i < j$) 的联合密度为

$$f_{ij}(x, y) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(x))^{i-1} (F(y) - F(x))^{j-i-1} \\ \times (1 - F(y))^{n-j} f(x) f(y), & x < y, \\ 0, & \text{其他}. \end{cases} \quad (2.3.6)$$

此处 $x = x_{(i)}$, $y = x_{(j)}$.

证 由于 $f_{ij}(x, y)$ 是 $(X_{(1)}, \dots, X_{(n)})$ 的联合密度 $g(x_{(1)}, \dots, x_{(n)})$ 的边缘密度, 因此利用式 (2.3.5) 可知当 $x < y$ 时,

$$\begin{aligned} & f_{ij}(x, y) \\ &= \int_{-\infty < x_1 < \dots < x_n < +\infty} n! f(x_1) \cdots f(x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_{j-1} dx_{j+1} \cdots dx_n \\ &= n! \int_{-\infty < x_1 < \dots < x_{i-1} < x} f(x_1) \cdots f(x_{i-1}) f(x) dx_1 \cdots dx_{i-1} \\ &\quad \times \int_{x < x_{i+1} < \dots < x_{j-1} < y} f(x_{i+1}) \cdots f(x_{j-1}) f(y) dx_{i+1} \cdots dx_{j-1} \\ &\quad \times \int_{y < x_{j+1} < \dots < x_n < +\infty} f(x_{j+1}) \cdots f(x_n) dx_{j+1} \cdots dx_n \\ &= n! \frac{f(x)}{(i-1)!} (F(x))^{i-1} \cdot \frac{f(y)}{(j-i-1)!} [F(y) - F(x)]^{j-i-1} \cdot \frac{(1-F(y))^{n-j}}{(n-j)!} \\ &= \frac{n! f(x) f(y)}{(i-1)!(j-i-1)!(n-j)!} (F(x))^{i-1} (F(y) - F(x))^{j-i-1} (1-F(y))^{n-j}. \end{aligned}$$

而当 $x \geq y$ 时 $f_{ij}(x, y) = 0$, 这就证明了定理.

用类似方法可求 r ($r > 2$) 个次序统计量 $(X_{(i_1)}, \dots, X_{(i_r)})$ ($i_1 < i_2 < \dots < i_k$) 的联合分布.

2.3.3 极差的分布

样本极差的定义在 1.3 节中已给出, 即 $R = X_{(n)} - X_{(1)}$ 称为样本极差. 为求样本极差的分布, 先求 $V = X_{(j)} - X_{(i)}$, $j > i$ 的分布, 当取 $j = n$, $i = 1$ 即可得到样本极差的分布. 具体方法如下.

作下列变换:

$$\begin{cases} V = X_{(j)} - X_{(i)}, \\ Z = X_{(i)} \end{cases} \iff \begin{cases} X_{(i)} = Z, \\ X_{(j)} = V + Z. \end{cases}$$

变换的 Jacobi 行列式为 $J = \left| \frac{\partial(X_{(i)}, X_{(j)})}{\partial(V, Z)} \right| = 1$, $(X_{(i)}, X_{(j)})$ 的联合分布密度由式 (2.3.6) 给出, 将其中 x 用 z 代替, y 用 $v+z$ 代替, 再乘上变换的 Jacobi 行列式的

绝对值 $|J|$, 得到 (V, Z) 的联合密度

$$g_{ij}(v, z) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(z))^{i-1} (F(v+z) - F(z))^{j-i-1} \\ \quad \times (1 - F(v+z))^{n-j} f(z) f(v+z), & v > 0, \\ 0, & \text{其他.} \end{cases} \quad (2.3.7)$$

从而易知 V 的密度为

$$g_V(v) = \int_{-\infty}^{\infty} g_{ij}(v, z) dz.$$

特别地, 在式 (2.3.7) 中取 $i = 1, j = n$ 得到 $(R, Z) = (V, Z)$ 的联合密度

$$g_{1n}(r, z) = \begin{cases} \frac{n!}{(n-2)!} (F(r+z) - F(z))^{n-2} f(r+z) f(z), & r > 0, \\ 0, & r \leq 0. \end{cases} \quad (2.3.8)$$

而 $R = X_{(n)} - X_{(1)}$ 的边缘密度为 $g_R(r) = \int_{-\infty}^{\infty} g_{1n}(r, z) dz.$

2.3.4 均匀分布情形

设 X_1, X_2, \dots, X_n i.i.d. \sim 均匀分布 $U(0, 1)$, 其分布函数和密度函数分别为

$$F(x) = \begin{cases} 0, & x \leq 0, \\ x, & 0 < x \leq 1, \\ 1, & x > 1 \end{cases} \quad \text{和} \quad f(x) = \begin{cases} 1, & 0 < x \leq 1, \\ 0, & \text{其他.} \end{cases}$$

设 $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 为样本 X_1, X_2, \dots, X_n 的次序统计量, 由式 (2.3.1) 可知次序统计量 $X_{(m)}$ 的密度函数为

$$f_m(x) = m \binom{n}{m} x^{m-1} (1-x)^{n-m} I_{(0,1)}(x). \quad (2.3.9)$$

由式 (2.3.6) 可知 $(X_{(i)}, X_{(j)})$ ($i < j$) 的联合密度为

$$f_{ij}(x, y) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!} x^{i-1} (y-x)^{j-i-1} (1-y)^{n-j}, & 0 < x < y < 1, \\ 0, & \text{其他.} \end{cases} \quad (2.3.10)$$

由式 (2.3.4) 可知 $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 的联合密度为

$$f(x_{(1)}, x_{(2)}, \dots, x_{(n)}) = \begin{cases} n!, & 0 < x_{(1)} < \dots < x_{(n)} < 1, \\ 0, & \text{其他.} \end{cases} \quad (2.3.11)$$

在式(2.3.7)中,令 $F(z)=z$, $0 < z < 1$, $F(v+z)=v+z$, $0 < v+z < 1$,得到在均匀分布 $U(0,1)$ 场合 (V,Z) 的联合密度

$$g_{ij}(v,z) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!} z^{i-1} v^{j-i-1} [1-(v+z)]^{n-j}, \\ \quad 0 < z < 1, 0 < v+z < 1, \\ 0, \quad \text{其他.} \end{cases} \quad (2.3.12)$$

此时 $V=X_{(j)}-X_{(i)}$ 的边缘密度,通过计算积分 $\int_0^{1-v} g_{ij}(v,z) dz$ 得

$$g_{ij}(v) = \begin{cases} \frac{n!}{(j-i-1)!(n-j+i)!} v^{j-i-1} (1-v)^{n-j+i}, & 0 < v < 1, \\ 0, & \text{其他.} \end{cases} \quad (2.3.13)$$

特别地,极差 $R=X_{(n)}-X_{(1)}$ 的密度函数 $g_R(r)=g_{1n}(r)$ 为将式(2.3.13)中的 v 换成 r ,将 j 和 i 分别用 n 和1代替得到

$$g_R(r) = \begin{cases} n(n-1)r^{n-2}(1-r), & 0 < r < 1, \\ 0, & \text{其他.} \end{cases} \quad (2.3.14)$$

2.4 χ^2 分布, t 分布和 F 分布

能求出统计量的抽样分布的确切而且具有简单表达式的情形并不多,一般都较难.所幸的是,在总体分布为正态情形,许多重要统计量的抽样分布可以求得,这些多与下面讨论的三个分布有密切关系.这三个分布在后面几章中有重要应用.下面将分别介绍它们的定义和有关性质.

2.4.1 χ^2 分布

定义 2.4.1 设 X_1, X_2, \dots, X_n i.i.d. $\sim N(0,1)$,则称

$$\xi = \sum_{i=1}^n X_i^2$$

是自由度为 n 的 χ^2 变量,其分布称为自由度为 n 的 χ^2 分布,记为 $\xi \sim \chi_n^2$.

χ^2 变量的概率密度函数由下面的定理给出.

定理 2.4.1 设随机变量 ξ 是自由度 n 的 χ^2 随机变量,则其概率密度函数为

$$g_n(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (2.4.1)$$

证 由于 X_1, X_2, \dots, X_n i.i.d. $\sim N(0, 1)$, 故其联合密度为

$$f(x_1, x_2, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2} \sum_{i=1}^n x_i^2\right\}.$$

令 r.v. $\xi = \sum_{i=1}^n X_i^2$ 的分布函数为 $G_n(x)$, 则有

$$G_n(x) = P\left(\sum_{i=1}^n X_i^2 < x\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \int \cdots \int \exp\left\{-\frac{1}{2} \sum_{i=1}^n x_i^2\right\} dx_1 \cdots dx_n.$$

作 n 维球坐标变换

$$\begin{cases} x_1 = \rho \cos \theta_1 \cos \theta_2 \cdots \cos \theta_{n-2} \cos \theta_{n-1}, \\ x_2 = \rho \cos \theta_1 \cos \theta_2 \cdots \cos \theta_{n-2} \sin \theta_{n-1}, \\ \cdots \\ x_n = \rho \sin \theta_1. \end{cases} \quad (2.4.2)$$

变换的 Jacobi 行列式的绝对值为

$$|J| = \left| \frac{\partial(x_1, x_2, \dots, x_n)}{\partial(\rho, \theta_1, \dots, \theta_{n-1})} \right| = \rho^{n-1} D(\theta_1 \cdots \theta_{n-1}),$$

其中 $D(\theta_1, \dots, \theta_{n-1})$ 表示 $\theta_1, \dots, \theta_{n-1}$ 的某个函数; $0 < \rho < \sqrt{x}$; $-\pi/2 < \theta_i < \pi/2$, $i = 1, \dots, n-2$; $-\pi < \theta_{n-1} < \pi$. 因此有

$$\begin{aligned} G_n(x) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_0^{\sqrt{x}} \rho^{n-1} e^{-\frac{\rho^2}{2}} d\rho \int_{-\pi}^{\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} D(\theta_1 \cdots \theta_{n-1}) d\theta_1 \cdots d\theta_{n-1} \\ &= C_n \int_0^{\sqrt{x}} \rho^{n-1} e^{-\frac{\rho^2}{2}} d\rho, \end{aligned}$$

其中

$$C_n = \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_{-\pi}^{\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} D(\theta_1 \cdots \theta_{n-1}) d\theta_1 \cdots d\theta_{n-1}.$$

令 $y = \rho^2$, 则 $d\rho = 1/(2\sqrt{y}) \cdot dy$, 故有

$$G_n(x) = \frac{1}{2} C_n \int_0^x y^{\frac{n}{2}-1} e^{-\frac{y}{2}} dy. \quad (2.4.3)$$

下面确定 C_n . 由于

$$1 = G_n(+\infty) = \frac{1}{2} C_n \int_0^\infty y^{\frac{n}{2}-1} e^{-\frac{y}{2}} dy = C_n 2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right),$$

故有 $C_n = 1/[2^{n/2-1}\Gamma(n/2)]$. 将其代入式 (2.4.3) 得

$$G_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^x y^{\frac{n}{2}-1} e^{-\frac{y}{2}} dy.$$

因此 ξ 的密度函数 $g_n(x)$ 为

$$g_n(x) = G'_n(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

定理证毕.

χ_n^2 的密度函数 $g_n(x)$ 形状如图 2.4.1 所示.

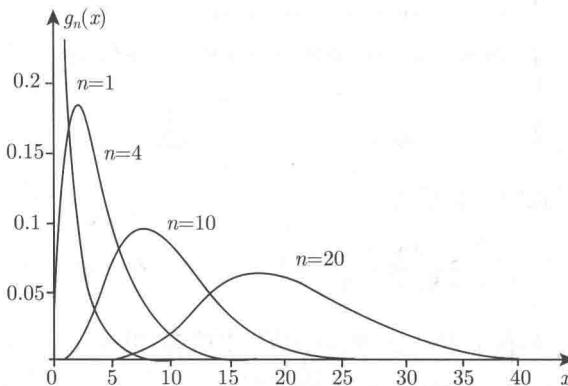


图 2.4.1 χ_n^2 的密度函数 $g_n(x)$ 形状图

χ_n^2 密度函数的支撑集 (即使密度函数为正的自变量的集合) 为 $(0, +\infty)$, 由图 2.4.1 可见当自由度 n 越大, χ_n^2 的密度曲线越趋于对称 (由中心极限定理知当 $n \rightarrow \infty$ 时, 它趋于正态分布), n 越小, 曲线越不对称. 当 $n = 1, 2$ 时曲线是单调下降趋于 0. 当 $n \geq 3$ 时曲线有单峰, 从 0 开始先单调上升, 在一定位置达到峰值, 然后单调下降趋于 0.

设 $\xi \sim \chi_n^2$, $0 < \alpha < 1$, 令 $P(\xi > c) = \alpha$, 则称 $c = \chi_n^2(\alpha)$ 为 χ_n^2 分布的上侧 α 分位数 (图 2.4.2). α 通常取较小的数, 如 $\alpha = 0.05, 0.01$ 等. 当 α 和 n 给定时, 可查附表 3 求出 $\chi_n^2(\alpha)$ 之值, 如 $\chi_{10}^2(0.01) = 23.209$, $\chi_6^2(0.05) = 12.592$ 等.

注 2.4.1 若记 $\Gamma(\alpha, \lambda)$ 为具有下列密度函数的概率分布

$$p(x; \alpha, \lambda) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

则自由度为 n 的 χ^2 分布与 Γ 分布的关系为 $\xi = \sum_{i=1}^n X_i^2 \sim \Gamma(n/2, 1/2)$. 也可以利用这一关系给出 χ^2 分布的定义, 即若 r.v. ξ 的概率密度函数为 $\Gamma(n/2, 1/2)$, 则称 ξ 为服从自由度为 n 的 χ^2 分布. 另一方面, 若 $Y \sim \Gamma(\alpha, \lambda)$, 则 $Z = 2\lambda Y \sim \chi_{2\alpha}^2$.

χ^2 变量具有下列性质:

- (1) 设 r.v. $\xi \sim \chi_n^2$, 则 ξ 的特征函数为 $\varphi(t) = (1 - 2it)^{-n/2}$;
- (2) r.v. ξ 的均值和方差分别为 $E(\xi) = n$, $D(\xi) = 2n$.

上述两条性质的证明留给读者作练习.

- (3) 设 $Z_1 \sim \chi_{n_1}^2$, $Z_2 \sim \chi_{n_2}^2$ 且 Z_1 和 Z_2 独立, 则 $Z_1 + Z_2 \sim \chi_{n_1+n_2}^2$.

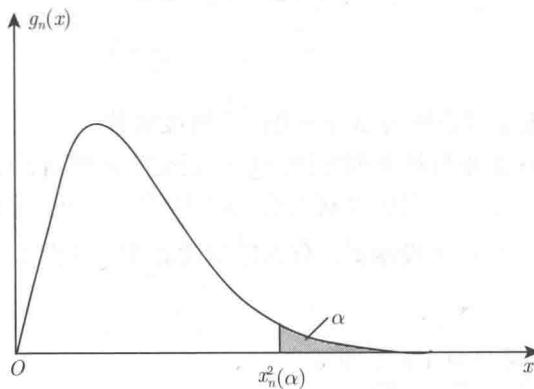


图 2.4.2 χ_n^2 的上侧 α 分位数

证 这一性质可用特征函数去证明. 此处从 χ^2 分布的定义出发给出另一个简单证明:

由定义 $Z_1 = X_1^2 + \cdots + X_{n_1}^2$, 其中 X_1, X_2, \dots, X_{n_1} i.i.d. $\sim N(0, 1)$, 同理 $Z_2 = X_{n_1+1}^2 + \cdots + X_{n_1+n_2}^2$, 其中 $X_{n_1+1}, X_{n_1+2}, \dots, X_{n_1+n_2}$ i.i.d. $\sim N(0, 1)$, 再由 Z_1 和 Z_2 的独立性可知 $X_1, X_2, \dots, X_{n_1}, X_{n_1+1}, \dots, X_{n_1+n_2}$ i.i.d. $\sim N(0, 1)$. 因此

$$Z_1 + Z_2 = X_1^2 + \cdots + X_{n_1}^2 + X_{n_1+1}^2 + \cdots + X_{n_1+n_2}^2,$$

则按定义有 $Z_1 + Z_2 \sim \chi_{n_1+n_2}^2$.

推广 若 $Z_i \sim \chi_{n_i}^2$, $i = 1, 2, \dots, k$ 且 Z_1, Z_2, \dots, Z_k 相互独立, 则 $\sum_{i=1}^k Z_i \sim \chi_{n_1+\cdots+n_k}^2$.

*非中心 χ^2 分布简介

下面将对非中心的 χ^2 分布的定义和性质作一简略的介绍.

定义 2.4.2 设随机变量 X_1, X_2, \dots, X_n 相互独立, $X_i \sim N(a_i, 1)$, a_i ($i = 1, \dots, n$) 不全为 0. 记 $Y = \sum_{i=1}^n X_i^2$, 则称 Y 的分布是自由度为 n 和非中心参数为

$\delta = \sqrt{\sum_{i=1}^n a_i^2}$ 的非中心 χ^2 分布, 记为 $Y \sim \chi_{n,\delta}^2$. 特别当 $\delta = 0$ 时称为中心的 χ^2 分布, 即前面所述 χ_n^2 分布.

若 $Y \sim \chi_{n,\delta}^2$, 则其密度函数为

$$g(x) = \begin{cases} e^{-\delta^2/2} \sum_{i=0}^{\infty} \frac{1}{i!} \left(\frac{\delta^2}{2}\right)^i \frac{x^{i+n/2-1}}{2^{i+n/2} \Gamma(n/2+i)} e^{-x/2}, & x > 0, \\ 0, & x \leq 0 \end{cases} \quad (2.4.4)$$

$$= \begin{cases} e^{-\delta^2/2} \sum_{i=0}^{\infty} \frac{(\delta^2/2)^i}{i!} \chi^2(x, 2i+n), & x > 0, \\ 0, & x \leq 0. \end{cases}$$

此处 $\chi^2(x, 2i+n)$ 表示自由度为 $2i+n$ 的 χ^2 密度函数.

密度函数 (2.4.4) 的推导按下列方法: ① 作正交变换使 $X_1^2 + \dots + X_n^2 = Y_1^2 + Z$, 其中 $Y_1 \sim N(\delta, 1)$, $Z \sim \chi_{n-1}^2$; ② 再利用求 r.v. 和的分布公式求出 $Y_1^2 + Z$ 的密度函数, 此即 $Y = \sum_{i=1}^n X_i^2$ 的密度函数. 有兴趣的读者不妨按此方法推导这一密度函数.

非中心的 χ^2 变量具有下列性质:

- (1) 若 $Y \sim \chi_{n,\delta}^2$, 则 Y 的特征函数为 $\varphi(t) = (1 - 2it)^{-\frac{n}{2}} e^{-\frac{i\delta^2 t}{1-2it}}$;
- (2) 若 $Y \sim \chi_{n,\delta}^2$, 则 $E(Y) = n + \delta^2$, $D(Y) = 2n + 4\delta^2$;
- (3) 若 Y_1, \dots, Y_k 相互独立, $Y_i \sim \chi_{n_i, \delta_i}^2$, $i = 1, 2, \dots, k$, 则 $\sum_{i=1}^k Y_i \sim \chi_{n,\delta}^2$, 此处
 $n = \sum_{i=1}^k n_i$, $\delta = \sqrt{\delta_1^2 + \delta_2^2 + \dots + \delta_k^2}$.

2.4.2 t 分布

定义 2.4.3 设 r.v. $X \sim N(0, 1)$, $Y \sim \chi_n^2$ 且 X 和 Y 独立, 则称

$$T = \frac{X}{\sqrt{Y/n}}$$

是自由度为 n 的 t 变量, 其分布称为自由度为 n 的 t 分布, 记为 $T \sim t_n$.

t 变量的概率密度函数由下面的定理给出.

定理 2.4.2 设随机变量 $T \sim t_n$, 则其密度函数为

$$t_n(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2}) \sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad -\infty < x < \infty. \quad (2.4.5)$$

定理的证明采用下列方法: ① 求出 $\sqrt{Y/n}$ 的密度函数, 其中 $Y \sim \chi_n^2$; ② 利用求独立随机变量商的密度函数公式求出 T 的密度函数. 定理的证明留给读者作为练习.

t_n 的密度函数与标准正态分布 $N(0, 1)$ 密度很相似, 它们都是关于原点对称, 单峰的偶函数, 在 $x = 0$ 处达到极大. 但 t_n 的峰值低于 $N(0, 1)$ 的峰值, t_n 的密度函数尾部都要比 $N(0, 1)$ 的两侧尾部粗一些, 如图 2.4.3 所示. 容易证明: $\lim_{n \rightarrow \infty} t_n(x) = \varphi(x)$, 此处 $\varphi(x)$ 是 $N(0, 1)$ 变量的密度函数.

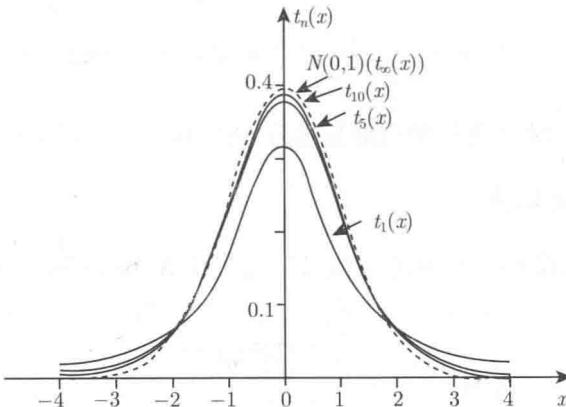


图 2.4.3 t_n 的密度函数 $t_n(x)$ 形状图

设 $T \sim t_n$, $0 < \alpha < 1$, 令 $P(|T| > c) = \alpha$, 则称 $c = t_n(\alpha/2)$ 为自由度为 n 的 t 分布的双侧上 α 分位数, 如图 2.4.4 所示. 当 α 和 n 给定时, 可查附表 2 求出 $t_n(\alpha)$, $t_n(\alpha/2)$ 等. 例如, $t_{12}(0.05) = 1.782$, $t_9(0.025) = 2.262$ 等.

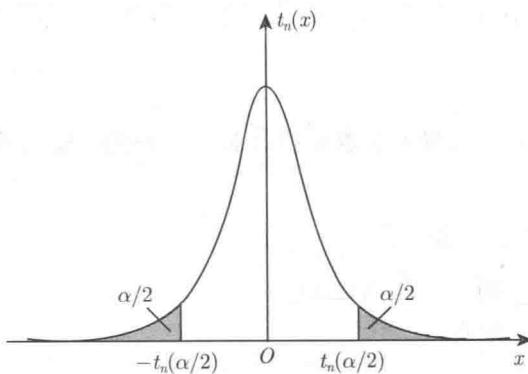


图 2.4.4 t_n 的双侧 α 分位数

t 分布是英国统计学家 W. S. Gosset 在 1908 年以笔名 Student 发表的论文中提出的, 故后人称为“学生氏 (student) 分布”或“ t 分布”.

t 变量具有下列的性质:

(1) 若 r.v. $T \sim t_n$, 则 $E(T^r)$ 只有当 $r < n$ ($n > 1$) 时存在, 且

$$E(T^r) = \begin{cases} n^{\frac{r}{2}} \frac{\Gamma(\frac{r+1}{2}) \Gamma(\frac{n-r}{2})}{\Gamma(\frac{n}{2}) \Gamma(\frac{1}{2})}, & r \text{ 为偶数,} \\ 0, & r \text{ 为奇数.} \end{cases}$$

特别当 $n \geq 2$ 时, $E(T) = 0$. 当 $n \geq 3$ 时, $D(T) = n/(n-2)$.

(2) 当 $n=1$ 时 t 分布就是柯西 (Cauchy) 分布, 此时式 (2.4.5) 变为

$$t_1(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < +\infty.$$

(3) 当 $n \rightarrow \infty$ 时, t 变量的极限分布为 $N(0, 1)$.

*非中心 t 分布简介

定义 2.4.4 设 r.v. $X \sim N(\delta, 1)$, $Y \sim \chi_n^2$ 且 X 和 Y 独立, 则称

$$Z = \frac{X}{\sqrt{Y/n}}$$

的分布是自由度为 n 和非中心参数为 δ 的非中心 t 分布, 记为 $Z \sim t_{n,\delta}$. 特别当 $\delta = 0$ 时的分布称为中心的 t 分布, 即前面所述的 t_n 分布.

非中心 t 分布的密度函数为

$$t_{n,\delta}(x) = \frac{n^{n/2}}{\sqrt{\pi} \Gamma(n/2)} \cdot \frac{e^{-\delta^2/2}}{(n+x^2)^{\frac{n+1}{2}}} \sum_{i=0}^{\infty} \Gamma\left(\frac{n+i+1}{2}\right) \frac{(\delta x)^i}{i!} \left(\frac{2}{n+x^2}\right)^{i/2},$$

$$-\infty < x < \infty. \quad (2.4.6)$$

密度函数 (2.4.6) 的推导方法也是利用求 r.v. 商的密度函数公式, 经过较复杂的计算可求得.

非中心 t 分布的性质如下:

- (1) 若 $Z_n \sim t_{n,\delta}$, 则 $Z_n \xrightarrow{\mathcal{L}} N(\delta, 1)$;
- (2) 若 $Z_n \sim t_{n,\delta}$, 则有

$$E(Z_n) = \delta \left(\frac{n}{2}\right)^{\frac{1}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}, \quad n \geq 2;$$

$$D(Z_n) = \frac{n(1+\delta^2)}{n-2} - \frac{\delta^2 n}{2} \left(\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}\right)^2, \quad n \geq 3.$$

2.4.3 F 分布

定义 2.4.5 设 r.v. $X \sim \chi_m^2$, $Y \sim \chi_n^2$ 且 X 和 Y 独立, 则称

$$F = \frac{X/m}{Y/n}$$

是自由度为 m 和 n (注意分子的自由度在前) 的 F 变量, 其分布称为自由度是 m 和 n 的 F 分布, 记为 $F \sim F_{m,n}$.

F 变量的概率密度函数由下面的定理给出.

定理 2.4.3 若 r.v. $Z \sim F_{m,n}$, 则其密度函数为

$$f_{m,n}(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1} (n+mx)^{-\frac{m+n}{2}}, & x > 0, \\ 0, & \text{其他.} \end{cases} \quad (2.4.7)$$

定理的证明方法如下: ① 分别求出 X/m 和 Y/n 的密度函数; ② 利用求随机变量商的密度函数公式求 F 的密度函数. 定理的具体证明留给读者作为练习.

自由度为 m, n 的 F 分布的密度函数如图 2.4.5 所示. 注意 F 分布的自由度 m 和 n 是有顺序的, 当 $m \neq n$ 时, 若将自由度 m 和 n 的顺序颠倒一下, 得到的是两个不同的 F 分布. 图 2.4.5 给出了几个不同自由度的密度函数曲线. 由图 2.4.5 看到对给定 $m = 10$, n 取不同值时密度曲线 $f_{m,n}(x)$ 的图形是偏态的, n 越小偏态越严重.

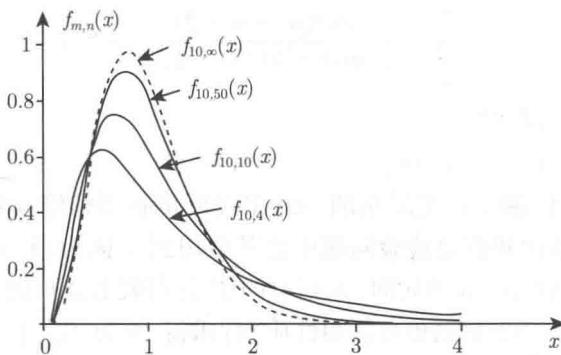
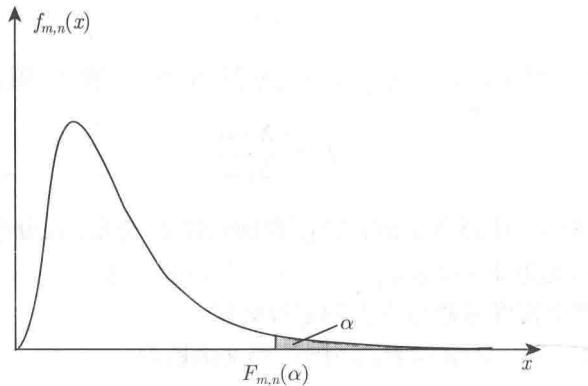


图 2.4.5 $F_{m,n}$ 的密度函数 $f_{m,n}(x)$ 形状图

设 $F \sim F_{m,n}$, $0 < \alpha < 1$, 令 $P(F > c) = \alpha$, 则称 $c = F_{m,n}(\alpha)$ 为 F 分布的上侧 α 分位数, 如图 2.4.6 所示. 当 m, n 和 α 给定时, 可以通过查附表 4 求出 $F_{m,n}(\alpha)$ 之值. 例如, $F_{4,10}(0.05) = 3.48$, $F_{10,15}(0.01) = 3.80$ 等. 这在区间估计和假设检验问题中常常用到.

图 2.4.6 $F_{m,n}$ 的上侧 α 分位数

F 变量具有下列的性质:

- (1) 若 $Z \sim F_{m,n}$, 则 $1/Z \sim F_{n,m}$.
- (2) 若 $Z \sim F_{m,n}$, 则对 $r > 0$ 有

$$E(Z^r) = \left(\frac{n}{m}\right)^r \frac{\Gamma\left(\frac{m}{2} + r\right) \Gamma\left(\frac{n}{2} - r\right)}{\Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{m}{2}\right)}, \quad 2r < n.$$

特别地,

$$\begin{aligned} E(Z) &= \frac{n}{n-2}, \quad n > 2, \\ D(Z) &= \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \quad n > 4. \end{aligned}$$

- (3) 若 $T \sim t_n$, 则 $T^2 \sim F_{1,n}$.
- (4) $F_{m,n}(1-\alpha) = 1/F_{n,m}(\alpha)$.

以上性质中 (1) 和 (3) 是显然的, (2) 和 (4) 的证明留给读者作为练习, 尤其性质 (4) 在区间估计和假设检验问题中会常常用到. 因为当 α 为较小的数, 如 $\alpha = 0.05$ 或 $\alpha = 0.01$, m, n 给定时, 从已有的 F 分布表上查不到 $F_{m,n}(1-0.05)$ 和 $F_{m,n}(1-0.01)$ 之值, 但它们的值可利用性质 (4) 求得, 因为 $F_{n,m}(0.05)$ 和 $F_{n,m}(0.01)$ 是可以通过查 F 分布表求得的.

*非中心 F 分布简介

定义 2.4.6 设 r.v. $X \sim \chi^2_{m,\delta}$, $Y \sim \chi^2_n$ 且 X 和 Y 独立, 则称

$$Z = \frac{X/m}{Y/n}$$

的分布是自由度为 m, n 和非中心参数为 δ 的非中心 F 分布, 记为 $Z \sim F_{m,n;\delta}$. 当 $\delta = 0$ 时, 称 Z 的分布为中心的 F 分布, 即前面定义的 $F_{m,n}$.

若 $Z \sim F_{m,n;\delta}$, 则 Z 的密度函数为

$$f_{m,n;\delta}(x) = \begin{cases} \frac{m^{\frac{n}{2}} n^{\frac{m}{2}}}{\Gamma(\frac{n}{2})} e^{-\frac{\delta^2}{2}} x^{\frac{m}{2}-1} \sum_{k=0}^{\infty} \frac{\left(\frac{\delta^2 m x}{2}\right)^k \Gamma\left(\frac{m+n}{2} + k\right)}{k! \Gamma\left(\frac{m}{2} + k\right) (mx + n)^{\frac{m+n}{2} + k}}, & x > 0, \\ 0, & \text{其他.} \end{cases} \quad (2.4.8)$$

密度函数 (2.4.8) 的推导方法: ① 求出 r.v. X/m 的密度函数; ② 求出 r.v. Y/n 的密度函数; ③ 利用求随机变量之商密度函数的公式, 经过复杂的计算可得.

非中心 F 分布具有下列性质:

- (1) 若 $X \sim t_{n,\delta}$, 则 $X^2 \sim F_{1,n;\delta}$.
- (2) 若 $Z_n \sim F_{m,n;\delta}$, $n = 1, 2, \dots$, δ 固定, 则当 $n \rightarrow \infty$ 时 $Z_n \xrightarrow{\mathcal{L}} \chi_{m,\delta}^2 / m$.
- (3) 若 $Z \sim F_{m,n;\delta}$, 则

$$E(Z) = \frac{n(m+\delta)}{m(n-2)}, \quad n > 2,$$

$$D(Z) = \frac{2n^2}{m^2(n-2)^2(n-4)} [(m+\delta^2)^2 + (n-2)(m+2\delta^2)], \quad n > 4.$$

2.4.4 几个重要推论

下面几个推论在正态总体假设检验和区间估计问题中有着重要应用.

推论 2.4.1 设 X_1, X_2, \dots, X_n 相互独立, $X_i \sim N(a_i, \sigma_i^2)$, $i = 1, 2, \dots, n$, 则

$$\sum_{i=1}^n \left(\frac{X_i - a_i}{\sigma_i} \right)^2 \sim \chi_n^2.$$

特别当 $a_i = a$, $\sigma_i^2 = \sigma^2$, $i = 1, \dots, n$ 时 $\sum_{i=1}^n (X_i - a)^2 / \sigma^2 \sim \chi_n^2$.

证 因为 $Y_i = (X_i - a_i) / \sigma_i \sim N(0, 1)$, $i = 1, 2, \dots, n$ 且 Y_1, \dots, Y_n 相互独立, 按定义 $\sum_{i=1}^n (X_i - a_i)^2 / \sigma_i^2 = \sum_{i=1}^n Y_i^2 \sim \chi_n^2$.

推论 2.4.2 设 X_1, X_2, \dots, X_n i.i.d. $\sim N(a, \sigma^2)$, 则

$$T = \frac{\sqrt{n}(\bar{X} - a)}{S} \sim t_{n-1}.$$

证 由定理 2.2.3 可知 $\bar{X} \sim N(a, \sigma^2/n)$, 将其标准化得 $\sqrt{n}(\bar{X} - a) / \sigma \sim N(0, 1)$. 又 $(n-1)S^2 / \sigma^2 \sim \chi_{n-1}^2$, 即 $S^2 / \sigma^2 \sim \chi_{n-1}^2 / (n-1)$ 且 \bar{X} 和 S^2 独立, 按定义有

$$T = \frac{\sqrt{n}(\bar{X} - a) / \sigma}{\sqrt{S^2 / \sigma^2}} = \frac{\sqrt{n}(\bar{X} - a)}{S} \sim t_{n-1}.$$

推论 2.4.3 设 X_1, X_2, \dots, X_m i.i.d. $\sim N(a_1, \sigma^2)$, Y_1, Y_2, \dots, Y_n i.i.d. $\sim N(a_2, \sigma^2)$, 且样本 X_1, X_2, \dots, X_m 与 Y_1, Y_2, \dots, Y_n 独立, 则

$$T = \frac{(\bar{X} - \bar{Y}) - (a_1 - a_2)}{S_w} \cdot \sqrt{\frac{mn}{n+m}} \sim t_{n+m-2},$$

其中 $(n+m-2)S_w^2 = (m-1)S_1^2 + (n-1)S_2^2$, 此处

$$S_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2.$$

证 由定理 2.2.3 可知 $\bar{X} \sim N(a_1, \sigma^2/m)$, $\bar{Y} \sim N(a_2, \sigma^2/n)$, 故有 $\bar{X} - \bar{Y} \sim N(a_1 - a_2, (n+m)\sigma^2/mn)$. 将其标准化得

$$\frac{(\bar{X} - \bar{Y}) - (a_1 - a_2)}{\sigma} \sqrt{\frac{mn}{m+n}} \sim N(0, 1). \quad (2.4.9)$$

又 $(m-1)S_1^2/\sigma^2 \sim \chi_{m-1}^2$, $(n-1)S_2^2/\sigma^2 \sim \chi_{n-1}^2$, 且二者独立, 再利用 χ^2 分布的性质可知

$$\frac{(m-1)S_1^2 + (n-1)S_2^2}{\sigma^2} \sim \chi_{n+m-2}^2. \quad (2.4.10)$$

再由式 (2.4.9) 和式 (2.4.10) 中 (\bar{X}, \bar{Y}) 与 (S_1^2, S_2^2) 相互独立, 由 t 变量的定义可知

$$\begin{aligned} T &= \frac{(\bar{X} - \bar{Y}) - (a_1 - a_2)}{\sigma} \sqrt{\frac{mn}{n+m}} \Bigg/ \sqrt{\frac{(m-1)S_1^2 + (n-1)S_2^2}{\sigma^2(n+m-2)}} \\ &= \frac{(\bar{X} - \bar{Y}) - (a_1 - a_2)}{S_w} \sqrt{\frac{nm}{n+m}} \sim t_{n+m-2}. \end{aligned}$$

推论 2.4.4 设 X_1, X_2, \dots, X_m i.i.d. $\sim N(a_1, \sigma_1^2)$, Y_1, Y_2, \dots, Y_n i.i.d. $\sim N(a_2, \sigma_2^2)$, 且样本 X_1, X_2, \dots, X_m 和 Y_1, Y_2, \dots, Y_n 独立, 则

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{m-1, n-1},$$

此处 S_1^2 和 S_2^2 定义如推论 2.4.3 所述.

证 由定理 2.2.3 可知 $(m-1)S_1^2/\sigma_1^2 \sim \chi_{m-1}^2$, $(n-1)S_2^2/\sigma_2^2 \sim \chi_{n-1}^2$, 且二者独立, 由 F 分布的定义可知

$$F = \frac{\frac{(m-1)S_1^2}{\sigma_1^2} / (m-1)}{\frac{(n-1)S_2^2}{\sigma_2^2} / (n-1)} = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{m-1, n-1}.$$

下面这一推论给出了服从指数分布随机变量的线性函数的分布与 χ^2 分布的关系. 这在指数分布总体参数的区间估计和假设检验问题中有重要应用.

推论 2.4.5 设 X_1, X_2, \dots, X_n i.i.d. 服从指数分布 $f(x, \lambda) = \lambda e^{-\lambda x} I_{(0, \infty)}(x)$, 则有

$$2\lambda n \bar{X} = 2\lambda \sum_{i=1}^n X_i \sim \chi_{2n}^2.$$

证 首先证明 $2\lambda X_1 \sim \chi_2^2$. 因为

$$F(y) = P(2\lambda X_1 < y) = P\left(X_1 < \frac{y}{2\lambda}\right) = \int_0^{\frac{y}{2\lambda}} \lambda e^{-\lambda x} dx,$$

所以

$$f(y) = F'(y) = \begin{cases} \frac{1}{2} e^{-\frac{y}{2}}, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

因此 $f(y)$ 即为自由度为 2 的 χ^2 密度, 即 $2\lambda X_1 \sim \chi_2^2$.

再利用 χ^2 分布的性质 (3), $2\lambda X_i \sim \chi_2^2$, $i = 1, 2, \dots, n$; 又它们相互独立, 故有
 $2\lambda \sum_{i=1}^n X_i \sim \chi_{2n}^2$.

2.5 统计量的极限分布

2.5.1 定义

本章引言中已指出, 在许多情形下统计量的精确分布很难求出, 因此要研究统计量的极限分布. 首先给出下列定义.

定义 2.5.1 当样本容量 n 趋向无穷时, 统计量的分布趋于一确定分布, 则后者的分布称为统计量的极限分布, 也常称为大样本分布.

当样本容量 n 充分大时, 极限分布可作为统计量的近似分布.

研究统计量的极限分布有下列意义:

(1) 为了获得统计推断方法的优良性, 常常要知道统计量的分布. 但统计量的精确分布一般很难求得, 建立统计量的极限分布, 提供了一种近似方法, 总比什么方法没有要好;

(2) 有时统计量的精确分布虽可求出, 但表达式过于复杂, 使用不方便. 若极限分布较简单, 宁可使用极限分布;

(3) 有些统计推断方法的优良性本身就是研究其极限性质, 如相合性、渐近正态性等.

定义 2.5.2 当样本容量 $n \rightarrow \infty$ 时, 一个统计量或统计推断方法的性质称为大样本性质 (large sample properties). 当样本大小固定时, 统计量或统计推断方法的性质称为小样本性质 (small sample properties).

在此要强调的是, 大样本性质和小样本性质 的差别不在于样本个数的多少, 而是在于所讨论的问题是在样本容量 $n \rightarrow \infty$ 时去考虑, 还是在样本容量 n 固定时去研究. 关于大样本性质的研究构成了数理统计的一个很重要的部分, 称为统计大样本理论. 统计大样本理论, 近几十年来发展很快, 成为第二次世界大战后数理统计发展的重要特点之一. 有些统计分支中, 如非参数统计, 大样本理论占据了主导地位.

2.5.2 几个例子

例 2.5.1 设 X_1, \dots, X_n i.i.d. $\sim F$, 其中总体 F 有均值 a_F 和方差 σ_F^2 . 设 $0 < \sigma_F^2 < \infty$, $\bar{X}_n = \sum_{i=1}^n X_i/n$ 为样本均值, 试讨论 \bar{X}_n 的大样本性质和小样本性质.

解 (1) 由柯尔莫哥洛夫 (Kolmogorov) 强大数定律有

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = a_F\right) = 1 \quad \text{等价地表示为} \quad \bar{X}_n \xrightarrow{\text{a.s.}} a_F \quad \text{当 } n \rightarrow \infty. \quad (2.5.1)$$

式 (2.5.1) 表示当样本容量 $n \rightarrow \infty$ 时, 估计量 \bar{X}_n 以概率 1 任意地接近估计值 a_F , 这个性质称为 \bar{X}_n 的强相合性, 它是一个大样本性质. 因为只有在 $n \rightarrow \infty$ 时这个性质才有意义.

(2) 按 Lindeberg 中心极限定理当 $n \rightarrow \infty$ 时有

$$\sqrt{n}(\bar{X}_n - a_F)/\sigma_F \xrightarrow{\mathcal{L}} N(0, 1). \quad (2.5.2)$$

其中 $\xrightarrow{\mathcal{L}}$ 表示依分布收敛. 式 (2.5.2) 刻画了 \bar{X}_n 的另一个大样本性质——渐近正态性. \bar{X}_n 作为总体均值的估计与 a_F 的偏差超过 c 的概率 $P(|\bar{X}_n - a_F| > c)$, 可以作为衡量这一估计量优良性的一项指标. 若 F 为正态分布 $N(a_F, \sigma_F^2)$, 则这一概率可以精确地算出 (当 σ_F^2 已知时, 可通过查标准正态分布表, 即附表 1 求出其值), 若 F 的分布类型根本不知道, 这概率无法计算. 但有了式 (2.5.2) 后, 至少在样本容量 n 较大时, 可用正态分布求得这一概率的近似值.

(3) 另一方面, 关于 \bar{X} 的小样本性质有 $E(\bar{X}_n) = a_F$, 即估计量 \bar{X}_n 的期望值等于被估计的未知参数 a_F . 这个性质称为 \bar{X}_n 的无偏性 (即 \bar{X}_n 为 a_F 的无偏估计). 这是一个小样本性质. 因为这个性质的意义是在样本大小 n 固定时去理解的.

例 2.5.2 设 $X \sim b(1, p)$, 令 X_1, \dots, X_n 为自总体 X 中抽取的简单样本. 试证 $\sqrt{n}(\bar{X} - p)/\sqrt{\bar{X}(1 - \bar{X})}$ 的极限分布为 $N(0, 1)$.

解 为了证明上述结果, 需要下列引理:

引理 2.5.1(Slutsky 引理) 令 $\{X_n\}$ 和 $\{Y_n\}$ 是两个随机变量的序列, 满足当 $n \rightarrow \infty$ 时 $X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{P} c$, $-\infty < c < \infty$ 为常数, 则有① $X_n \pm Y_n \xrightarrow{\mathcal{L}} X \pm c$, ② $X_n Y_n \xrightarrow{\mathcal{L}} cX$, ③ $X_n / Y_n \xrightarrow{\mathcal{L}} X/c$ ($c \neq 0$).

证 见文献 [21] 1.5.4 节.

利用上述引理证明所求问题如下: 令 $\hat{p} = \bar{X} = \sum_{i=1}^n X_i/n$, 由中心极限定理和大数定律可知, 当 $n \rightarrow \infty$ 时有

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \xrightarrow{\mathcal{L}} N(0, 1), \quad \frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{P} 1.$$

故由 Slutsky 引理可知, 当 $n \rightarrow \infty$ 时有

$$\frac{\sqrt{n}(\bar{X} - p)}{\sqrt{\bar{X}(1-\bar{X})}} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \cdot \frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{\mathcal{L}} N(0, 1).$$

问题得证.

***例 2.5.3** 设总体 X 有密度 f , X_1, \dots, X_n i.i.d. $\sim X$. 令总体 X 的中位数为 $\xi_{1/2}$, 即 $\xi_{1/2}$ 满足 $\int_{-\infty}^{\xi_{1/2}} f(x)dx = 1/2$, 且假定 f 在 $\xi_{1/2}$ 点连续非 0 (由此可知 $\xi_{1/2}$ 为 f 唯一的中位数). 设 $m_{1/2}$ 为样本中位数, 则当 $n \rightarrow \infty$ 时有

$$2\sqrt{n}f(\xi_{1/2})(m_{1/2} - \xi_{1/2}) \xrightarrow{\mathcal{L}} N(0, 1), \quad (2.5.3)$$

此处样本中位数 $m_{1/2}$ 的定义由式 (1.3.1) 给出.

证 此例是文献 [1] 第 538 页定理 6.11 的特例, 有兴趣者可参看其证明. 此例表明当 n 充分大时, 样本中位数 $m_{1/2}$ 的分布可用 $N(\xi_{1/2}, 1/[4nf^2(\xi_{1/2})])$ 来近似.

* 2.6 指 数 族

在统计理论问题中, 许多统计推断方法的优良性, 对一类范围广泛的统计模型 (亦称为分布族) 有较满意的结果. 这类分布族称为指数型分布族 (简称指数族). 常见的分布, 如正态分布、二项分布、Poisson 分布、负二项分布、指数分布和 Gamma 分布等都属于这类分布族, 这些表面上看来各不相同的分布族, 其实都可以统一在一种包罗更广的一类称为指数族的模式中. 当然引进这种分布族的理由, 主要不在于谋求形式上的统一, 而在于这种统一抓住了它们的共性, 因此许多统计理论问题, 对指数族获得较彻底的解决. 本节介绍指数族的定义及简单性质.

2.6.1 定义与例子

定义 2.6.1 设 $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$ 是定义在样本空间 \mathcal{X} 上的分布族, 其中 Θ 为参数空间. 若其概率函数 $f(x, \theta)$ 可表示成如下形式:

$$f(x, \theta) = C(\theta) \exp \left\{ \sum_{i=1}^k Q_i(\theta) T_i(x) \right\} h(x), \quad (2.6.1)$$

则称此分布族为指数型分布族 (简称指数族, exponential family), 其中 k 为正整数, $C(\theta) > 0$ 和 $Q_i(\theta)$ ($i = 1, 2, \dots, k$) 都是定义在参数空间 Θ 上的函数, $h(x) > 0$ 和 $T_i(x)$ ($i = 1, 2, \dots, k$) 都是定义在样本空间 \mathcal{X} 上的函数.

指数族的一个重要性质是族中的所有分布具有共同的支撑集 ($G(x)$ 称为概率函数 $f(x, \theta)$ 的支撑集, 若 $G(x) = \{x : f(x, \theta) > 0\}$). 由定义可见指数族的支撑集 $\{x : f(x, \theta) > 0\} = \{x : h(x) > 0\}$ 与 θ 无关. 任一分布族若其支撑集与 θ 有关, 则族中分布不再具有共同支撑集, 因而必不是指数族.

例 2.6.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从正态分布 $N(\mu, \sigma^2)$ 中抽取的简单样本, 则样本分布族是指数族.

解 样本 \mathbf{X} 的联合密度为

$$f(\mathbf{x}; \mu, \sigma^2) = (\sqrt{2\pi}\sigma)^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\}. \quad (2.6.2)$$

记 $\theta = (\mu, \sigma^2)$, 则参数空间为 $\Theta = \{\theta = (\mu, \sigma^2) : -\infty < \mu < +\infty, \sigma^2 > 0\}$. 将式 (2.6.2) 改写为

$$\begin{aligned} f(\mathbf{x}, \theta) &= (\sqrt{2\pi}\sigma)^{-n} e^{-\frac{n\mu^2}{2\sigma^2}} \exp \left\{ \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right\} \\ &= C(\theta) \exp\{Q_1(\theta)T_1(\mathbf{x}) + Q_2(\theta)T_2(\mathbf{x})\} h(\mathbf{x}), \end{aligned} \quad (2.6.3)$$

其中 $C(\theta) = (\sqrt{2\pi}\sigma)^{-n} \exp\{-n\mu^2/(2\sigma^2)\}$, $Q_1(\theta) = \mu/\sigma^2$, $Q_2(\theta) = -1/(2\sigma^2)$, $T_1(\mathbf{x}) = \sum_{i=1}^n x_i$, $T_2(\mathbf{x}) = \sum_{i=1}^n x_i^2$, $h(\mathbf{x}) \equiv 1$. 因此, 由定义可知上述样本分布族是指数族.

特别当取样本容量 $n = 1$ 时, X_1 的密度 (即正态密度) 函数为

$$\begin{aligned} f(x, \theta) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\mu^2}{2\sigma^2}} \exp \left\{ \frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2 \right\} \\ &= c(\theta) \exp\{Q_1(\theta)T_1(x) + Q_2(\theta)T_2(x)\} h(x), \end{aligned}$$

其中 $\theta = (\mu, \sigma^2)$, $C(\theta) = (\sqrt{2\pi}\sigma)^{-1} e^{-\frac{\mu^2}{2\sigma^2}}$, $T_1(x) = x$, $T_2(x) = x^2$, $h(x) = 1$, 而 $Q_1(\theta)$, $Q_2(\theta)$ 与式 (2.6.3) 相同, 按定义, 正态分布族 $\{N(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$ 是指数族. 这说明样本分布族是否为指数族, 不依赖于样本大小 n .

例 2.6.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 Gamma 分布 $\Gamma(\gamma, \lambda)$ 中抽取的简单样本, 则样本分布族是指数族.

解 样本 \mathbf{X} 的联合密度为

$$f(\mathbf{x}; \gamma, \lambda) = \prod_{i=1}^n \left[\frac{\lambda^\gamma}{\Gamma(\gamma)} x_i^{\gamma-1} \exp\{-\lambda x_i\} I_{(0, \infty)}(x_i) \right]^n. \quad (2.6.4)$$

记 $\theta = (\gamma, \lambda)$, 则参数空间为 $\Theta = \{\theta = (\gamma, \lambda) : \gamma > 0, \lambda > 0\}$. 将上式改写为

$$\begin{aligned} f(\mathbf{x}, \theta) &= \frac{\lambda^{n\gamma}}{(\Gamma(\gamma))^n} \exp \left\{ -\lambda \sum_{i=1}^n x_i + (\gamma - 1) \sum_{i=1}^n \log x_i \right\} \prod_{i=1}^n I_{(0, \infty)}(x_i) \\ &= C(\theta) \exp\{Q_1(\theta)T_1(\mathbf{x}) + Q_2(\theta)T_2(\mathbf{x})\} h(\mathbf{x}), \end{aligned} \quad (2.6.5)$$

其中 $C(\theta) = \lambda^{n\gamma}/(\Gamma(\gamma))^n$, $Q_1(\theta) = -\lambda$, $Q_2(\theta) = \gamma - 1$, $T_1(\mathbf{x}) = \sum_{i=1}^n x_i$, $T_2(\mathbf{x}) = \sum_{i=1}^n \log x_i$, $h(\mathbf{x}) = \prod_{i=1}^n I_{(0, \infty)}(x_i)$. 按定义可知上述样本分布族是指数族.

与例 2.6.1 末尾的说明相似, 当取样本容量 $n = 1$ 时, 将 X_1 的密度函数表成指数族的形式, 易见 Gamma 分布族 $\{\Gamma(\gamma, \lambda) : \gamma > 0, \lambda > 0\}$ 仍是指数族.

例 2.6.3 二项分布族 $\{b(n, \theta) : 0 < \theta < 1\}$ 是指数族.

解 设 $X \sim$ 二项分布 $b(n, \theta)$, 其概率函数为

$$\begin{aligned} p(x, \theta) &= P_\theta(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \\ &= \binom{n}{x} \left(\frac{\theta}{1 - \theta} \right)^x (1 - \theta)^n, \quad x = 0, 1, 2, \dots, n, \end{aligned} \quad (2.6.6)$$

其中样本空间 $\mathcal{X} = \{0, 1, 2, \dots, n\}$, 参数空间 $\Theta = \{\theta : 0 < \theta < 1\} = (0, 1)$. 将上式改写为

$$\begin{aligned} p(x, \theta) &= (1 - \theta)^n \exp \left\{ x \log \frac{\theta}{1 - \theta} \right\} \cdot \binom{n}{x} \\ &= C(\theta) \exp\{Q_1(\theta)T_1(x)\} h(x), \end{aligned} \quad (2.6.7)$$

其中 $C(\theta) = (1 - \theta)^n$, $Q_1(\theta) = \log[\theta/(1 - \theta)]$, $T_1(x) = x$, $h(x) = \binom{n}{x}$, 按定义二项分布族 $\{b(n, \theta) : 0 < \theta < 1\}$ 也是指数族.

例 2.6.4 Poisson 分布族 $\{P(\theta) : \theta > 0\}$ 是指数族.

解 设 $X \sim$ Poisson 分布 $P(\theta)$, 其概率函数为

$$\begin{aligned} p(x, \theta) &= P_\theta(X = x) = \frac{e^{-\theta} \theta^x}{x!} = \frac{e^{-\theta} \exp\{x \log \theta\}}{x!} \\ &= C(\theta) \exp\{Q_1(\theta)T_1(x)\} h(x), \quad x = 0, 1, 2, \dots, \end{aligned} \quad (2.6.8)$$

其中样本空间 $\mathcal{X} = \{0, 1, 2, \dots\}$, 参数空间 $\Theta = \{\theta : \theta > 0\} = (0, \infty)$, $C(\theta) = e^{-\theta}$, $Q_1(\theta) = \log \theta$, $T_1(x) = x$, $h(x) = 1/x!$, 按定义 Poisson 分布族是指数族.

例 2.6.5 均匀分布族 $\{U(0, \theta), \theta > 0\}$ 和双参数指数分布不是指数族.

解 由指数族的定义可知, 其支撑集 $\{x : f(x, \theta) > 0\} = \{x : h(x) > 0\}$ 与 θ 无关. 而均匀分布族 $\{U(0, \theta), \theta > 0\}$ 的支撑集为 $\{x : f(x, \theta) > 0\} = (0, \theta)$ 与 θ 有关, 因此它不是指数族.

双参数指数分布族的密度函数如下:

$$p(x; \mu, \sigma) = \frac{1}{\sigma} \exp \left\{ -\frac{x-\mu}{\sigma} \right\} I_{[x>\mu]}, \quad -\infty < \mu < +\infty, \quad \sigma > 0, \quad (2.6.9)$$

其中 μ 和 σ 为两个参数, 它的支撑集为 $\{x : p(x; \mu, \sigma) > 0\} = (\mu, \infty)$ 与未知参数 μ 有关, 故它也不是指数族. 但若 μ 已知, 如 $\mu = 0$, 则单参数指数分布族 $\text{Exp}(1/\sigma)$ 属于指数族.

2.6.2 指数族的自然形式及自然参数空间

在指数族的定义 $C(\theta) \exp \left\{ \sum_{i=1}^k Q_i(\theta) T_i(x) \right\} h(x)$ 中, 若用 φ_i 代替 $Q_i(\theta)$, 而将 $C(\theta)$ 表成 φ 的函数 $C^*(\varphi)$, $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_k)$, 则指数族的表达式变为 $C^*(\varphi) \exp \left\{ \sum_{i=1}^k \varphi_i T_i(x) \right\} h(x)$. 再改 φ_i 为 θ_i , $i = 1, 2, \dots, k$, 得到如下指数族的自然形式 (或称为标准形式) 的定义.

定义 2.6.2 如果指数族有下列形式:

$$f(x, \theta) = C^*(\theta) \exp \left\{ \sum_{i=1}^k \theta_i T_i(x) \right\} h(x), \quad (2.6.10)$$

则称它为指数族的自然形式 (natural form). 此时集合

$$\Theta^* = \left\{ (\theta_1, \theta_2, \dots, \theta_k) : \int_{\mathcal{X}} \exp \left\{ \sum_{i=1}^k \theta_i T_i(x) \right\} h(x) dx < \infty \right\} \quad (2.6.11)$$

称为自然参数空间 (natural parametric space).

例 2.6.6 将例 2.6.1 中样本分布族表示为指数族的自然形式, 并求出其自然参数空间.

解 令 $\theta = (\mu, \sigma^2)$. 由例 2.6.1 可知

$$\begin{aligned} f(x; \mu, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{n\mu^2}{2\sigma^2}} \exp \left\{ \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right\} \\ &= C(\theta) \exp \{Q_1(\theta)T_1(x) + Q_2(\theta)T_2(x)\} h(x), \end{aligned}$$

其中 $C(\theta) = (2\pi\sigma^2)^{-n/2} \exp\{-n\mu^2/(2\sigma^2)\}$, 而 $Q_1(\theta) = \mu/\sigma^2$, $Q_2(\theta) = -1/(2\sigma^2)$ 以及 $T_1(\mathbf{x}) = \sum_{i=1}^n x_i$, $T_2(\mathbf{x}) = \sum_{i=1}^n x_i^2$, $h(\mathbf{x}) \equiv 1$ 皆与式 (2.6.3) 中相同.

令 $\varphi_1 = Q_1(\theta)$, $\varphi_2 = Q_2(\theta)$, 解出 $\sigma^2 = -1/(2\varphi_2)$, $\mu^2/\sigma^2 = -\varphi_1^2/(2\varphi_2)$. 因此 $C(\theta)$ 变为 $C^*(\varphi) = (-\varphi_2/\pi)^{n/2} \exp\{n\varphi_1^2/(4\varphi_2)\}$, $\varphi = (\varphi_1, \varphi_2)$, 故式 (2.6.3) 变为下面的自然形式:

$$f(\mathbf{x}, \varphi) = C^*(\varphi) \exp\{\varphi_1 T_1(\mathbf{x}) + \varphi_2 T_2(\mathbf{x})\} h(\mathbf{x}), \quad (2.6.12)$$

其自然参数空间为

$$\Theta^* = \{\varphi = (\varphi_1, \varphi_2) : -\infty < \varphi_1 < +\infty, -\infty < \varphi_2 < 0\}. \quad (2.6.13)$$

例 2.6.7 将例 2.6.3 中二项分布族表成指数族的自然形式, 并求出其自然参数空间.

解 由例 2.6.3 可知

$$f(x, \theta) = (1 - \theta)^n \exp\left\{\log \frac{\theta}{1 - \theta} \cdot x\right\} \binom{n}{x}, \quad x = 0, 1, 2, \dots, n,$$

参数空间 $\Theta = \{\theta : 0 < \theta < 1\}$. 令 $\log[\theta/(1 - \theta)] = \varphi$, 故知 $-\infty < \varphi < +\infty$. 由 $\varphi = \log[\theta/(1 - \theta)]$ 解出 $1 - \theta = 1/(1 + e^\varphi)$. 故由式 (2.6.7) 获得其自然形式

$$f(x, \varphi) = (1 + e^\varphi)^{-n} \exp\{\varphi \cdot x\} \binom{n}{x} = C^*(\varphi) \exp\{\varphi T_1(x)\} h(x), \quad (2.6.14)$$

其中 $C^*(\varphi) = (1 + e^\varphi)^{-n}$, 而 $T_1(x)$, $h(x)$ 与式 (2.6.7) 中相同, 其自然参数空间为

$$\Theta^* = \{\varphi : -\infty < \varphi < +\infty\} = (-\infty, +\infty). \quad (2.6.15)$$

2.6.3 指数族的性质

自然参数空间具有下述重要性质.

定理 2.6.1 在指数族的自然形式下, 自然参数空间为凸集.

证明的方法如下: 设 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ 为 k 维向量. 任给 $\boldsymbol{\theta}^{(1)} = (\theta_1^1, \dots, \theta_k^1)$, $\boldsymbol{\theta}^{(0)} = (\theta_1^0, \dots, \theta_k^0)$ 皆属于自然参数空间 Θ^* , 设 $0 < \alpha < 1$, 令 $\boldsymbol{\theta} = \alpha \boldsymbol{\theta}^{(1)} + (1 - \alpha) \boldsymbol{\theta}^{(0)}$ (即 $\theta_i = \alpha \theta_i^1 + (1 - \alpha) \theta_i^0$, $i = 1, 2, \dots, k$), 若能证明 $\boldsymbol{\theta} \in \Theta^*$, 则按凸集的定义, 定理 2.6.1 得证. 具体证明留作练习.

指数族具有良好的分析性质, 见下述定理.

定理 2.6.2 设指数族的自然形式中, 自然参数空间有内点, 其内点集为 Θ_0 . 设 $g(x)$ 为任一实函数, 使得积分

$$G(\boldsymbol{\theta}) = \int_{\mathcal{X}} g(x) \exp\left\{\sum_{j=1}^k \theta_j T_j(x)\right\} h(x) dx$$

在 Θ_0 内存在有限, 则 $G(\theta)$ 的任意阶偏导数在 Θ_0 内存在且可在积分号下求得, 即

$$\frac{\partial^m G(\theta)}{\partial \theta_1^{m_1} \cdots \partial \theta_k^{m_k}} = \int_{\mathcal{X}} \frac{\partial^m}{\partial \theta_1^{m_1} \cdots \partial \theta_k^{m_k}} \left[g(x) \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) \right\} h(x) \right] dx,$$

其中 $\sum_{j=1}^k m_j = m$, 即对 $G(\theta)$ 关于 θ 的任意阶偏导数可在积分下求得.

此定理一般的形式叙述及其证明查看文献 [1] 第 21 页定理 1.2.1.

2.7 充分统计量

2.7.1 引言与定义

统计量的充分性是数理统计的一个基本概念, 它是由 R.A. Fisher 在其 1922 年的奠基性工作中提出来的.

我们知道, 统计量是对样本的简化, 希望达到: ① 简化的程度高; ② 信息的损失少. 一个统计量能集中样本中信息的多少, 与统计量的具体形式有关, 也依赖于问题的统计模型. 最好的情况是统计量把样本中的全部信息都集中起来, 也就是说信息无损失, 称这样的统计量为充分统计量. 请看下例.

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 0-1 分布中抽取的简单样本, 即 $P(X_i = 1) = \theta$, $P(X_i = 0) = 1 - \theta$, $0 < \theta < 1$. 记 $T(\mathbf{X}) = \sum_{i=1}^n X_i$, 如果目的仅仅为了推断 θ , 则 $T(\mathbf{X})$ 是充分统计量. 从直观上看, 知道了 $T(\mathbf{X})$ 推断 θ 的效果, 与知道样本 (X_1, \dots, X_n) 一样, 因为 $T(\mathbf{X})$ 表示事件 A 在 n 次独立的 Bernoulli 试验中成功的次数. 而 (X_1, \dots, X_n) 比 $T(\mathbf{X})$ 更详细的地方在于指出了事件 A 是在哪几次试验中发生的. 由于各次试验是在同样条件下独立进行的, 所以 A 在哪几次试验中发生是随机的, 因此知道了 A 出现的总次数后, 再知道上述细节对推断 θ 不会有更多的用处. 故从直观上, 知道了 $T(\mathbf{X})$ 后与知道样本 (X_1, \dots, X_n) 对推断 θ 的效果相同.

一大堆原始资料, 加工成统计量 $T(\mathbf{X})$ 后, 一般来说在信息上会有损失. 但也有可能, 将样本 \mathbf{X} 加工成 $T(\mathbf{X})$ 时抓住了问题的实质, 即 $T(\mathbf{X})$ 中保留了样本 \mathbf{X} 中所含参数 θ 的全部信息, 所丢掉的只是无关紧要的东西. 如果一个统计量满足这个要求, 即使忘掉了样本 \mathbf{X} 也能恢复参数 θ 的信息, 则称此统计量为充分的. 仍看上例, 样本 \mathbf{X} 的分布为

$$P_\theta(\mathbf{X} = \mathbf{x}) = P_\theta(X_1 = x_1, \dots, X_n = x_n) = \theta^t(1 - \theta)^{n-t}, \quad (2.7.1)$$

其中 $t = \sum_{i=1}^n x_i$. 而 $T(\mathbf{X}) = \sum_{i=1}^n X_i \sim \text{二项分布 } b(n, \theta)$, 故有

$$P_\theta(T(\mathbf{X}) = t) = \binom{n}{t} \theta^t (1-\theta)^{n-t}, \quad t = 0, 1, 2, \dots, n, \quad 0 < \theta < 1. \quad (2.7.2)$$

可见基于式 (2.7.2) 推断参数 θ 和基于式 (2.7.1) 推断参数 θ 的效果是相同的. 因此 $T(\mathbf{X})$ 为充分统计量.

关于样本 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 的信息可以设想成如下的公式:

$$\begin{aligned} & \{\text{样本 } \mathbf{X} \text{ 中的信息}\} \\ &= \{T(\mathbf{X}) \text{ 中所含样本的信息}\} \\ &+ \{\text{在知道 } T(\mathbf{X}) \text{ 后样本 } \mathbf{X} \text{ 尚含有的剩余信息}\}. \end{aligned}$$

故 $T(\mathbf{X})$ 为充分统计量的要求归结为: 要求后一项信息为 0. 用统计的语言来描述, 即要求 $P_\theta(\mathbf{X} \in A | T = t)$ 与 θ 无关, 其中 A 为任一事件. 因此得到如下的定义.

定义 2.7.1 设样本 \mathbf{X} 的分布族为 $\{f(\theta, \mathbf{x}), \theta \in \Theta\}$, Θ 是参数空间. 令 $T = T(\mathbf{X})$ 为一统计量, 若在已知 T 的条件下, 样本 \mathbf{X} 的条件分布与参数 θ 无关, 则称 $T(\mathbf{X})$ 为 θ 的充分统计量 (sufficient statistic).

实际应用时条件分布用条件概率 (离散情形) 或条件密度 (连续情形) 来代替.

例 2.7.1 设 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 为从 0-1 分布中抽取的简单样本, 则 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为充分统计量.

证 记 $T = T(\mathbf{X})$, 按定义只要证明下列条件概率与 θ 无关. 当 $\sum_{i=1}^n x_i = t_0$ 时有

$$\begin{aligned} & P(X_1 = x_1, \dots, X_n = x_n | T = t_0) \\ &= \frac{P(X_1 = x_1, \dots, X_n = x_n, T = t_0)}{P(T = t_0)} \\ &= \frac{P\left(X_1 = x_1, \dots, X_n = t_0 - \sum_{i=1}^{n-1} x_i\right)}{P(T = t_0)} = \frac{\theta^{t_0} (1-\theta)^{n-t_0}}{\binom{n}{t_0} \theta^{t_0} (1-\theta)^{n-t_0}} = \frac{1}{\binom{n}{t_0}}, \end{aligned}$$

因此有

$$P(X_1 = x_1, \dots, X_n = x_n | T = t_0) = \begin{cases} \frac{1}{\binom{n}{t_0}}, & \sum_{i=1}^n x_i = t_0, \\ 0, & \sum_{i=1}^n x_i \neq t_0. \end{cases}$$

上述条件概率与 θ 无关, 因此 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为 θ 的充分统计量.

例 2.7.2 设 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 为从正态总体 $N(\theta, 1)$ 中抽取的简单样本, 则 $T(\mathbf{X}) = \sum_{i=1}^n X_i/n = \bar{X}$ 为 θ 的充分统计量.

证 如果直接取计算在给定 T 时 \mathbf{X} 的条件概率密度, 则计算比较复杂. 采用如下办法: 作正交变换

$$(Y_1, \dots, Y_n)' = \mathbf{A}(X_1, \dots, X_n)', \quad (2.7.3)$$

其中正交阵

$$\mathbf{A} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

由代数学中的 Schmidt 正交化方法可知上述正交阵 \mathbf{A} 是存在的. 变换的 Jacobi 行列式的绝对值为 $|J| = 1$. 由正交变换 (2.7.3) 可知

$$\begin{cases} Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i = \sqrt{n}\bar{X}, \\ Y_j = \sum_{k=1}^n a_{jk} X_k, \quad j = 2, \dots, n. \end{cases}$$

由定理 2.2.3 的证明过程可知 $\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n X_i^2$, 且 Y_1, Y_2, \dots, Y_n 是相互独立的, $Y_1 \sim N(\sqrt{n}\theta, 1)$, $Y_i \sim N(0, 1)$, $i = 2, \dots, n$.

显然, \bar{X} 对原样本 \mathbf{X} 的充分性等价于 Y_1 对 (Y_1, Y_2, \dots, Y_n) 的充分性. 因此只要证明给定 $Y_1 = y_1$ 时, (Y_1, Y_2, \dots, Y_n) 的条件密度与 θ 无关即可. 下面来证明这一事实.

易见 Y_1, \dots, Y_n 的联合密度为

$$f(y_1, \dots, y_n) = (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=2}^n y_i^2 - \frac{1}{2}(y_1 - \sqrt{n}\theta)^2 \right\}.$$

Y_1 的边缘密度函数为

$$f_{Y_1}(y_1) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y_1 - \sqrt{n}\theta)^2 \right\},$$

给定 $Y_1 = y_1$ 时, (Y_1, \dots, Y_n) 的条件密度是

$$f(y_1, \dots, y_n | y_1) = \frac{f(y_1, \dots, y_n)}{f_{Y_1}(y_1)} = (2\pi)^{-\frac{n-1}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=2}^n y_i^2 \right\}, \quad (2.7.4)$$

它与 θ 无关. 所以 $T(\mathbf{X}) = \bar{X}$ 是 θ 的充分统计量.

例 2.7.3 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从指数分布 $\text{Exp}(\theta)$ 中抽取的简单样本, 其密度函数为 $f(x, \theta) = \theta e^{-\theta x} I_{[x>0]}$, 则 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为 θ 的充分统计量.

证 \mathbf{X} 的联合密度为

$$f(\mathbf{x}, \theta) = \theta^n \exp \left\{ -\theta \sum_{i=1}^n x_i \right\} I_{[x_i > 0, i=1,2,\dots,n]},$$

作变换

$$\begin{cases} Y_1 = X_1, \\ \dots \\ Y_{n-1} = X_{n-1}, \\ Y_n = X_1 + X_2 + \dots + X_n = T. \end{cases}$$

变换的 Jacobi 行列式的绝对值为 $|J| = 1$, 这是一个一一对应的变换.

显然, $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 对原样本 \mathbf{X} 的充分性等价于 $Y_n = T$ 对 (Y_1, Y_2, \dots, Y_n) 的充分性. 因此只要证明给定 $Y_n = y_n$ (即 $T = t$) 时, (Y_1, Y_2, \dots, Y_n) 的条件密度与 θ 无关即可. 下面来证明这一事实.

易见 $(Y_1, \dots, Y_{n-1}, Y_n)$ 的联合密度为

$$f(y_1, y_2, \dots, y_{n-1}, t, \theta) = \theta^n e^{-\theta t} I_{[y_i > 0, i=1,2,\dots,n-1, y_1+\dots+y_{n-1} < t]}.$$

由于 $T(\mathbf{X}) = \sum_{i=1}^n X_i \sim \Gamma(n, \theta)$, 因此 $T = T(\mathbf{X})$ 有密度函数

$$f_T(t) = \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t} I_{[t>0]} = \frac{\theta^n}{(n-1)!} t^{n-1} e^{-\theta t} I_{[t>0]}.$$

给定 $T = t$ 时, (Y_1, \dots, Y_n) 的条件密度为

$$\begin{aligned} f(y_1, \dots, y_n | t) &= \frac{f(y_1, \dots, y_{n-1}, t)}{f_T(t)} \\ &= \frac{(n-1)! \theta^n e^{-\theta t} \cdot I_{[y_i > 0, i=1,\dots,n-1, y_1+\dots+y_{n-1} < t]}}{\theta^n t^{n-1} e^{-\theta t} I_{[t>0]}} \\ &= \frac{(n-1)! I_{[y_i > 0, i=1,\dots,n-1, y_1+\dots+y_{n-1} < t]}}{t^{n-1} I_{[t>0]}} \end{aligned}$$

与 θ 无关, 因此 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为 θ 的充分统计量.

例 2.7.4 在例 2.7.2 中, 令 $T(\mathbf{X}) = X_1$, 则 $T(\mathbf{X})$ 不是充分统计量.

证 在 $T = T(\mathbf{X}) = X_1$ 的条件下, (X_1, \dots, X_n) 的条件密度为

$$\begin{aligned} f(x_1, x_2, \dots, x_n | x_1) &= \frac{f(x_1, \dots, x_n)}{f_T(x_1)} \\ &= f(x_2, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}} \right)^{n-1} \exp \left\{ -\frac{1}{2} \sum_{i=2}^n (x_i - \theta)^2 \right\}, \end{aligned}$$

与 θ 有关, 因此 $T(\mathbf{X}) = X_1$ 不是充分统计量.

此例中 $T(\mathbf{X}) = X_1$ 不是充分统计量, 这个道理是显然的, 因为 $T(\mathbf{X})$ 只使用了一个观察值 X_1 , 把其余观察值 X_2, \dots, X_n 全丢掉了, 它当然不能把 X_1, X_2, \dots, X_n 的全部信息集中起来.

前面的例子表明, 从定义出发验证一个统计量是充分的, 计算太复杂. 幸好有下面的判别方法, 在应用上很方便.

2.7.2 充分性的判别准则——因子分解定理

因子分解定理是由 R.A. Fisher 在 20 世纪 20 年代提出来的, 它的最一般形式和严格数学证明是 Halmos 和 Savage 在 1949 年作出来的.

定理 2.7.1 (因子分解定理) 设样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的概率函数 $f(\mathbf{x}, \theta)$ 依赖于参数 θ , $T = T(\mathbf{X})$ 是一个统计量, 则 T 为充分统计量的充要条件是 $f(\mathbf{x}, \theta)$ 可以分解为

$$f(\mathbf{x}, \theta) = g(t(\mathbf{x}), \theta)h(\mathbf{x}) \quad (2.7.5)$$

的形状. 注意此处函数 $h(\mathbf{x}) = h(x_1, \dots, x_n)$ 不依赖于 θ , $t(\mathbf{x})$ 为 $T(\mathbf{X})$ 的观察值.

这里概率函数是指若 \mathbf{X} 为连续型, 则 $f(\mathbf{x}, \theta)$ 是其密度函数; 若 \mathbf{X} 是离散型, 则 $f(\mathbf{x}, \theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n)$, 即样本 \mathbf{X} 的概率分布.

证 为确定计, 考虑 \mathbf{X} 有密度的情形. 设统计量 $T = (T_1, \dots, T_k)$, T_1, \dots, T_k 都是一维的随机变量, k 一般是较小的自然数, 并可以找到 $n - k$ 维统计量 $\mathbf{Y} = (Y_1, \dots, Y_{n-k})$ 使得变换

$$\mathbf{X} = (X_1, \dots, X_n) \rightarrow (\mathbf{T}, \mathbf{Y}) = (T_1, \dots, T_k, Y_1, \dots, Y_{n-k})$$

是个一一对应的变换, 且有一阶连续偏导数. 假定 \mathbf{X} 的样本空间 \mathcal{X} 和 \mathbf{T} 的样本空间 \mathcal{T} 皆为欧氏的, 即 $\mathcal{X} = R_n$, $\mathcal{T} = R_k$. 由变换是一一对应可知

$$X_i = X_i(\mathbf{T}, \mathbf{Y}) = X_i(T_1, \dots, T_k, Y_1, \dots, Y_{n-k}), \quad i = 1, 2, \dots, n;$$

$$T_j = T_j(X_1, \dots, X_n), \quad j = 1, 2, \dots, k;$$

$$Y_l = Y_l(X_1, \dots, X_n), \quad l = 1, 2, \dots, n - k.$$

变换的 Jacobi 行列式的绝对值为 $|J| = \left| \frac{\partial(x_1, \dots, x_n)}{\partial(t, y)} \right| \triangleq w(t, y)$, 此处 $t = t(\mathbf{x})$ 为 \mathbf{T} 的观察值, \mathbf{y} 为 \mathbf{Y} 的观察值.

充分性的证明. 已知因子分解式成立, 即

$$f(\mathbf{x}, \theta) = g(t(\mathbf{x}), \theta)h(\mathbf{x}).$$

故 $(\mathbf{T}, \mathbf{Y}) = (T_1, \dots, T_k, Y_1, \dots, Y_{n-k})$ 的联合密度为

$$\begin{aligned} k(\mathbf{t}, \mathbf{y}; \theta) &= f(\mathbf{x}, \theta)|J| = g(t(\mathbf{x}), \theta)h(\mathbf{x})w(\mathbf{t}, \mathbf{y}) \\ &= g(\mathbf{t}, \theta)h(x_1(\mathbf{t}, \mathbf{y}), \dots, x_n(\mathbf{t}, \mathbf{y}))w(\mathbf{t}, \mathbf{y}) \\ &= g(\mathbf{t}, \theta)\mu(\mathbf{t}, \mathbf{y}), \end{aligned}$$

此处 $\mu(\mathbf{t}, \mathbf{y}) = h(x_1(\mathbf{t}, \mathbf{y}), \dots, x_n(\mathbf{t}, \mathbf{y}))w(\mathbf{t}, \mathbf{y})$ 与 θ 无关, $\mathbf{T} = T(\mathbf{X})$ 的边缘密度为

$$V(\mathbf{t}; \theta) = \int_{R_{n-k}} k(\mathbf{t}, \mathbf{y}; \theta) d\mathbf{y} = g(\mathbf{t}, \theta) \int_{R_{n-k}} \mu(\mathbf{t}, \mathbf{y}) d\mathbf{y}.$$

故给定 $\mathbf{T} = \mathbf{t}$ 时 \mathbf{Y} 的条件密度为

$$q(\mathbf{y}|\mathbf{t}) = \frac{k(\mathbf{t}, \mathbf{y}; \theta)}{V(\mathbf{t}; \theta)} = \frac{g(\mathbf{t}, \theta)\mu(\mathbf{t}, \mathbf{y})}{g(\mathbf{t}, \theta) \int_{R_{n-k}} \mu(\mathbf{t}, \mathbf{y}) d\mathbf{y}} = \frac{\mu(\mathbf{t}, \mathbf{y})}{\int_{R_{n-k}} \mu(\mathbf{t}, \mathbf{y}) d\mathbf{y}},$$

与 θ 无关, 故 $\mathbf{T} = T(\mathbf{X})$ 为充分统计量.

必要性的证明. 此时已知 $\mathbf{T} = T(\mathbf{X})$ 为充分统计量, 因此给定 $T = \mathbf{t}$ 时 \mathbf{Y} 的条件密度 $q(\mathbf{y}|\mathbf{t})$ 与 θ 无关. (\mathbf{T}, \mathbf{Y}) 的联合密度为

$$k(\mathbf{t}, \mathbf{y}; \theta) = q(\mathbf{y}|\mathbf{t})g(\mathbf{t}, \theta).$$

故通过前面的一一对应变换可知 (X_1, \dots, X_n) 的联合密度为

$$\begin{aligned} f(\mathbf{x}, \theta) &= k(\mathbf{t}, \mathbf{y}; \theta) \left| \frac{\partial(\mathbf{t}, \mathbf{y})}{\partial(x_1, \dots, x_n)} \right| = g(\mathbf{t}, \theta)q(\mathbf{y}|\mathbf{t}) \left| \frac{\partial(\mathbf{t}, \mathbf{y})}{\partial(x_1, \dots, x_n)} \right| \\ &= g(\mathbf{t}, \theta) \cdot h(\mathbf{x}). \end{aligned} \tag{2.7.6}$$

将 $\mathbf{t} = (t_1(x_1, \dots, x_n), \dots, t_k(x_1, \dots, x_n))$ 和 $\mathbf{y} = (y_1(x_1, \dots, x_n), \dots, y_{n-k}(x_1, \dots, x_n))$ 代入到式 (2.7.6) 中 $q(\mathbf{t}|\mathbf{y}) \left| \frac{\partial(\mathbf{t}, \mathbf{y})}{\partial(x_1, \dots, x_n)} \right|$ 的表达式中, 可见它是 x_1, \dots, x_n 的函数, 用 $h(\mathbf{x})$ 表示, 即 $h(\mathbf{x}) = q(\mathbf{t}|\mathbf{y}) \left| \frac{\partial(\mathbf{t}, \mathbf{y})}{\partial(x_1, \dots, x_n)} \right|$. 显见 $h(\mathbf{x})$ 与 θ 无关. 因此由式 (2.7.6) 可知因子分解定理成立. 定理证毕.

推论 2.7.1 设 $T = T(\mathbf{X})$ 为 θ 的充分统计量, $S = \varphi(T)$ 是单值可逆函数, 则 $S = \varphi(T)$ 也是 θ 的充分统计量.

证 由于 $S = \varphi(T)$ 为单值可逆函数, 故

$$\{\mathbf{X} : T(\mathbf{X}) = t_0\} = \{\mathbf{X} : S = \varphi(T) = s_0\}$$

表示相同的事件, 故对任一事件 A ,

$$P(A|T = t_0) = P(A|S = s_0)$$

与 θ 无关, 按定义 2.7.1 可知 $S = \varphi(T)$ 也是充分统计量.

利用因子分解定理也不难证明推论 2.7.1, 这留给读者作为练习. 利用这一推论可知若 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 是充分统计量, 则 \bar{X} 也是充分统计量.

例 2.7.5 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从正态总体 $N(a, \sigma^2)$ 中抽取的简单样本, 令 $\theta = (a, \sigma^2)$, 则 (\bar{X}, S^2) 为 θ 的充分统计量, 此处 \bar{X}, S^2 分别为样本均值和样本方差.

证 样本 \mathbf{X} 的联合密度为

$$\begin{aligned} f(\mathbf{x}, \theta) &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2 \right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n x_i^2 - 2a \sum_{i=1}^n x_i + na^2 \right) \right\} \\ &= g(t(\mathbf{x}), \theta) \cdot h(\mathbf{x}). \end{aligned}$$

此处 $h(\mathbf{x}) \equiv 1$, 由因子分解定理可知 $T(\mathbf{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$ 为 θ 的充分统计量.

由于 $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$ 与 (\bar{X}, S^2) 为一一对应的变换, 由推论 2.7.1 可知 (\bar{X}, S^2) 也是 θ 的充分统计量.

例 2.7.6 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $b(1, \theta)$ 中抽取的简单样本, 则 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 是 θ 的充分统计量.

证 样本 \mathbf{X} 的联合分布是

$$\begin{aligned} f(\mathbf{x}, \theta) &= P_\theta(X_1 = x_1, \dots, X_n = x_n) \\ &= \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} = g(t(\mathbf{x}), \theta)h(\mathbf{x}), \end{aligned}$$

其中 $h(\mathbf{x}) \equiv 1$, 由因子分解定理可知 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为 θ 的充分统计量.

例 2.7.7 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从均匀分布 $U(0, \theta)$ 中抽取的简单样本, 则 $T(\mathbf{X}) = X_{(n)} = \max\{X_1, \dots, X_n\}$ 为 θ 的充分统计量.

证 样本 \mathbf{X} 的联合密度为

$$f(\mathbf{x}, \theta) = \frac{1}{\theta^n} I_{(0, \theta)}(x_{(n)}) = g(t(\mathbf{x}), \theta)h(\mathbf{x}),$$

其中 $h(\mathbf{x}) \equiv 1$. 由因子分解定理可知 $T(\mathbf{X}) = X_{(n)}$ 为 θ 的充分统计量.

例 2.7.8 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从均匀分布 $U(\theta - 1/2, \theta + 1/2)$ 中抽取的简单样本, 其中 $-\infty < \theta < +\infty$, θ 是区间 $[\theta - 1/2, \theta + 1/2]$ 的中点, 也是总体分布的均值. 利用因子分解定理, 验证样本均值 \bar{X} 不是充分统计量.

证 记 $X_{(1)} = \min\{X_1, \dots, X_n\}$, $X_{(n)} = \max\{X_1, \dots, X_n\}$, $T(\mathbf{X}) = (X_{(1)}, X_{(n)})$. 样本 \mathbf{X} 的联合密度为

$$f(\mathbf{x}, \theta) = \begin{cases} 1, & \theta - 1/2 < x_{(1)} \leq x_{(n)} < \theta + 1/2, \\ 0, & \text{其他} \end{cases} = g(t(\mathbf{x}), \theta)h(\mathbf{x}),$$

其中 $h(\mathbf{x}) \equiv 1$, 因此由因子分解定理可知 $T(\mathbf{X})$ 是充分统计量. 由于因子分解定理中的条件是个充要条件, 若 \bar{X} 是充分统计量, 则它必能用因子分解式表出. 上面说明 $f(\mathbf{x}, \theta)$ 不能表成 $g(\bar{x}, \theta)h(\mathbf{x})$ 的形式, 因此 \bar{X} 不是充分统计量.

例 2.7.9 若 $\mathbf{X} = (X_1, \dots, X_n)$ 为从指数族 (2.6.1) 中抽取的简单样本, 则 $T(\mathbf{X}) = (T_1(\mathbf{X}), \dots, T_k(\mathbf{X}))$ 为充分统计量.

证 样本 \mathbf{X} 的联合密度为

$$f(\mathbf{x}, \theta) = C(\theta) \exp \left\{ \sum_{i=1}^k Q_i(\theta) t_i(\mathbf{x}) \right\} h(\mathbf{x}) = g(t(\mathbf{x}), \theta)h(\mathbf{x}),$$

其中 $g(t(\mathbf{x}), \theta) = C(\theta) \exp \left\{ \sum_{i=1}^k Q_i(\theta) t_i(\mathbf{x}) \right\}$, $t(\mathbf{x})$ 为 $T(\mathbf{X})$ 的观察值. 由因子分解定理立得 $T(\mathbf{X}) = (T_1(\mathbf{X}), \dots, T_k(\mathbf{X}))$ 为 θ 的充分统计量.

* **例 2.7.10** 次序统计量的充分性. 设 $\mathcal{F} = \{F\}$ 为一维分布族, 这里对分布函数 F 没有任何限制. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从某个 F 中抽出的简单样本, $T(\mathbf{X}) = (X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 为次序统计量, 则次序统计量 $T(\mathbf{X})$ 是充分的.

证 特别地, 若 F 有密度 f , 则

$$f(x_1, \dots, x_n) = f(x_1) \cdots f(x_n) = f(x_{(1)}) \cdots f(x_{(n)})h(\mathbf{x}),$$

其中 $h(\mathbf{x}) \equiv 1$, f 起着参数 θ 的作用, 则由因子分解定理可知 $T(\mathbf{X}) = (X_{(1)}, \dots, X_{(n)})$ 为充分统计量.

若 \mathcal{F} 为取值至多可数个值的离散分布族, 充分性的证明见文献 [4] 第 50 页例 1.6.5. 一般情形下的严格证明见文献 [1] 第 63 页例 1.5.4.

*2.7.3 极小充分统计量

一个分布族 \mathcal{F} 的充分统计量往往不止一个, 那么在使用中应该如何挑选呢? 我们知道, 统计量是由样本加工而来的, 如本节引言所述, 对样本的加工显然可以提出两条要求: ① 在加工中, 样本所含参数 θ 的信息损失越少越好. 若加工中此种信息毫无损失, 那就是充分性的要求; ② 加工中, 所得统计量越简化越好, 简化的程度可以用统计量的维数来衡量, 也可以用函数关系来表示. 例如, 对一个二维统计量 $T_1(\mathbf{X}) = \left(\sum_{i=1}^m X_i, \sum_{i=m+1}^n X_i \right)$, 再进一步加工得到一维统计量 $T_2 = \sum_{i=1}^n X_i$. 直观上容易看出, T_2 比 T_1 简化. 而且可以看出, T_2 是 T_1 的函数. 一般来说, 若 T 与 S 是两个统计量且 T 是 S 的函数, 即 $T = q(S)$, 那么由函数的定义可知 T 比 S 简化.

定义 2.7.2 设 T 是分布族 \mathcal{F} 的充分统计量, 若对 \mathcal{F} 的任一充分统计量 $S(\mathbf{X})$, 存在一个函数 $q_S(\cdot)$ 使得 $T(\mathbf{X}) = q_S(S(\mathbf{X}))$, 则称 $T(\mathbf{X})$ 是此分布族的极小充分统计量.

虽然这定义表面上看来颇复杂, 实质上是一个简单的概念: 先把样本加工成 $S(\mathbf{X})$, 然后通过 $T = q_S(S)$ 的方式, 再将半成品 $S(\mathbf{X})$ 加工成 $T(\mathbf{X})$. 这两步加工中信息都没有损失.

常用的出现在前面例子中的充分统计量都是极小的, 它们常可用因子分解定理求出来.

可以证明: 一个充分完全统计量 (统计量完全性的概念见 2.8 节) 必是极小充分的, 反之不必对 (见文献 [17]).

*2.8 完全统计量

2.8.1 定义和例子

这个概念与正交函数理论中的完全性概念相似, 但其统计背景不像充分统计量那样好说明. 以后会通过有关问题看出这个概念的意义.

定义 2.8.1 设 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$ 为一分布族, Θ 是参数空间. 设 $T = T(X)$ 为一统计量, 若对任何满足条件

$$E_\theta \varphi(T(X)) = 0, \quad \text{一切 } \theta \in \Theta \tag{2.8.1}$$

的 $\varphi(T(X))$, 都有

$$P_\theta(\varphi(T(X)) = 0) = 1, \quad \text{一切 } \theta \in \Theta, \tag{2.8.2}$$

则称 $T(X)$ 是一完全统计量 (complete statistic).

由定义 2.8.1 可见, 若 $T(X)$ 是完全统计量, 则它的任一 (可测) 函数 $\delta(T)$ 也是完全统计量.

注 2.8.1 统计量 $T(X)$ 的完全性不仅取决于 T 的形状, 还取决于样本 X 的分布族. 完全性 (亦称完备性) 这个名称, 来源于正交函数理论中的一个类似概念. 为简单计, 设统计量 $T(X)$ 有密度函数 $g_\theta(t)$, 则式 (2.8.1) 可写为

$$\int \varphi(t) g_\theta(t) dt = 0, \quad \text{一切 } \theta \in \Theta. \quad (2.8.3)$$

积分 (2.8.3) 形式上可看成 “ φ 与 g_θ 正交”. 于是条件 (2.8.3) \Rightarrow (2.8.2) 可说成是 “若 φ 与函数系 $\{g_\theta, \theta \in \Theta\}$ 正交, 则 φ 必为 0”. 在正交函数论中, 若 M 表示一正交函数系, 且不存在与 M 正交的非零函数, 则称 M 为完全正交函数系. 由式 (2.8.3) \Rightarrow (2.8.2) 可以看出, 这里的完全性正好与正交函数系的完全性相当. 不过不是称密度函数系 $\{g_\theta, \theta \in \Theta\}$ 完全, 而称统计量 T 完全. 由于 $\{g_\theta, \theta \in \Theta\}$ 是由统计量 T 决定的, 这种称呼不影响实质.

统计量的完全性可以通过分布族的完全性来定义, 有兴趣的读者请查看文献 [1] 第 76 页.

例 2.8.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $b(1, \theta)$ 中抽取的简单样本, 则 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 是完全统计量.

证 显然, $T(\mathbf{X}) \sim b(n, \theta)$, 故有

$$P(T(\mathbf{X}) = k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

设 $\varphi(t)$ 为任一实函数, 满足 $E_\theta \varphi(T) = 0$, $0 < \theta < 1$, 此即

$$\begin{aligned} & \sum_{k=0}^n \varphi(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} = 0 \\ \Leftrightarrow & \sum_{k=0}^n \varphi(k) \binom{n}{k} \left(\frac{\theta}{1-\theta}\right)^k = 0, \quad 0 < \theta < 1. \end{aligned}$$

令 $\theta/(1-\theta) = \delta$, 则上式等价于

$$\sum_{k=0}^n \left[\varphi(k) \binom{n}{k} \right] \delta^k = 0, \quad 0 < \delta < \infty.$$

上式左边是 δ 的多项式, 故必有

$$\varphi(k) \binom{n}{k} = 0, \quad k = 0, 1, 2, \dots, n,$$

此处 $\{k : k = 0, 1, 2, \dots, n\}$ 为 T 的支撑集, 由上式可见 $\varphi(T) = 0$, a.s. P_θ 成立. 这就证明了 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 是完全统计量.

例 2.8.2 设 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 为从均匀分布 $U(0, \theta)$ 中抽取的简单样本, 则 $T(\mathbf{X}) = X_{(n)} = \max\{X_1, \dots, X_n\}$ 为完全统计量.

证 $T(\mathbf{X}) = X_{(n)}$ 的密度函数为

$$g_\theta(t) = \begin{cases} nt^{n-1}/\theta^n, & 0 < t < \theta, \\ 0, & \text{其他.} \end{cases}$$

设 $\varphi(t)$ 为 t 的任一实函数, 满足 $E_\theta \varphi(T) = 0$, 此即

$$\frac{n}{\theta^n} \int_0^\theta \varphi(t)t^{n-1} dt = 0, \quad \text{一切 } \theta > 0,$$

即

$$\int_0^\theta \varphi(t)t^{n-1} dt = 0, \quad \text{一切 } \theta > 0.$$

对上式两边关于 θ 求导得

$$\theta^{n-1} \varphi(\theta) = 0 \Leftrightarrow \varphi(\theta) = 0, \quad \text{一切 } \theta > 0.$$

改 θ 为 t , 即 $\varphi(t) = 0, t > 0$, 故 $T(\mathbf{X}) = X_{(n)}$ 为完全统计量.

例 2.8.3 设 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 为从正态总体 $N(\theta, 1)$ 中抽取的简单样本, 则 $T(\mathbf{X}) = \bar{X}$ 为完全统计量.

证 显然 $T(\mathbf{X}) = \bar{X} = \sum_{i=1}^n X_i/n \sim N(\theta, 1/n)$, 设 $\varphi(t)$ 为 t 的任一实函数, 满足 $E_\theta \varphi(T) = 0, -\infty < \theta < \infty$, 即

$$\sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} \varphi(t) e^{-\frac{n(t-\theta)^2}{2}} dt = \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} \varphi(t) e^{-\frac{nt^2}{2}} \cdot e^{-\frac{n\theta^2}{2}} \cdot e^{nt\theta} dt = 0.$$

所以

$$\int_{-\infty}^{\infty} \varphi(t) e^{-\frac{nt^2}{2}} \cdot e^{nt\theta} dt = 0, \quad -\infty < \theta < \infty.$$

令 $z = n\theta$, 则

$$G(z) = \int_{-\infty}^{\infty} \varphi(t) e^{-\frac{nt^2}{2}} e^{tz} dt.$$

将 z 视为复数, $G(z)$ 为复平面上的解析函数, 且 $G(z)$ 当 z 取实数时为 0, 由解析函数的唯一性定理, $G(z)$ 在整个复平面上为 0, 特别取 $z = i\mu$, 则

$$G(\mu) = \int_{-\infty}^{\infty} \varphi(t) e^{-\frac{nt^2}{2}} \cdot e^{-i\mu t} dt = 0.$$

由 Fourier 变换的逆变换公式, 可知

$$\varphi(t) e^{-nt^2/2} = 0.$$

故有 $\varphi(t) = 0$, $|t| < \infty$, 因此 $T(\mathbf{X}) = \bar{X}$ 为完全统计量.

由上面的几个例子可见, 从定义出发判断一个统计量的完全性较复杂, 下列定理将给判别统计量的完全性带来极大的方便.

2.8.2 指数族中统计量的完全性

定理 2.8.1 设样本 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 的概率函数

$$f(\mathbf{x}, \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp \left\{ \sum_{i=1}^k \theta_i T_i(\mathbf{x}) \right\} h(\mathbf{x}), \quad \boldsymbol{\theta} = (\theta_1, \dots, \theta_k) \in \Theta^*$$

为指数族的自然形式. 令 $T(\mathbf{X}) = (T_1(\mathbf{X}), \dots, T_k(\mathbf{X}))$, 若自然参数空间 Θ^* 作为 R_k 的子集有内点, 则 $T(\mathbf{X})$ 是完全统计量.

证 见文献 [1] 第 80 页定理 1.6.1.

例 2.8.4 利用定理 2.8.1 证明例 2.8.1 中的统计量 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为完全统计量.

证 将样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的联合分布表成指数族的自然形式

$$f(\mathbf{x}, \theta) = \theta^{T(\mathbf{x})} (1 - \theta)^{n - T(\mathbf{x})} = (1 + e^\varphi)^{-n} \exp\{\varphi \cdot T(\mathbf{x})\} h(\mathbf{x}),$$

其中 $h(\mathbf{x}) \equiv 1$, $\varphi = \log[\theta/(1 - \theta)]$, 自然参数空间为

$$\Theta^* = \{\varphi : -\infty < \varphi < +\infty\} = (-\infty, \infty),$$

显然自然参数空间 Θ^* 作为 R_1 的子集有内点, 故由定理 2.8.1 可知 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为完全统计量.

例 2.8.5 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从正态总体 $N(a, \sigma^2)$ 中抽取的简单样本, 参数空间 $\Theta = \{\theta = (a, \sigma^2) : -\infty < a < +\infty, \sigma^2 > 0\}$, 则 (\bar{X}, S^2) 为完全统计量, 其中 \bar{X}, S^2 分布为样本均值和样本方差.

证 由例 2.6.6 可知样本 \mathbf{X} 的分布属于指数族, 其自然形式为

$$f(\mathbf{x}, \varphi) = C^*(\varphi) \exp \left\{ \varphi_1 T_1(\mathbf{x}) + \varphi_2 T_2(\mathbf{x}) \right\} h(\mathbf{x}),$$

其中 $h(\mathbf{x}) \equiv 1$, $\varphi_1 = a/\sigma^2$, $\varphi_2 = -\frac{1}{2\sigma^2}$, $\varphi = (\varphi_1, \varphi_2)$. 自然参数空间为

$$\Theta^* = \left\{ (\varphi_1, \varphi_2) : -\infty < \varphi_1 < \infty, -\infty < \varphi_2 < 0 \right\},$$

Θ^* 作为 R_2 的子集显然有内点, 由定理 2.8.1 可知 $(T_1(\mathbf{X}), T_2(\mathbf{X})) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$ 为完全统计量. 由于 $T^*(\mathbf{X}) = (\bar{X}, S^2)$ 为 $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$ 的函数, 由定义可知 $T^*(\mathbf{X})$ 为完全统计量.

例 2.8.6 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从均匀分布 $U(\theta - 1/2, \theta + 1/2)$ 中抽取的简单样本, 则 $T(\mathbf{X}) = (X_{(1)}, X_{(n)})$ 是充分统计量, 但不是完全统计量.

证 $T(\mathbf{X}) = (X_{(1)}, X_{(n)})$ 的充分性在例 2.7.8 中已证. 下面来证明它不是完全的.

要证明一个统计量 $T(\mathbf{X})$ 不是完全的, 只要找到一个实函数 $\varphi(t)$ 使得 $E_\theta \varphi(T) = 0$, 但 “ $\varphi(T) = 0$, a.s. P_θ ” 是不成立的即可.

令 $Y_i = X_i - (\theta - 1/2)$, $i = 1, 2, \dots, n$, 则 Y_1, \dots, Y_n i.i.d. $\sim U(0, 1)$ 与 θ 无关. 而此时 $Z = X_{(n)} - X_{(1)} = Y_{(n)} - Y_{(1)}$ 的分布也与 θ 无关. 找常数 $a < b$ 使得

$$P(Z < a) = P(Z > b) > 0.$$

定义

$$\varphi(t) = \begin{cases} 1, & Z < a, \\ -1, & Z > b, \\ 0, & \text{其他}, \end{cases}$$

则易见 $\varphi(t)$ 满足 $E_\theta \varphi(T) = 0$, 但 $\varphi(t) \not\equiv 0$ (即 $P(\varphi(t) \neq 0) > 0$). 按定义 $T(\mathbf{X}) = (X_{(1)}, X_{(n)})$ 不是完全统计量.

2.8.3 有界完全统计量及其性质

定义 2.8.2 若对任何满足

$$E_\theta \varphi(T(\mathbf{X})) = 0, \quad \text{一切 } \theta \in \Theta$$

的有界 (或 a.s. 有界) 的函数 $\varphi(\cdot)$ 都有

$$P_\theta(\varphi(T(\mathbf{X})) = 0) = 1, \quad \text{一切 } \theta \in \Theta,$$

则称 $T(\mathbf{X})$ 为有界完全统计量.

由定义可见, 一个“完全统计量”必为“有界完全统计量”, 反之不必对.

定理 2.8.2 (Basu 定理) 设 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$ 为一分布族, Θ 是参数空间. 样本 $\mathbf{X} = (X_1, \dots, X_n)$ 是从分布族 \mathcal{F} 中抽取的简单样本, 设 $T(\mathbf{X})$ 是一有界完全统计量, 且是充分统计量. 若 r.v. $V(\mathbf{X})$ 的分布与 θ 无关, 则对任何 $\theta \in \Theta$, $V(\mathbf{X})$ 与 $T(\mathbf{X})$ 独立.

证 见文献 [1] 第 88 页定理 1.6.4.

对指数族, 定理 2.8.2 有如下推论.

推论 2.8.1 设样本 \mathbf{X} 的分布族为指数族, 即

$$f(\mathbf{x}, \theta) = C(\theta) \exp \left\{ \sum_{j=1}^k \theta_j T_j(\mathbf{x}) \right\} h(\mathbf{x}),$$

而自然参数空间 Θ^* 作为 R_k 的子集有内点. 若 r.v. $V(\mathbf{X})$ 的分布与 θ 无关, 对任何 θ , 则 $V(\mathbf{X})$ 与 $T(\mathbf{X}) = (T_1(\mathbf{X}), \dots, T_k(\mathbf{X}))$ 独立.

例 2.8.7 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从 $N(\theta, 1)$ 中抽取的简单样本, $R(\mathbf{X}) = X_{(n)} - X_{(1)}$ 称为极差, 则 $T(\mathbf{X}) = \bar{X} = \sum_{i=1}^n X_i/n$ 与 $R(\mathbf{X})$ 独立.

证 由于正态分布 $N(\theta, 1)$ 为指数族, 自然参数空间 $\Theta^* = \{\theta : -\infty < \theta < \infty\}$ 作为 R_1 的子集有内点. 故 $T(\mathbf{X})$ 为充分完全统计量.

令 $Y_i = X_i - \theta$, 则 $Y_i \sim N(0, 1)$, $i = 1, 2, \dots, n$. 因此 Y_1, \dots, Y_n i.i.d. $\sim N(0, 1)$ 与 θ 无关. 从而 $Y_{(n)} - Y_{(1)}$ 的分布也与 θ 无关. 故

$$R(\mathbf{X}) = X_{(n)} - X_{(1)} = Y_{(n)} - Y_{(1)}$$

的分布与 θ 无关, 由推论 2.8.1 可知 $T(\mathbf{X})$ 与 $R(\mathbf{X})$ 独立.

例 2.8.8 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态总体 $\{N(0, \sigma^2), \sigma > 0\}$ 中抽取的简单样本, 则 $T(\mathbf{X}) = \sum_{i=1}^n X_i^2$ 与随机变量 $\sum_{i=1}^n \lambda_i X_i^2 / \sum_{i=1}^n X_i^2$ 独立, 此处 $\lambda_1, \dots, \lambda_n$ 为任意实数.

证 由于正态分布 $N(0, \sigma^2)$ 为指数族, 记 $\varphi = -1/(2\sigma^2)$, 则自然参数空间 $\Theta^* = \{\varphi : -\infty < \varphi < 0\}$ 作为 R_1 子集有内点. 故 $T(\mathbf{X})$ 为充分完全统计量.

令 $Y_i = X_i/\sigma$, 则 $Y_i \sim N(0, 1)$, $i = 1, 2, \dots, n$, 故 Y_1, \dots, Y_n i.i.d. $\sim N(0, 1)$ 与 φ 无关. 而

$$V(\mathbf{X}) = \frac{\sum_{i=1}^n \lambda_i X_i^2}{\sum_{i=1}^n X_i^2} = \frac{\sum_{i=1}^n \lambda_i Y_i^2}{\sum_{i=1}^n Y_i^2},$$

显然它的分布与 φ 无关, 即与 σ^2 无关, 由推论 2.8.1 可知 $T(\mathbf{X})$ 与 $V(\mathbf{X})$ 独立.

习 题 2

- 设从正态总体 $N(20, 9)$ 中分别抽取容量为 10 和 15 的两组独立样本, 记这两组样本的样本均值和样本方差分别为 \bar{X}, \bar{Y} 和 S_X^2, S_Y^2 . (1) 求两样本均值差的绝对值大于 0.3 的概率. (2) 求 $9S_X^2 + 14S_Y^2$ 大于 164 的概率.
- 设 X_1, \dots, X_n 和 Y_1, \dots, Y_n 是分别从正态总体 $N(\mu, \sigma^2)$ 中抽取的两组简单样本, 且二者相互独立. 令 \bar{X} 和 \bar{Y} 分别为这两组样本的样本均值, 试确定样本大小 n 的近似值, 使得 $P(|\bar{X} - \bar{Y}| > \sigma) \approx 0.01$.
- 设 X_1, \dots, X_n 为从下列总体中抽取的简单样本, 利用特征函数试求样本均值 \bar{X} 的分布:
 - 正态总体 $N(a, \sigma^2)$;
 - 参数为 λ 的 Poisson 总体 $P(\lambda)$;
 - 参数为 λ 的指数分布.
- 设 X_1, \dots, X_n 是从两点分布 $b(1, p)$ 中抽取的简单样本, $0 < p < 1$, \bar{X} 和 S_n^2 为样本均值和样本方差, 求 $S_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$ 的分布 (提示: $S_n^2 = \bar{X}(1 - \bar{X})$).
- 设 X_1, X_2 为取自正态总体 $X \sim N(0, \sigma^2)$ 的一个样本, 证明统计量 X_1/X_2 和 $\sqrt{X_1^2 + X_2^2}$ 是相互独立的.
- 设 X_1, \dots, X_n i.i.d. $\sim F$, $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 是其次序统计量, 已知 $P(X_{(m)} < x) = \sum_{i=m}^n \binom{n}{i} (F(x))^i (1 - F(x))^{n-i}$, 证明下列恒等式:

$$\sum_{i=m}^n \binom{n}{i} (F(x))^i (1 - F(x))^{n-i} = m \binom{n}{m} \int_0^{F(x)} t^{m-1} (1-t)^{n-m} dt.$$

(提示: 为证 $\sum_{i=m}^n \binom{n}{i} p^i (1-p)^{n-i} = m \binom{n}{m} \int_0^p t^{m-1} (1-t)^{n-m} dt$, 注意 $p=0$ 时两边相等, 两边对 p 的导数也一样.)

- 设 r.v. X 的分布函数为 $F(x)$, 密度函数为 $f(x)$, X_1, \dots, X_n i.i.d. $\sim f$, 证明

$$\int_{a < x_1 < \dots < x_n < b} \cdots \int f(x_1) \cdots f(x_n) dx_1 \cdots dx_n = \frac{1}{n!} [F(b) - F(a)]^n.$$

次序统计量 $(X_{(1)}, \dots, X_{(n)})$ 的联合分布由式 (2.3.4) 给出, $X_{(m)}$ 的分布可视为 $(X_{(1)}, \dots, X_{(n)})$ 联合分布的边缘分布, 利用这一事实和上述恒等式求出 $X_{(m)}$ 的密度函数.

8. 设 $X_{(1)}, \dots, X_{(n)}$ 为从均匀分布 $U(0, 1)$ 中抽取的次序统计量,
- 样本容量 n 为多大时, 才能使 $P(X_{(n)} \geq 0.99) \geq 0.95$?
 - 求极差 $R_n = X_{(n)} - X_{(1)}$ 的密度函数;
 - 证明统计量 $Z = 2n(1 - R_n)$ 极限分布为 χ^2_4 .
9. 设总体 $X \sim N(0, 1)$, 令 F 为其分布函数. 从这总体中获取一组样本观察值: $X_1 = 0, X_2 = 0.2, X_3 = 0.25, X_4 = -0.3, X_5 = -0.1, X_6 = 2, X_7 = 0.15, X_8 = 1, X_9 = -0.7, X_{10} = -1$.
- 求上述样本的经验分布函数;
 - 计算 $E\{F(X_{(6)})\}, D\{F(X_{(6)})\}$, $X_{(6)}$ 为容量是 10 的次序统计量.
 - 计算容量 $n = 10$ 的样本中次序统计量 $X_{(6)}$ 的分布函数在 0.2 处的值.

10. 设总体 X 服从威布尔分布, 其分布函数为

$$F(x) = \begin{cases} 1 - e^{-(x/\beta)^\alpha}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

其中 α 为形状参数, β 为刻度参数. (X_1, \dots, X_n) 为从此总体中抽取的简单样本, 试证 $Y = \min(X_1, \dots, X_n)$ 仍服从威布尔分布并指出 Y 的分布的形状参数和刻度参数是什么?

11. 设 X_1, \dots, X_n i.i.d. $\sim N(a, \sigma^2)$, 且 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, 又设 $X_{n+1} \sim N(a, \sigma^2)$ 且与 X_1, \dots, X_n 独立, 试求统计量 $\frac{X_{n+1} - \bar{X}}{S_n} \sqrt{\frac{n-1}{n+1}}$ 的分布.
12. 设 X_1, \dots, X_m i.i.d. $\sim N(\mu_1, \sigma^2), Y_1, \dots, Y_n$ i.i.d. $\sim N(\mu_2, \sigma^2)$, 且 X_1, \dots, X_m 和 Y_1, \dots, Y_n 相互独立, \bar{X} 和 \bar{Y} 分别表示它们的样本均值, S_{1m}^2 和 S_{2n}^2 定义类似上题中的 S_n^2 . α 和 β 是两个给定的实数, 试求

$$T = \frac{\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)}{\sqrt{\frac{mS_{1m}^2 + nS_{2n}^2}{n+m-2} \cdot \left(\frac{\alpha^2}{m} + \frac{\beta^2}{n}\right)}}$$

的分布.

13. 设 X_1, \dots, X_n i.i.d. $\sim N(a, \sigma^2), \bar{X} = \sum_{i=1}^n X_i/n, S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$, 记 $\xi = (X_1 - \bar{X})/S$. 试找出 ξ 与 t 分布的联系, 因而定出 ξ 的密度 (提示: 作正交变换 $Y_1 = \sqrt{n}\bar{X}, Y_2 = \sqrt{\frac{n}{n-1}}(X_1 - \bar{X}), Y_i = \sum_{j=1}^n c_{ij}X_j, i = 3, \dots, n$, 利用定理 2.2.3).

- *14. 设 X_1, \dots, X_n 独立, $X_i \sim N(0, \sigma_i^2), i = 1, \dots, n$, 定义

$$\xi = \sum_{i=1}^n \frac{(X_i - Z)^2}{\sigma_i^2}, \quad \text{其中 } Z = \sum_{i=1}^n \frac{X_i}{\sigma_i^2} / \sum_{i=1}^n \frac{1}{\sigma_i^2},$$

求 ξ 的分布 (提示: 作适当的正交变换).

15. 若 X 服从自由度为 n 的 χ^2 分布, 求证
 (1) X 的特征函数为 $\varphi(t) = (1 - 2it)^{-\frac{n}{2}}$;
 (2) $EX = n$, $D(X) = 2n$;
 (3) 如果 X_1, \dots, X_k 相互独立且 $X_i \sim \chi_{n_i}^2$, 则 $\sum_{i=1}^k X_i \sim \chi_n^2$, 此处 $n = \sum_{i=1}^k n_i$.
16. 设 X_1, \dots, X_n 是从总体 $X \sim \chi_m^2$ 中抽取的大小为 n 的简单样本, 试求样本均值 \bar{X} 的概率分布.
17. 计算自由度为 n 的 χ^2 分布的变异系数和峰度.
18. 设 $X \sim \chi_{2n}^2$, Y 服从参数为 λ 的 Poisson 分布 $P(\lambda)$, 则有 $P(X < 2\lambda) = P(Y \geq n)$.
19. 设 $\xi \sim \chi_n^2$, r 为常数, 计算 $E(\xi^r)$ 和 $D(\xi^r)$, 并指出 r 取哪些值时这些量才存在. 又问 ξ 的密度的最大值在哪一点达到?
- *20. 设 $\mathbf{X} = (X_1, \dots, X_n)'$ 服从 n 元正态分布, $E\mathbf{X} = \boldsymbol{\mu}_{n \times 1}$, $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}_{n \times n} > 0$, 则 $U = (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$ 服从自由度为 n 的 χ^2 分布.
21. 若 r.v. X 服从自由度 n 的 t 分布, $n > 1$, 证明对 $r < n$, $E(X^r)$ 存在且

$$E(X^r) = \begin{cases} n^{\frac{r}{2}} \cdot \frac{\Gamma(\frac{r+1}{2}) \Gamma(\frac{n-r}{2})}{\Gamma(\frac{n}{2}) \Gamma(\frac{1}{2})}, & r \text{ 为偶数,} \\ 0, & r \text{ 为奇数.} \end{cases}$$

22. 设 r.v. $\xi_n \sim t_n$, 计算 ξ_n 的方差 $D(\xi_n)$, 并说明 n 取何值时, $D(\xi_n)$ 才存在.
23. 证明当 $n \rightarrow \infty$ 时 t_n 分布趋于标准正态分布.
24. 设 r.v. X 服从参数为 α, p 的 Gamma 分布 $\Gamma(p, \alpha)$, 求证
 (1) $\Gamma(p, \alpha)$ 的特征函数为 $\varphi(t) = \left(\frac{\alpha}{\alpha - it} \right)^p$;
 (2) $E(X) = p/\alpha$, $D(X) = p/\alpha^2$;
 (3) 如果 $X_i \sim \Gamma(p_i, \alpha)$, $i = 1, 2, \dots, k$ 且 X_1, \dots, X_n 相互独立, 记 $p = \sum_{i=1}^n p_i$, 则 $\sum_{i=1}^n X_i \sim \Gamma(p, \alpha)$;
 (4) 若取 $\alpha = 1/2$, $p = n/2$, 则 $\Gamma(n/2, 1/2)$ 是 χ_n^2 分布.
- *25. 设 Y 服从参数为 λ 的 Poisson 分布 $P(\lambda)$, 当给定 $Y = k$ 时 X 的条件分布为 χ_{n+2k}^2 , 则 X 服从自由度为 n , 非中心参数为 $\sqrt{2\lambda}$ 的非中心 χ^2 分布.
26. 某电子元件寿命服从指数分布, 其密度函数为

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda}, & 0 < x < \infty, \\ 0, & x \leq 0. \end{cases}$$

从这批产品中抽取 n 个作寿命试验, 规定到第 r 个 ($0 < r \leq n$) 电子元件失效时就停止试验. 这样获得前 r 个次序统计量 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(r)}$ 和 n 个电子元件总试验时间 $T = \sum_{i=1}^r X_{(i)} + (n-r)X_{(r)}$, 证明 $2T/\lambda$ 服从自由度为 $2r$ 的 χ^2 分布, 即 $2T/\lambda \sim \chi_{2r}^2$.

27. 设总体 X 服从双指分布, 其分布函数为

$$F(x) = \begin{cases} 1 - \exp\left\{-\frac{x-\mu}{\sigma}\right\}, & x > \mu, \\ 0, & x \leq \mu, \end{cases}$$

其中 $-\infty < \mu < +\infty$, $0 < \sigma < +\infty$, $X_{(1)} \leq \cdots \leq X_{(n)}$ 是样本 X_1, \dots, X_n 的次序统计量, 试证明 $\frac{2(n-i+1)}{\sigma}(X_{(i)} - X_{(i-1)})$ 服从自由度为 2 的 χ^2 分布 ($i = 2, \dots, n$).

28. 设 $X_1 \sim \Gamma(\alpha_1, \lambda)$, $X_2 \sim \Gamma(\alpha_2, \lambda)$, 且 X_1 与 X_2 独立, 则 $Y_1 = X_1 + X_2$ 与 $Y_2 = X_1/(X_1 + X_2)$ 亦独立, 且 $Y_1 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$, $Y_2 \sim \text{Be}(\alpha_1, \alpha_2)$.

29. 证明定理 2.4.2.

30. 证明定理 2.4.3.

31. 设 r.v. $\xi \sim F_{m,n}$, 求 $E(\xi)$ 和 $D(\xi)$, 并指出 m, n 取什么值时这些量才存在?

32. 若 $X \sim F_{m,n}$, 记 $A = F_{m,n}(1-\alpha)$ (即 $\int_0^A f_{m,n}(x)dx = \alpha$), 证明

$$F_{m,n}(1-\alpha) = \frac{1}{F_{n,m}(\alpha)}.$$

33. 设总体 X 的 $2k$ 阶原点矩 $\alpha_{2k} = E(X^{2k}) < \infty$, 证明样本的 k 阶原点矩 $A_k = \sum_{i=1}^n X_i^{2k}/n$ 的渐近分布为 $N(\alpha_k, (\alpha_{2k} - \alpha_k^2)/n)$ (即证明 $\sqrt{n}(A_k - \alpha_k)/\sqrt{\alpha_{2k} - \alpha_k^2} \xrightarrow{\mathcal{L}} N(0, 1)$).

34. 设 $\xi_n \sim \chi_n^2$, 证明当 $n \rightarrow \infty$ 时 $(\xi_n - n)/\sqrt{2n} \xrightarrow{\mathcal{L}} N(0, 1)$. 利用这一事实给出 χ_n^2 的 p 分位点与 $N(0, 1)$ 的 p 分位点之间的一个近似关系.

35. 设 X_1, \dots, X_n 为来自均匀分布总体 $U(0, \theta)$ 的简单样本, 根据中心极限定理, 求样本均值 \bar{X} 的渐近分布.

36. 若 X_1, \dots, X_n 为取自 Poisson 总体 $P(\lambda)$ 的样本. 试证 $(\bar{X} - \lambda)/\sqrt{\bar{X}/n}$ 的极限分布为 $N(0, 1)$.

37. 设 X_1, \dots, X_m 和 Y_1, \dots, Y_n 分别为取自正态总体 $N(\mu, \sigma_1^2)$ 和 $N(\mu, \sigma_2^2)$ 的样本, 试证
明当 m, n 都趋于无穷时, 统计量 $(\bar{X} - \bar{Y})/\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}$ $\xrightarrow{\mathcal{L}} N(0, 1)$. 其中 S_X^2 和 S_Y^2 分别为两组样本的样本方差.

38. 分别举出一个单参数指数族和多参数指数族的例子.

39. 将负二项分布和指数分布写成指数族的标准形式, 并求出其自然参数空间.

40. 设指数族的自然形式为 $f(x, \theta) = C(\theta) \exp\left\{\sum_{j=1}^k \theta_j T_j(x)\right\} h(x)$, 证明

$$E_\theta(T_j(x)) = -\frac{\partial \log C(\theta)}{\partial \theta_j} = -\frac{1}{C(\theta)} \frac{\partial C(\theta)}{\partial \theta_j}, \quad \text{Cov}(T_j(x), T_s(x)) = -\frac{\partial^2 \log C(\theta)}{\partial \theta_j \partial \theta_s}.$$

41. 证明指数族的自然参数空间为凸集.

42. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从 Poisson 分布 $P(\lambda)$ 中抽取的简单样本, 从定义出发证明 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为充分统计量, 再用因子分解定理证明之.
43. 设 X_1, \dots, X_n i.i.d. \sim 几何分布, 试用两种方法证明 $T(x) = \sum_{i=1}^n X_i$ 是充分统计量:
(1) 从定义出发; (2) 用因子分解定理.
44. 设 $T = T(X)$ 是充分统计量, 又 $S(X) = G(T(X))$, 而函数 $S = G(T)$ 是一对一的 (即 $T_1 \neq T_2 \Rightarrow G(T_1) \neq G(T_2)$), 则 S 也是充分统计量.
45. 设 X_1, \dots, X_n i.i.d. \sim 指数分布 $f(x, \lambda) = \lambda e^{-\lambda x} I_{[x>0]}$, 用因子分解定理证明 \bar{X} 是充分统计量.
46. 设 X_1, \dots, X_n i.i.d. $\sim N(\theta, \theta^2)$, $\theta > 0$, 问 \bar{X} 是否为充分统计量?

47. 设 X_1, \dots, X_m i.i.d. $\sim N(a, \sigma^2)$, Y_1, \dots, Y_n i.i.d. $\sim N(b, \sigma^2)$ 且两组样本独立. 记 $\bar{X} = \sum_{i=1}^m X_i/m$, $\bar{Y} = \sum_{j=1}^n Y_j/n$, 而

$$S^2 = \frac{1}{n+m-2} \left[\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2 \right].$$

证明 (\bar{X}, \bar{Y}, S^2) 为充分完全统计量.

48. 设 X_1, \dots, X_n 是从下列总体中抽取的样本, 其密度函数为

$$f(x, \theta) = \frac{1}{2\theta} \exp\left\{-\frac{|x|}{\theta}\right\}, \quad -\infty < x < +\infty, \theta > 0,$$

证明 $T = \sum_{i=1}^n |X_i|$ 是 θ 的充分完全统计量.

49. 设 X_1, \dots, X_n 为取自下列指数分布总体的简单样本,

$$f(x, \theta) = \exp\{-(x - \theta)\}, \quad x > \theta,$$

其中 $-\infty < \theta < +\infty$ 为位置参数. 证明 $T(\mathbf{X}) = X_{(1)}$ 为充分完全统计量.

50. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从均匀分布 $U(-\theta/2, \theta/2)$, $0 < \theta < \infty$ 中抽取的简单样本, 证明 $(X_{(1)}, X_{(n)})$ 为充分统计量, 但不是完全的.
51. 设 X_1, \dots, X_n i.i.d. \sim 均匀分布 $U(\theta, 2\theta)$, $\theta > 0$, 证明 $(X_{(1)}, X_{(n)})$ 是充分统计量, 但不是完全的.

52. 设 X_1, \dots, X_n i.i.d. \sim 两参数的指数分布, 其密度函数为

$$f(x; \lambda, \mu) = \lambda^{-1} \exp\left\{-\frac{x - \mu}{\lambda}\right\} I_{[x > \mu]},$$

其中 $-\infty < \mu < +\infty$, $0 < \lambda < +\infty$. 设 $X_{(1)} \leq \dots \leq X_{(n)}$ 是样本的次序统计量, 证明:

(1) $(X_{(1)}, \sum_{i=1}^n X_{(i)})$ 是 (λ, μ) 的充分统计量. *(2) $X_{(1)}$ 与 $\sum_{i=1}^n (X_i - X_{(1)})$ 独立.

53. 设 X_1, \dots, X_n 为来自 $N(a, \sigma^2)$ 中抽取的 i.i.d. 样本, 证明 \bar{X} 和 $X_{(n)} - X_{(1)}$ 独立.
54. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从 $N(a, \sigma_1^2)$ 中抽取的简单样本, $\mathbf{Y} = (Y_1, \dots, Y_n)$ 是从 $N(b, \sigma_2^2)$ 中抽取的简单样本, 且合样本 $X_1, \dots, X_n; Y_1, \dots, Y_n$ 相互独立, 证明 $T(\mathbf{X}, \mathbf{Y}) = (\bar{X}, \bar{Y}, Q_1^2, Q_2^2)$ 与

$$r(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) / \sqrt{Q_1^2 \cdot Q_2^2}$$

独立, 此处 $\bar{X} = \sum_{i=1}^n X_i/n$, $\bar{Y} = \sum_{j=1}^n Y_j/n$, $Q_1^2 = \sum_{i=1}^n (X_i - \bar{X})^2$, $Q_2^2 = \sum_{j=1}^n (Y_j - \bar{Y})^2$.

- *55. 设 X_1, \dots, X_n i.i.d. $\sim N(a, \sigma^2)$, \bar{X} 为样本均值, $\xi = f(x_1, \dots, x_n)$ 满足条件 $f(x_1+c, \dots, x_n+c) = f(x_1, \dots, x_n)$, 对任何常数 c , 证明 ξ 与 \bar{X} 独立. (提示: 方法一, 用 Basu 定理; 方法二, 作类似定理 2.2.3 的正交变换把 ξ 表示为 Y_1, \dots, Y_n 的函数, 证明此函数只依赖于 Y_2, \dots, Y_n , 而 \bar{X} 只依赖于 Y_1 , 且 Y_1, \dots, Y_n 相互独立.)

第3章 点 估 计

3.1 引 言

3.1.1 参数估计问题

数理统计的任务是用样本去推断总体. 参数估计是统计推断的一种重要形式.

设有参数分布族 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$, 其中 Θ 是参数空间, $f(x, \theta)$ 的分布形式已知, 但其分布与未知参数 θ 有关. X_1, \dots, X_n 是从总体 \mathcal{F} 中抽出的简单随机样本, 要利用样本对未知参数 θ 或其函数 $g(\theta)$ 作出估计. 在参数分布族的场合, 把参数理解为定义在参数空间 Θ 上的实值函数 $g(\theta)$, 一个重要的特例是 $g(\theta) = \theta$. 例如, X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma^2)$, 记 $\theta = (\mu, \sigma^2)$, 希望利用样本对 μ 和 σ^2 或其函数 $g(\theta) = \mu/\sigma^2$ 的值作出估计, 这就是参数估计问题.

有时样本分布族 $\mathcal{F} = \{F\}$ 是非参数分布族, 其中 F 的分布形式未知, 但其均值、方差等都是刻画总体数字特征的量, 也都是参数. 因此在非参数分布族场合, 把参数理解为分布族 \mathcal{F} 上的泛函 $g(F)$, 如 $g(F)$ 为总体均值, 方差或中位数等, 希望利用样本对 $g(F)$ 作出估计, 这也属于参数估计问题. 例如, 从某城市居民中抽取一部分, 对其年收入作调查, 获得样本 X_1, \dots, X_n , 要对该城市居民的年人均收入作出估计, 就属于这类问题.

参数估计 (parameter estimation) 问题常有两类: 点估计和区间估计. 点估计就是用样本函数的一个具体数值去估计一个未知参数. 区间估计就是用样本函数的两个值构成的区间去估计未知参数的取值范围. 例如, 在某市居民年人均收入的调查中, 估计该市居民的年人均收入为 18250 元, 这是一个点估计. 若估计年人均收入在 16350 元到 19850 元之间, 这就是一个区间估计. 点估计与区间估计是互为补充的参数估计形式. 本章主要是讨论参数的点估计问题, 区间估计问题在第 4 章讨论.

3.1.2 点估计

定义 3.1.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从某个总体中抽取的样本, $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n)$ 是样本的函数, 用 $\hat{g}(\mathbf{X})$ 作为 $g(\theta)$ 的估计, 称为点估计 (point estimation).

本章将首先讨论点估计的方法. 对于同一个未知参数 θ (为方便计, 此处以 $g(\theta) = \theta$ 为例) 的估计量可以有很多. 例如, 设 X_1, \dots, X_n 是取自某总体 $F \in \mathcal{F}$ 的一组简单样本. 对此总体的均值 $\theta = E_F(X)$ 可以给出几个估计量

$$\hat{\theta}_1 = \bar{X} = \frac{1}{n}(X_1 + \dots + X_n),$$

$$\hat{\theta}_2 = \frac{1}{2}(X_{(1)} + X_{(n)}),$$

$$\hat{\theta}_3 = m_{1/2},$$

其中 $X_{(1)}$ 和 $X_{(n)}$ 为样本最小和最大次序统计量, $m_{1/2}$ 为样本中位数, 还可以给出其他的估计量. 这就产生了一个问题: 采用哪一个估计量作为 θ 的点估计较好呢? 这就涉及评价一个估计量优劣的标准问题. 标准不同, 回答也不同. 在经典估计理论中, 用来评价估计量好坏的标准有: 无偏性、有效性、相合性和渐近正态性等. 下面将分别叙述这几个准则.

点估计是数理统计学中内容很丰富的一个分支. 它主要包括导出估计量的一般方法, 制定评价估计量优良性的种种合理准则, 寻求某种特定准则下的最优估计量. 本章后面几节将分别介绍矩法、极大似然估计方法、一致最小方差无偏估计和 C-R 不等式及有效估计等内容. 3.6 节介绍了概率密度函数及其导数的核估计方法, 这属于点估计的非参数方法. 有关点估计的内容, 除一部分在本章集中讨论外, 还将在第 7 章关于参数估计的 Bayes 方法中占有一席之地.

3.1.3 点估计的优良性准则

1. 无偏性

在评价估计量好坏时, 一般总希望估计量 $\hat{\theta}(\mathbf{X})$ 的平均值与 θ 越接近越好, 即 $E(\hat{\theta}(\mathbf{X}) - \theta)$ 越小越好. 由于 $\hat{\theta}(\mathbf{X})$ 是随机变量, $\hat{\theta}(\mathbf{X})$ 的值有时比 θ 的真值大, 有时比 θ 的真值小, 希望 $\hat{\theta}(\mathbf{X})$ 在大量重复使用时, 在平均意义下 $\hat{\theta}(\mathbf{X})$ 与 θ 的偏差很小. 期望值 $E(\hat{\theta}(\mathbf{X}) - \theta) = 0$ 时就得到无偏性的概念. 将其一般化, 用 $g(\theta)$ 代替 θ , 用 $\hat{g}(\mathbf{X})$ 代替 $\hat{\theta}(\mathbf{X})$, 得到如下定义.

定义 3.1.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $\{f(x, \theta), \theta \in \Theta\}$ 中抽取的样本, $g(\theta)$ 是定义于参数空间 Θ 上的已知函数. $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n)$ 是 $g(\theta)$ 的一个估计量, 如果

$$E_\theta(\hat{g}(\mathbf{X})) = g(\theta), \quad \theta \in \Theta,$$

则称 $\hat{g}(\mathbf{X})$ 为 $g(\theta)$ 的一个无偏估计 (unbiased estimation). 记 $\hat{g}(\mathbf{X}) = \hat{g}_n(\mathbf{X})$, 若 $E_\theta(\hat{g}_n(\mathbf{X})) \neq g(\theta)$, 但

$$\lim_{n \rightarrow \infty} E_\theta(\hat{g}_n(\mathbf{X})) = g(\theta), \quad \theta \in \Theta,$$

则称 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的渐近无偏估计 (asymptotically unbiased estimation).

无偏性的含义有两个: 第一个含义是无系统偏差. 由于样本的随机性, $\hat{g}(\mathbf{X})$ 是样本的函数, 因此它是一随机变量, 用估计量 $\hat{g}(\mathbf{X})$ 去估计 $g(\theta)$, 对某些样本, $\hat{g}(\mathbf{X})$ 与 $g(\theta)$ 相比偏低; 对另一些样本, $\hat{g}(\mathbf{X})$ 偏高; 无偏性表明, 把这些正负偏差在概率上平均起来, 其值为 0. 如用一杆秤去称东西, 误差的来源有两个: ① 这杆秤自身结构上有问题, 用它称东西总是倾向于偏高或总是倾向于偏低, 这属于系统误差; ② 另一种误差是随机误差, 由不可控制的因素产生, 如温度、湿度和工作人员心理波动等影响造成的, 这属于随机误差. 无偏性相当于要求无系统误差. 随机误差总是存在的, 大量重复使用无偏估计量, 误差有时为正, 有时为负, 但随机误差可以在重复使用中正负相抵消.

无偏性的另一个含义是: 要求估计量大量重复使用, 在多次重复使用下给出接近真值 $g(\theta)$ 的估计. 设想这样一种情况: 每天抽样对 $g(\theta)$ 进行估计, 第 i 天的样本为 $\mathbf{X}^{(i)} = (X_1^{(i)}, \dots, X_n^{(i)})$, 估计值为 $\hat{g}(\mathbf{X}^{(i)})$, 一共作了 m 天. 设 $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}$ 是独立同分布的样本, 如 $\hat{g}(\mathbf{X})$ 有无偏性, 按大数定律有

$$P\left(\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \hat{g}(\mathbf{X}^{(i)}) = g(\theta)\right) = 1.$$

也就是说, 尽管一次估计结果 $\hat{g}(\mathbf{X}^{(i)})$ 不一定恰好等于 $g(\theta)$, 但在大量重复使用时, 多次估计的算术平均值, 可以任意接近 $g(\theta)$. 如果这一估计量 $\hat{g}(\mathbf{X})$ 只使用一次, 无偏性这个概念就失去意义.

不妨设想一个例子. 某工厂生产一种产品, 从长期来看废品率稳定在 p_0 上. 设商店每日从这工厂进货一批, 每批 N 件, 每件单价 a 元. 双方议定: 若某日抽样中 p_0 的估计值为 \hat{p} , 则商店付给工厂 $N(1 - \hat{p})a$ 元. 这时, 就一日的情况而言, \hat{p} 相对 p_0 可能偏高, 也可能偏低, 因而有一方要吃一点亏. 但从长远看, 若 \hat{p} 为 p_0 的无偏估计, 则平均说来哪一方也不吃亏. 无偏性保证从长远来看是公平的. 如果应用中无这经常的重复使用性, 无偏性也就失去意义.

在点估计理论中, 目前无偏性仍占有重要的地位. 除了历史因素外, 还有两个原因. 一是无偏性的要求只涉及一阶矩 (均值), 在数学处理时较方便; 二是在没有其他合理准则可循时, 人们心理上觉得: 一个具有无偏性的估计, 总比没有这种性质的估计要好些.

例 3.1.1 设 X_1, \dots, X_n 是取自期望为 μ , 方差为 σ^2 的总体的一个样本. 显然样本均值 \bar{X} 是 μ 的无偏估计. 证明样本方差 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1)$ 是 σ^2 的无偏估计.

证 显然

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} \left[\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \right] = \frac{n}{n-1} [E(X_1^2) - E(\bar{X}^2)] \\ &= \frac{n}{n-1} \left[(\sigma^2 + \mu^2) - (\sigma^2/n + \mu^2) \right] = \sigma^2, \end{aligned}$$

故样本方差 S^2 是 σ^2 的无偏估计.

2. 有效性

在应用中, 同一个参数的无偏估计常常不止一个, 那么选用哪一个无偏估计更好呢? 为了解决好这一问题, 就要讨论估计量的有效性 (efficiency). 设 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 为 θ 的两个无偏估计, 由无偏性可知它们的一阶原点矩相等, 比较它们的二阶中心矩——方差, 方差越小越好.

举一个例子来说明这个问题. 到商店购买电视机, 看中了其中的两种品牌, 它们分别由甲、乙两厂生产, 外观、音质和画面都不错. 根据市场调查, 甲乙两厂生产的这两种电视机平均使用寿命相同, 都是 20 年. 甲厂生产的电视机质量较稳定, 最低使用寿命有 18 年, 最高可使用 22 年; 而乙厂生产的电视机质量稳定性要差一些, 最差的使用 10 年就坏了, 但最好的可使用 30 年. 选用哪一个厂家生产的电视机呢? 若将电视机的使用寿命视为随机变量, 甲乙两厂生产的电视机使用寿命的均值相等, 但乙厂的质量不稳定, 即方差较大. 从稳健的角度出发, 显然愿意购买甲厂生产的电视机, 其风险较小, 即方差较小, 质量稳定.

定义 3.1.3 设 $\hat{g}_1(\mathbf{X}) = \hat{g}_1(X_1, \dots, X_n)$ 和 $\hat{g}_2(\mathbf{X}) = \hat{g}_2(X_1, \dots, X_n)$ 为 $g(\theta)$ 的两个不同无偏估计量, 若

$$D_\theta(\hat{g}_1(\mathbf{X})) \leq D_\theta(\hat{g}_2(\mathbf{X})), \quad \text{一切 } \theta \in \Theta,$$

且至少存在一个 $\theta \in \Theta$, 使得严格不等号成立, 则称估计量 $\hat{g}_1(\mathbf{X})$ 比 $\hat{g}_2(\mathbf{X})$ 有效.

从这个定义出发可以看出, 在均值相等的条件下, 方差越小的估计量越有效. 例如, X_1, \dots, X_n 是取自总体 F 的一个简单样本, 设总体均值 μ 和总体方差 σ^2 都存在, 则 $\hat{\theta}_1 = X_1$ 和 $\hat{\theta}_2 = \bar{X}$ 都是总体均值 μ 的无偏估计量, 它们的方差分别是

$$D(\hat{\theta}_1) = \sigma^2, \quad D(\hat{\theta}_2) = \frac{1}{n} \sigma^2.$$

后者方差更小, 可见 \bar{X} 比 X_1 更有效, 且 n 越大, \bar{X} 对 μ 的估计就越有效. 这就是在物体的称重问题中, 为什么要将物体称 n 次, 用其平均值作为物重的理由.

3. 相合性

大量实践表明, 随着样本容量 n 的增加, 估计量 $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n)$ 与被估计参数 $g(\theta)$ 的偏差越来越小, 这是一个良好估计量应具有的性质. 试想, 若不然,

无论做多少次试验, 也不能把 $g(\theta)$ 估计到任意指定的精确程度, 这样的估计量显然是不可取的.

定义 3.1.4 设对每个自然数 n , $\hat{g}_n(\mathbf{X}) = \hat{g}_n(X_1, \dots, X_n)$ 是 $g(\theta)$ 一个估计量, 若 $\hat{g}_n(\mathbf{X})$ 依概率收敛到 $g(\theta)$, 即对任何 $\theta \in \Theta$ 及 $\varepsilon > 0$ 有

$$\lim_{n \rightarrow \infty} P_\theta(|\hat{g}_n(\mathbf{X}) - g(\theta)| \geq \varepsilon) = 0,$$

则称 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的弱相合估计 (weakly consistent estimation). 若对任何 $\theta \in \Theta$ 有

$$P_\theta \left(\lim_{n \rightarrow \infty} \hat{g}_n(\mathbf{X}) = g(\theta) \right) = 1,$$

则称 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的强相合估计 (strongly consistent estimation). 若 $r > 0$ 和对任何 $\theta \in \Theta$, 有

$$\lim_{n \rightarrow \infty} E_\theta |\hat{g}_n(\mathbf{X}) - g(\theta)|^r = 0,$$

则称 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的 r 阶矩相合估计 (consistent estimation in r 'th mean). 当 $r = 2$ 时称为均方相合估计 (consistent estimation in quadratic mean).

估计量的相合性是对大样本问题提出的要求, 是估计量的一种大样本性质.

由概率论中关于这几种收敛性的关系, 可知上述三种相合性有如下关系: 强相合 \Rightarrow 弱相合, 反之不必对; 对任何 $r > 0$ 有: r 阶矩相合 \Rightarrow 弱相合, 反之不必对; 又强相合与 r 阶矩相合之间没有包含关系.

估计量的大样本性质, 还有渐近正态性, 将在本章后面有关小节中给出定义.

例 3.1.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从均匀分布 $U(0, \theta)$ 中抽取的简单样本, θ 为未知参数. 证明 $T(\mathbf{X}) = \left(\prod_{i=1}^n X_i \right)^{1/n}$ 是 $g(\theta) = \theta e^{-1}$ 的强相合估计.

证 令 $Y_i = \ln X_i$, $i = 1, \dots, n$, 则 Y_1, \dots, Y_n 相互独立, 具有下列共同分布

$$f(y, \theta) = \frac{1}{\theta} e^y I_{(-\infty, \ln \theta)}(y)$$

且

$$E(Y_1) = \frac{1}{\theta} \int_{-\infty}^{\ln \theta} y e^y dy = \frac{1}{\theta} \left[y e^y \Big|_{-\infty}^{\ln \theta} - \int_{-\infty}^{\ln \theta} e^y dy \right] = \ln \theta - 1.$$

故由强大数定律可知

$$\ln T(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \ln X_i = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y} \xrightarrow{\text{a.s.}} E(Y_1) = \ln \theta - 1.$$

因此有

$$T(\mathbf{X}) = e^{\bar{Y}} \xrightarrow{\text{a.s.}} e^{\ln \theta - 1} = \theta \cdot e^{-1}.$$

3.2 矩 估 计

3.2.1 矩法和矩估计量

设 X_1, \dots, X_n 是从总体 F 中抽取的简单随机样本. 这时, 样本矩可用来估计 F 的相应的总体矩. 即样本 k 阶原点矩

$$a_{nk} = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots \quad (3.2.1)$$

是总体 k 阶原点矩 $\alpha_k = E(X^k)$ 的“自然”的矩估计量. 特别总体均值 $\alpha_1 = E(X)$ 的“自然”的矩估计量是样本均值 $a_{n1} = \bar{X}$.

样本 k 阶中心矩

$$m_{nk} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad k = 2, 3, \dots \quad (3.2.2)$$

是总体 k 阶中心矩 $\mu_k = E[X - E(X)]^k$ 的“自然”的矩估计量. 特别总体方差 $\mu_2 = E(X - EX)^2$ 的“自然”的矩估计量是 $m_{n2} = S_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / n$, 它与样本方差 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1)$ 只差一个常数因子.

用 a_{nk}, m_{nk} 分别估计 α_k 和 μ_k 是一种基于直观的方法, 它的依据是 a_{nk} 是 α_k 的无偏估计, 即

$$Ea_{nk} = \frac{1}{n} \sum_{i=1}^n E(X_i^k) = \frac{1}{n} \sum_{i=1}^n \alpha_k = \alpha_k. \quad (3.2.3)$$

但用 m_{nk} 估计 μ_k , 一般不是无偏的, 当样本大小 n 较大时, 偏差不显著, 且必要时可作一些修正, 使之成为无偏估计. 请看例 3.2.1.

例 3.2.1 设 $\mu_2 = \sigma^2$ 是总体 X 的方差, 令 X_1, \dots, X_n i.i.d. $\sim X$, 则 $S_n^2 = m_{n2}$ 不是 σ^2 的无偏估计.

证 由例 3.1.1 可知

$$\begin{aligned} E(S_n^2) &= E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{n-1}{n} E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\ &= \frac{n-1}{n} E(S^2) = \frac{n-1}{n} \sigma^2. \end{aligned} \quad (3.2.4)$$

因而 m_{n2} 不是 σ^2 的无偏估计, 且是系统地偏低. 将其修正, 只需用

$$m_{n2}^* = \frac{n}{n-1} m_{n2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = S^2 \quad (3.2.5)$$

代替 m_{n2} , 就得到 $E(m_{n2}^*) = E(S^2) = \sigma^2$, 即 S^2 为总体方差的无偏估计. 这就是用 S^2 作为样本方差的定义, 而不用 m_{n2} 的理由所在. 但当 $k \geq 4$, 就不能通过这样简单的修正得出 μ_k 的无偏估计.

一般地, 样本的 k 阶中心矩可以用样本的原点矩表出 (令 $a_{n0} = 1$)

$$\begin{aligned} m_{nk} &= \frac{1}{n} \sum_{i=1}^n (X_i - a_{n1})^k = \frac{1}{n} \sum_{i=1}^n \sum_{r=0}^k \binom{k}{r} (-1)^{k-r} X_i^r a_{n1}^{k-r} \\ &= \sum_{r=0}^k \left(\frac{1}{n} \sum_{i=1}^n X_i^r \right) (-1)^{k-r} \binom{k}{r} a_{n1}^{k-r} \\ &= \sum_{r=0}^k (-1)^{k-r} \binom{k}{r} a_{nr} a_{n1}^{k-r}. \end{aligned} \quad (3.2.6)$$

下面给出矩法及矩估计量的定义.

定义 3.2.1 设有总体分布族 $\{f(x, \theta), \theta \in \Theta\}$, Θ 是参数空间, $g(\theta)$ 是定义在 Θ 上参数 θ 的函数, 它可以表示为总体分布的某些矩的函数, 即

$$g(\theta) = G(\alpha_1, \dots, \alpha_k; \mu_2, \dots, \mu_s). \quad (3.2.7)$$

设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从上述分布族中抽取的简单样本, 将式 (3.2.7) 中的 α_i 和 μ_j 分别用它们“自然”的矩估计量 a_{ni} 和 m_{nj} 代替, 得

$$\hat{g}(\mathbf{X}) = G(a_{n1}, \dots, a_{nk}; m_{n2}, \dots, m_{ns}), \quad (3.2.8)$$

则 $\hat{g}(\mathbf{X})$ 作为 $g(\theta)$ 的估计量, 称为 $g(\theta)$ 的矩估计量 (moment estimate). 这种求矩估计量的方法称为矩法 (moment method of estimation).

3.2.2 若干例子

例 3.2.2 设 X_1, \dots, X_n 是从具有成功概率 θ 的两点分布总体 $b(1, \theta)$ 中抽取的简单样本, 求 θ 和 $g(\theta) = \theta(1 - \theta)$ 的矩估计量.

解 设 $X \sim b(1, \theta)$, 则有 $P(X = x) = \theta^x(1 - \theta)^{1-x}$, $x = 0, 1$. 由于 $E(X) = \theta$, 所以 θ 的矩估计量就是 $\hat{\theta}(X_1, \dots, X_n) = \bar{X}$, 按定义 $g(\theta)$ 的矩估计量是

$$\hat{g}(X_1, \dots, X_n) = \bar{X}(1 - \bar{X}).$$

例 3.2.3 设 X_1, \dots, X_n i.i.d. \sim 均匀分布 $U(\theta_1, \theta_2)$, 参数 $\theta = (\theta_1, \theta_2)$, 其中 $-\infty < \theta_1 < \theta_2 < +\infty$. 求 θ_1 和 θ_2 的矩估计量.

解 设 $X \sim U(\theta_1, \theta_2)$, 由均匀分布的性质可知

$$E(X) = \alpha_1 = (\theta_1 + \theta_2)/2,$$

$$D(X) = \mu_2 = (\theta_2 - \theta_1)^2/12.$$

解此方程组得

$$\theta_1 = \alpha_1 - \sqrt{3\mu_2}, \quad \theta_2 = \alpha_1 + \sqrt{3\mu_2}.$$

将上式中 α_1 和 μ_2 分别用 \bar{X} 和 m_{n2} 代入得

$$\begin{aligned}\hat{\theta}_1(X_1, \dots, X_n) &= \bar{X} - \sqrt{3m_{n2}} = \bar{X} - \sqrt{3S_n}, \\ \hat{\theta}_2(X_1, \dots, X_n) &= \bar{X} + \sqrt{3S_n},\end{aligned}$$

其中 S_n^2 由式 (3.2.4) 给出 (亦可用 S 代替 S_n , 此处 S^2 由式 (3.2.5) 给出).

例 3.2.4 设总体分布有概率密度

$$f(x, \theta) = \begin{cases} 2\sqrt{\frac{\theta}{\pi}} \exp\{-\theta x^2\}, & x > 0, \\ 0, & x \leq 0,\end{cases}$$

其中 $\theta > 0$ 为未知参数. 这个分布称为 Maxwell 分布, 在气体分子动力学中有应用. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为抽自此总体的简单随机样本, 求 $g(\theta) = 1/\theta$ 的矩估计量.

解 设 $X \sim f(x, \theta)$, 由方程

$$\alpha_1 = E(X) = 2\sqrt{\frac{\theta}{\pi}} \int_0^\infty x e^{-\theta x^2} dx = \frac{1}{\sqrt{\pi\theta}},$$

解得 $g(\theta) = 1/\theta = \pi\alpha_1^2$, 将 α_1 用 \bar{X} 代替得

$$\hat{g}_1(\mathbf{X}) = \pi\bar{X}^2.$$

另一方面, 由另一方程

$$\alpha_2 = E_\theta(X^2) = 2\sqrt{\frac{\theta}{\pi}} \int_0^\infty x^2 e^{-\theta x^2} dx = \frac{1}{2\theta},$$

解得 $g(\theta) = 1/\theta = 2\alpha_2$, 将 α_2 用 $a_{n2} = \sum_{i=1}^n X_i^2/n$ 代入, 得 $g(\theta)$ 的矩估计

$$\hat{g}_2(\mathbf{X}) = 2a_{n2} = \frac{2}{n} \sum_{i=1}^n X_i^2.$$

由此可见矩估计不唯一. 这两个矩估计量 $\hat{g}_1(\mathbf{X})$ 和 $\hat{g}_2(\mathbf{X})$ 中哪一个更好? 以后可以证明基于 a_{n2} 的估计量 $\hat{g}_2(\mathbf{X})$, 在 $g(\theta)$ 一切无偏估计类中是方差最小者. 基于 a_{n1} 的估计 $\hat{g}_1(\mathbf{X})$ 不是 $g(\theta)$ 的无偏估计.

例 3.2.5 设 X_1, \dots, X_n 为从总体 X 中抽取的 i.i.d. 样本, 求总体 X 的变异系数、偏度和峰度

$$\nu = \frac{\sqrt{\mu_2}}{\alpha_1}, \quad \beta_1 = \frac{\mu_3}{\mu_2^{3/2}}, \quad \beta_2 = \frac{\mu_4}{\mu_2^2} - 3$$

的矩估计量.

解 总体变异系数, 偏度和峰度的定义在 1.3 节中已给出, 此处不再解释其意义. 将上述表达式中的 α_1, μ_2, μ_3 和 μ_4 分别用它们的矩估计 a_{n1}, m_{n2}, m_{n3} 和 m_{n4} 代替, 得到 ν, β_1, β_2 的矩估计, 即由式 (1.3.2)- 式 (1.3.4) 给出的样本变异系数, 样本偏度和样本峰度为

$$\hat{\nu}(X_1, \dots, X_n) = \frac{\sqrt{m_{n2}}}{a_{n1}},$$

$$\hat{\beta}_1(X_1, \dots, X_n) = \frac{m_{n3}}{m_{n2}^{3/2}},$$

$$\hat{\beta}_2(X_1, \dots, X_n) = \frac{m_{n4}}{m_{n2}^2} - 3.$$

例 3.2.6 设总体分布有概率密度

$$f(x, \theta) = \begin{cases} \frac{\theta_2}{\Gamma((1+\theta_1)/\theta_2)} x^{\theta_1} \exp\{-x^{\theta_2}\}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

参数 $\theta = (\theta_1, \theta_2)$ 的变化范围 $-1 < \theta_1 < \infty, \theta_2 > 0$. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为抽自此总体的简单随机样本, 求 θ_1 和 θ_2 的矩估计量.

解 由简单计算得到总体分布的前两阶矩为

$$\begin{aligned} \alpha_1 &= \Gamma\left(\frac{2+\theta_1}{\theta_2}\right) / \Gamma\left(\frac{1+\theta_1}{\theta_2}\right), \\ \alpha_2 &= \Gamma\left(\frac{3+\theta_1}{\theta_2}\right) / \Gamma\left(\frac{1+\theta_1}{\theta_2}\right). \end{aligned} \quad (3.2.9)$$

按矩估计方法, 用 a_{n1} 和 a_{n2} 分别代替式 (3.2.9) 中的 α_1 和 α_2 , 用 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别代替 θ_1 和 θ_2 , 得到如下的方程组:

$$\begin{aligned} a_{n1} &= \Gamma\left(\frac{2+\hat{\theta}_1}{\hat{\theta}_2}\right) / \Gamma\left(\frac{1+\hat{\theta}_1}{\hat{\theta}_2}\right), \\ a_{n2} &= \Gamma\left(\frac{3+\hat{\theta}_1}{\hat{\theta}_2}\right) / \Gamma\left(\frac{1+\hat{\theta}_1}{\hat{\theta}_2}\right). \end{aligned}$$

其解就是 θ_1 和 θ_2 的矩估计. 但此处得不出 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 的简单解析表达式, 而只能用数值方法. 此例说明不是所有的矩估计都有解析表达式.

矩估计方法也可以用于多维样本, 请看下例.

例 3.2.7 设 $(X_i, Y_i), i = 1, 2, \dots, n$ 为从一个二维总体中抽取的简单随机样本, 求总体分布的协方差 σ_{12} 和相关系数 ρ 的矩估计.

解 按定义 σ_{12} 的矩估计量是样本协方差, 即

$$m_{12} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}), \quad (3.2.10)$$

其中 $\bar{X} = \sum_{i=1}^n X_i / n$, $\bar{Y} = \sum_{j=1}^n Y_j / n$, 而 ρ 的矩估计量是样本相关系数, 即

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{j=1}^n (Y_j - \bar{Y})^2}}. \quad (3.2.11)$$

3.2.3 矩估计的无偏性和渐近无偏性

矩法是由 K. Pearson 在 1894 年提出的点估计的古老方法. 它的特点是直观性强, 用此法获得估计量简便、易行, 且不要求事先知道总体的分布, 在一定条件下矩估计量还具有相合性和渐近正态性. 它的缺点是: 在参数分布族场合, 没有充分利用提供的有关参数的信息, 小样本性质不突出. 此外, 矩估计量不具唯一性.

下面研究矩估计的三方面的性质: 小样本性质有无偏性, 大样本性质有相合性和渐近正态性. 其大样本性质将放在下面两小节考虑.

估计量的无偏性和渐近无偏性的定义在 3.1 节中已给出, 下面讨论矩估计的无偏性和渐近无偏性.

(1) 样本 k 阶原点矩 a_{nk} 是总体 k 阶原点矩 α_k ($k = 1, 2, \dots$) 的无偏估计, 式 (3.2.3) 已给出了证明.

(2) 对 $k \geq 2$, 样本 k 阶中心矩不是总体 k 阶中心矩的无偏估计.

(i) 由例 3.2.1 可知

$$E(m_{n2}) = \frac{n-1}{n} \mu_2,$$

将其修正, 得

$$S^2 = \left(\frac{n}{n-1} \right) m_{n2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

是 μ_2 的无偏估计.

(ii) 经过计算可知样本 3 阶中心矩 m_{n3} 也不是总体 3 阶中心矩 μ_3 的无偏估计, 事实上

$$E(m_{n3}) = \frac{(n-1)(n-2)}{n^2} \mu_3. \quad (3.2.12)$$

将其修正, 得

$$m_{n3}^* = \frac{n^2}{(n-1)(n-2)} m_{n3},$$

它是 μ_3 的无偏估计.

(iii) 更进一步, 可以证明对 $\nu \geq 4$ 有

$$E(m_{n\nu}) = \mu_\nu + O\left(\frac{1}{n}\right), \quad (3.2.13)$$

因此对 $\nu \geq 4$, $m_{n\nu}$ 也不是总体 ν 阶中心矩 μ_ν 的无偏估计.

(3) 当待估函数 $g(\theta)$ 为总体分布一些矩的函数时, 即

$$g(\theta) = G(\alpha_1, \dots, \alpha_k; \mu_2, \dots, \mu_s) \quad (3.2.14)$$

时, 矩估计

$$\hat{g}(X_1, \dots, X_n) = G(a_{n1}, \dots, a_{nk}; m_{n2}, \dots, m_{ns}) \quad (3.2.15)$$

一般不是 $g(\theta)$ 的无偏估计, 例 3.2.4 就是一例. 但若 $g(\theta)$ 为若干总体原点矩的线性组合时, 即

$$g(\theta) = h(\alpha_1, \dots, \alpha_m) = \sum_{i=1}^m c_i \alpha_i$$

时, 其矩估计

$$\hat{g}(X_1, \dots, X_n) = \sum_{i=1}^m c_i a_{ni}$$

是 $g(\theta)$ 的无偏估计.

(4) 矩估计一般具有渐近无偏性. 由前面第 2 段可见 $m_{n\nu}$ ($\nu \geq 2$) 是总体 ν 阶中心矩 μ_ν 的渐近无偏估计, 如

$$\lim_{n \rightarrow \infty} E(m_{n3}) = \lim_{n \rightarrow \infty} \frac{(n-1)(n-2)}{n^2} \mu_3 = \mu_3.$$

当待估参数函数 $g(\theta)$ 是式 (3.2.14) 形式时, 其矩估计 (3.2.15) 是否具有渐近无偏性? 可以证明对函数 G 加上适当的条件, 这个结论是对的 (见文献 [3] 第 338 页).

例 3.2.8 设 X_1, \dots, X_n i.i.d. \sim 指数分布 $\text{Exp}(\lambda)$, 其密度为 $f(x, \lambda) = \lambda e^{-\lambda x} \cdot I_{[x>0]}$. 求 λ 的矩估计量 $\hat{\lambda}$, 并讨论它的无偏性.

解 由于 $\alpha_1 = \int_0^\infty x f_\lambda(x) dx = 1/\lambda$, 解得 $\lambda = 1/\alpha_1$, 将 α_1 用其矩估计量 $\bar{X} = a_{n1}$ 代入, 得到 λ 的矩估计量为

$$\hat{\lambda}(X_1, \dots, X_n) = \frac{1}{\bar{X}}.$$

由于 $Y = \sum_{i=1}^n X_i \sim \text{Gamma 分布 } \Gamma(n, \lambda)$,

$$E(\hat{\lambda}) = nE\left(\frac{1}{Y}\right) = n \int_0^\infty \frac{1}{y} \cdot \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y} dy = \frac{n}{n-1} \lambda,$$

可见 $\hat{\lambda}(X_1, \dots, X_n) = 1/\bar{X}$ 不是 λ 的无偏估计. 因此矩估计量不一定都具有无偏性. 但 $\lim_{n \rightarrow \infty} E(\hat{\lambda}) = \lim_{n \rightarrow \infty} [n/(n-1) \cdot \lambda] = \lambda$, 故 $\hat{\lambda}(X_1, \dots, X_n)$ 是 λ 的渐近无偏估计. 对 $\hat{\lambda}$ 略作修正, 可得 λ 的一个无偏估计

$$\hat{\lambda}^*(X_1, \dots, X_n) = \frac{n-1}{n} \cdot \hat{\lambda}(X_1, \dots, X_n) = \frac{n-1}{n} \cdot \frac{1}{\bar{X}}.$$

3.2.4 矩估计的相合性

估计量的几种相合性的定义已在 3.1 节中给出. 一般说来, 矩估计在较一般的条件下具有相合性. 此处给出矩估计的强相合性, 显然相应的矩估计的弱相合性也成立.

(1) 样本 k 阶原点矩是总体 k 阶原点矩的强相合估计. 设 X_1, \dots, X_n 是从总体 F 中抽取的简单随机样本, a_{nk} 为样本的 k 阶原点矩, α_k 为总体的 k 阶原点矩. 由独立同分布场合的柯尔莫哥洛夫强大数律可知 $a_{nk} \xrightarrow{\text{a.s.}} \alpha_k$, $k = 1, 2, \dots$, 即

$$P\left(\lim_{n \rightarrow \infty} a_{nk} = \alpha_k\right) = 1, \quad k = 1, 2, \dots$$

(2) 样本 k 阶中心矩是总体 k 阶中心矩的强相合估计. 这一结论的证明用到下列结论.

引理 3.2.1 设函数 $f(y_1, \dots, y_k)$ 在 (c_1, c_2, \dots, c_k) 处连续, 若 $y_{n1} \xrightarrow{\text{a.s.}} c_1, \dots, y_{nk} \xrightarrow{\text{a.s.}} c_k$, 则 $f(y_{n1}, \dots, y_{nk}) \xrightarrow{\text{a.s.}} f(c_1, c_2, \dots, c_k)$.

这一引理的证明留作练习. 利用引理 3.2.1 证明上述结论: 由于

$$\mu_k = \sum_{r=0}^k (-1)^{k-r} \binom{k}{r} \alpha_r \alpha_1^{k-r} = f(\alpha_1, \alpha_2, \dots, \alpha_k),$$

显见 $f(\cdot)$ 是其变元的连续函数. 由 $a_{ni} \xrightarrow{\text{a.s.}} \alpha_i$, $i = 1, \dots, k$, 利用引理 3.2.1 及式 (3.2.6) 得

$$m_{nk} = f(a_{n1}, \dots, a_{nk}) \xrightarrow{\text{a.s.}} f(\alpha_1, \dots, \alpha_k) = \mu_k, \quad k = 2, 3, \dots,$$

这就证明了结论.

(3) 设 $g(\theta)$ 有式 (3.2.14) 的形式, 其矩估计为式 (3.2.15), 关于此类矩估计的强相合性有下述定理.

定理 3.2.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 F 中抽取的简单随机样本, 待估函数 $g(\theta) = G(\alpha_1, \dots, \alpha_k, \mu_2, \dots, \mu_s)$, 其矩估计量为 $\hat{g}_n(\mathbf{X}) = G(a_{n1}, \dots, a_{nk}, m_{n2}, \dots, m_{ns})$, 且 G 为其变元的连续函数, 则 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 之强相合估计.

证 由 $a_{ni} \xrightarrow{\text{a.s.}} \alpha_i$, $i = 1, 2, \dots, k$, 利用引理 3.2.1 可知 $m_{nj} \xrightarrow{\text{a.s.}} \mu_j$, $j = 2, 3, \dots, s$. 再由 G 是其变元的连续函数, 再利用引理 3.2.1, 立即可得 $\hat{g}_n(\mathbf{X}) \xrightarrow{\text{a.s.}} g(\theta)$.

由这一定理可得出一些常见估计的相合性. 例如, 正态总体 $N(a, \sigma^2)$ 中, 样本均值 \bar{X} 和样本方差 S^2 分别是 a 和 σ^2 的强相合估计. 也不难证明 S^2 是 σ^2 的均方相合估计(留作习题). 其实对任何 $r > 0$, S^2 是 σ^2 的 r 阶矩相合估计. 例 3.2.5 中定义的偏度、峰度和变异系数的矩估计都是强相合的.

例 3.2.9 证明例 3.2.3 中 θ_1 和 θ_2 的矩估计量分别是 θ_1 和 θ_2 的强相合估计.

证 在例 3.2.3 中已求出 θ_1 和 θ_2 的矩估计量分别为

$$\hat{\theta}_1 = \hat{\theta}_1(X_1, \dots, X_n) = \bar{X} - \sqrt{3}S_n,$$

$$\hat{\theta}_2 = \hat{\theta}_2(X_1, \dots, X_n) = \bar{X} + \sqrt{3}S_n.$$

由强大数定律可知 \bar{X} 和 $S_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / n$ 分别为 $\alpha_1 = E(X)$ 和 $\mu_2 = D(X)$ 的强相合估计, 又 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别是 \bar{X} 和 S_n^2 的连续函数, 故由引理 3.2.1 可知

$$\hat{\theta}_1 \xrightarrow{\text{a.s.}} \alpha_1 - \sqrt{3\mu_2} = \theta_1,$$

$$\hat{\theta}_2 \xrightarrow{\text{a.s.}} \alpha_1 + \sqrt{3\mu_2} = \theta_2.$$

证毕.

*3.2.5 矩估计的渐近正态性

本段将在很一般的条件下, 给出矩估计是相合渐近正态估计. 下面首先给出定义.

定义 3.2.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $\{f(x, \theta), \theta \in \Theta\}$ 中抽取的简单样本, $\hat{g}_n(\mathbf{X})$ 是 $g(\theta)$ 的矩估计量. 若存在与样本大小 n 有关的, 定义于参数空间 Θ 上的函数 $A_n(\theta)$ 和 $B_n(\theta)$, 其中 $B_n(\theta)$ 在 Θ 上处处大于 0, 使当 $n \rightarrow \infty$ 时,

$$\frac{\hat{g}_n(\mathbf{X}) - A_n(\theta)}{B_n(\theta)} \xrightarrow{\mathcal{L}} N(0, 1),$$

且 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的弱相合估计, 则称 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的相合渐近正态估计 (consistent asymptotic normal estimation, CAN 估计).

就是说 CAN 估计是既相合, 其分布又渐近服从正态分布的那种估计. 本段提出两个重要结果, 即在很一般的条件下, 矩估计为 CAN 估计.

(1) 设样本 X_1, \dots, X_n 为从总体 $\{f(x, \theta), \theta \in \Theta\}$ 中抽取的简单样本, $g(\theta)$ 是定义在参数空间 Θ 上的实函数, 它可以表示为形式: $g(\theta) = G(\alpha_1, \dots, \alpha_k)$ (若 G 是 $\alpha_1, \dots, \alpha_k, \mu_2, \dots, \mu_s$ 的函数, 不妨令 $s \leq k$, 可将 μ_j 用 $\alpha_1, \alpha_2, \dots, \alpha_s$ 表出, 则 G 仍可表示为 $\alpha_1, \dots, \alpha_k$ 的函数), 而 $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n) = G(a_{n1}, \dots, a_{nk})$ 为 $g(\theta)$ 的矩估计. 再设总体的 $2k$ 阶原点矩存在, 且 G 对其各变元的一阶偏导数存在、连续, 令

$$\begin{aligned} b_{ij} &= \alpha_{i+j} - \alpha_i \alpha_j, \quad i, j = 1, 2, \dots, k; \quad \mathbf{B} = (b_{ij}) \text{ 为 } k \times k \text{ 的方阵,} \\ d_i &= \frac{\partial G(\alpha_1, \dots, \alpha_k)}{\partial \alpha_i}, \quad i = 1, 2, \dots, k; \quad \mathbf{d} = (d_1, \dots, d_k)', \\ b^2 &= \mathbf{d}' \mathbf{B} \mathbf{d}. \end{aligned}$$

定理 3.2.2 在上述记号和条件下, $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的 CAN 估计, 即 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta)$ 的弱相合估计, 且有

$$\sqrt{n}(\hat{g}_n(\mathbf{X}) - G(\alpha_1, \dots, \alpha_k)) \xrightarrow{\mathcal{L}} N(0, b^2), \quad n \rightarrow \infty.$$

证 由定理 3.2.1 可知 $\hat{g}_n(\mathbf{X})$ 为 $g(\theta) = G(\alpha_1, \dots, \alpha_k)$ 的强相合估计, 显然它也是弱相合估计, 其渐近正态性的证明见文献 [2] 第 82 页定理 2.6 的证明.

(2) 在一些情况下, $g(\theta)$ 可表示为一两个中心矩的函数, 还可能包含总体均值 α_1 , $g(\theta)$ 的表达式较简单. 若将中心矩用原点矩表出, $g(\theta)$ 的表达式则显得复杂, 因此有必要给出这种情形下渐近正态性的结果. 一般地, 将 $g(\theta)$ 表达成如下形式:

$$g(\theta) = H(\alpha_1, \mu_{t_2}, \dots, \mu_{t_r}), \quad (3.2.16)$$

其矩估计量为 $H(\bar{X}, m_{nt_2}, \dots, m_{nt_r})$, 使用与定理 3.2.2 基本相同的证明方法, 可得如下结果.

定理 3.2.3 设式 (3.2.16) 中的函数 H 在点 $(\alpha_1, \mu_{t_2}, \dots, \mu_{t_r})$ 的邻域内有一阶偏导数, 且此偏导数在点 $(\alpha_1, \mu_{t_2}, \dots, \mu_{t_r})$ 处连续, 则有

$$\sqrt{n} \left(H(\bar{X}, m_{nt_2}, \dots, m_{nt_r}) - H(\alpha_1, \mu_{t_2}, \dots, \mu_{t_r}) \right) \xrightarrow{\mathcal{L}} N(0, b^2). \quad (3.2.17)$$

此处

$$b^2 = \sum_{i=1}^r \sum_{j=1}^r \sigma_{ij} H_i H_j, \quad (3.2.18)$$

其中

$$H_1 = \frac{\partial H}{\partial \alpha_1}, \quad H_i = \frac{\partial H}{\partial \mu_{t_i}}, \quad i = 2, 3, \dots, r;$$

$$\begin{aligned}\sigma_{11} &= \mu_2, \quad \sigma_{1i} = \sigma_{i1} = \mu_{t_i+1} - t_i \mu_{t_i-1} \mu_2, \quad i = 2, \dots, r; \\ \sigma_{ij} &= \mu_{t_i+t_j} - t_i \mu_{t_i-1} \mu_{t_j+1} - t_j \mu_{t_i+1} \mu_{t_j-1} - \mu_{t_i} \mu_{t_j} \\ &\quad + t_i t_j \mu_2 \mu_{t_i-1} \mu_{t_j-1}, \quad i, j = 2, 3, \dots, r.\end{aligned}$$

如果 $g(\theta)$ 有 $H(\mu_{t_2}, \dots, \mu_{t_r})$ 的形状, 即与 α_1 无关, 则式 (3.2.17) 仍成立, 只需把式 (3.2.18) 所确定的 b^2 改为 $\sum_{i=2}^r \sum_{j=2}^r \sigma_{ij} H_i H_j$ 即可.

例 3.2.10 继续考虑例 3.2.5. 被估计的量 $g(\theta)$ 为偏度 β_1 , 峰度 β_2 和变异系数 ν , 其矩估计量由例 3.2.5 可知为

$$\hat{\beta}_1 = \frac{m_{n3}}{m_{n2}^{3/2}}, \quad \hat{\beta}_2 = \frac{m_{n4}}{m_{n2}^2} - 3, \quad \hat{\nu} = \sqrt{m_{n2}} / \bar{X}.$$

讨论它们的渐近正态性.

解 按式 (3.2.17) 和式 (3.2.18), 对这三个矩估计量分别算得 b^2 之值为

$$b^2(\beta_1) = 6, \quad b^2(\beta_2) = 24, \quad b^2(\nu) = \nu^2/2 + \nu^4.$$

于是根据定理 3.2.3, 有

$$\begin{aligned}\sqrt{n}(\hat{\beta}_1 - \beta_1) &\xrightarrow{\mathcal{L}} N(0, 6), \\ \sqrt{n}(\hat{\beta}_2 - \beta_2) &\xrightarrow{\mathcal{L}} N(0, 24), \\ \sqrt{n}(\hat{\nu} - \nu) &\xrightarrow{\mathcal{L}} N(0, \nu^2/2 + \nu^4).\end{aligned}$$

值得注意的是 $\hat{\beta}_1, \hat{\beta}_2$ 的极限分布的方差与被估计的参数值无关, 这点对 β_1, β_2 的大样本推断有用.

3.3 极大似然估计

3.3.1 引言及定义

极大似然法是在参数分布族场合下常用的参数估计方法. 设有一参数分布族 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$, 其中 Θ 为参数空间. 令 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 \mathcal{F} 中抽取的简单随机样本, $f(\mathbf{x}, \theta) = f(x_1, \dots, x_n, \theta)$ 为样本 \mathbf{X} 的概率函数, 即当总体分布为连续型时, $f(\mathbf{x}, \theta)$ 表示样本 \mathbf{X} 的密度函数; 当总体分布为离散形时, $f(\mathbf{x}, \theta)$ 为样本 \mathbf{X} 的概率分布, 即 $f(\mathbf{x}, \theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n)$. 首先引入似然函数的概念.

定义 3.3.1 设 $f(\mathbf{x}, \theta) = f(x_1, \dots, x_n, \theta)$ 为样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的概率函数. 当 \mathbf{x} 固定时, 把 $f(\mathbf{x}, \theta)$ 看成 θ 的函数, 称为似然函数(likelihood function),

记为

$$L(\theta, \mathbf{x}) = f(\mathbf{x}, \theta), \quad \theta \in \Theta, \quad \mathbf{x} \in \mathcal{X}, \quad (3.3.1)$$

其中 Θ 为参数空间, \mathcal{X} 为样本空间. 称 $\log L(\theta, \mathbf{x})$ 为对数似然函数, 记为 $l(\theta, \mathbf{x})$.

注 3.3.1 似然函数和概率函数是同一表达式 (3.3.1), 但表示两种不同含意. 当把 θ 固定, 将其看成定义在样本空间 \mathcal{X} 上的函数时, 称为概率函数; 当把 \mathbf{x} 固定, 将其看成定义在参数空间 Θ 上的函数时, 称为似然函数. 这是两个不同的概念.

为解释极大似然原理, 考虑下列一个简单的实例.

例 3.3.1 设罐子里有许多黑球和红球. 假定已知它们的比例是 1 : 3, 但不知道是黑球多还是红球多. 也就是说抽出一个黑球的概率或者是 $1/4$ 或者是 $3/4$. 如果有放回地从罐子中抽 n 个球, 要根据抽样数据, 说明抽到黑球的概率是 $1/4$, 还是 $3/4$.

解 将此问题用统计模型来表述. 令 X_i 表示第 i 次抽球的结果, 即

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 次抽出为黑球,} \\ 0, & \text{否则.} \end{cases}$$

记每次抽样中抽到黑球的概率为 θ , 此处 θ 只取可能的两个值 $\theta_1 = 1/4$ 和 $\theta_2 = 3/4$ 之一. 记 $X = \sum_{i=1}^n X_i$, 则 $X \sim b(n, \theta)$. 即样本分布族 $\mathcal{F} = \{f(x, \theta_1), f(x, \theta_2)\}$, 其中 $f(x, \theta_1)$ 为 $b(n, \theta_1)$, $f(x, \theta_2)$ 为 $b(n, \theta_2)$. 要根据抽样结果对 θ 作出估计, 即 θ 取值为 $1/4$ 还是 $3/4$? 或说样本来自总体 $f(x, \theta_1)$ 还是 $f(x, \theta_2)$?

显然, 当样本 X 给定时, 似然函数为

$$L(\theta, x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

为简单计, 取 $n = 3$. 当 $x = 0, 1, 2, 3$ 时似然函数取值如表 3.3.1 所示.

表 3.3.1

x	0	1	2	3
$L(\theta_1, x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$
$L(\theta_2, x)$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$

由表 3.3.1 可见:

当 $x = 0, 1$ 时, $L(\theta_1, x) > L(\theta_2, x)$;

当 $x = 2, 3$ 时, $L(\theta_2, x) > L(\theta_1, x)$.

因此得出结论: 当样本观察值 $x = \sum_{i=1}^3 x_i$ 取值为 0, 1 时认为样本来自总体 $f(x, \theta_1)$,

即取参数 θ 的估计值为 $\hat{\theta}_1 = 1/4$; 当 $x = 2, 3$ 时认为样本来自总体 $f(x, \theta_2)$, 即取 θ 的估计值为 $\hat{\theta}_2 = 3/4$.

将此例模型化如下: 若样本 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$ 中抽取的简单随机样本, 其中 $\Theta = \{\theta_1, \theta_2\}$. 此即等价地说分布族 \mathcal{F} 中只有两个总体 $f(x, \theta_1), f(x, \theta_2)$. 一旦获得了样本 \mathbf{x} , 如何用极大似然方法求出真参数 θ 的估计值呢? 上例表明, 若

$$L(\theta_1, \mathbf{x}) > L(\theta_2, \mathbf{x}),$$

则倾向于认为样本 \mathbf{X} 来自总体 $f(x, \theta_1)$ (即真参数 θ 为 θ_1) 的理由比认为样本 \mathbf{X} 来自总体 $f(x, \theta_2)$ (即真参数 θ 为 θ_2) 的理由更充分些. 或者说, 真参数 θ 为 θ_1 的“似然性”更大些. 这样, 自然把“似然性”最大(即看起来最像)的那个值作为真参数 θ 的估计值. 这正是“极大似然”一词的由来.

更一般地, 若样本 \mathbf{X} 的分布族 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$, 参数空间 Θ 为 R_1 的有限子集或无限子集. 当样本 \mathbf{x} 给定时, 若 $\hat{\theta}^*$ 使似然函数 $L(\hat{\theta}^*, \mathbf{x})$ 为似然函数的集合 $\{L(\theta, \mathbf{x}), \text{一切 } \theta \in \Theta\}$ 中最大者, 即参数 θ 的真值为 $\hat{\theta}^*$ 的“似然性”比参数空间 Θ 中任何其他参数值的“似然性”都大, 则取“似然性”最大的 $\hat{\theta}^*$ 作为 θ 的估计值, 这一方法得到的参数 θ 的估计, 称为“极大似然估计”. 将这一直观想法用数学语言来描述, 得到如下定义.

定义 3.3.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从参数分布族 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$ 中抽取的简单随机样本, $L(\theta, \mathbf{x})$ 是似然函数, 若存在统计量 $\hat{\theta}^* = \hat{\theta}^*(\mathbf{X})$, 满足条件

$$L(\hat{\theta}^*, \mathbf{x}) = \sup_{\theta \in \Theta} L(\theta, \mathbf{x}), \quad \mathbf{x} \in \mathcal{X}, \tag{3.3.2}$$

或等价地使得

$$l(\hat{\theta}^*, \mathbf{x}) = \sup_{\theta \in \Theta} l(\theta, \mathbf{x}), \quad \mathbf{x} \in \mathcal{X}, \tag{3.3.3}$$

则称 $\hat{\theta}^*(\mathbf{X})$ 为 θ 的极大似然估计 (maximum likelihood estimation, MLE). 若待估函数是 $g(\theta)$, 则定义 $g(\hat{\theta}^*(\mathbf{X}))$ 为 $g(\theta)$ 的 MLE.

极大似然估计是 R.A. Fisher 在 1912 年的一项工作中提出来的. 在正态分布这个特殊情况下, 这方法可追溯到 Gauss 在 19 世纪初关于最小二乘法的工作. Fisher 后来在 1922 年工作, 尤其在 1925 年发表的 *Theory of Statistical Estimation* 一文中对这一估计作了许多研究. 因此这个方法应归功于 R.A. Fisher.

3.3.2 极大似然估计的求法及例

获得参数 θ 的极大似然估计有下列两种方法.

1. 用微积分中求极值的方法

若似然函数 $L(\theta, \mathbf{x})$ 是 θ 的连续可微函数, 则可用微积分中求极值点的方法去求 θ 的 MLE, 即找使 $L(\theta, \mathbf{x})$ 达到最大时 θ 的值. 由于 $L(\theta, \mathbf{x})$ 和 $\log L(\theta, \mathbf{x}) = l(\theta, \mathbf{x})$ 具有相同的极值点, 可用 $l(\theta, \mathbf{x})$ 来代替 $L(\theta, \mathbf{x})$. 因为 $L(\theta, \mathbf{x})$ 一般是若干个函数的乘积, $l(\theta, \mathbf{x})$ 为若干个函数之和而较易于处理.

设 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ 为参数向量 (特别当 $k=1$, $\boldsymbol{\theta}$ 为参数). 若 $l(\theta, \mathbf{x})$ 的极大值在参数空间 Θ 的内点处 (而非边界点) 达到, 则此点必为似然方程组

$$\frac{\partial l(\theta, \mathbf{x})}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k$$

的解.

因此求极大似然估计首先求似然方程组的解 $\hat{\boldsymbol{\theta}}^*$. 但此解是否一定是 $\boldsymbol{\theta}$ 的 MLE 呢? $\hat{\boldsymbol{\theta}}^*$ 满足似然方程, 只是 MLE 的必要条件, 而非充分条件. 一般只有满足下列条件: ① 似然函数的极大值在参数空间 Θ 内部达到; ② 似然方程只有唯一解, 则似然方程之解 $\hat{\boldsymbol{\theta}}^*$ 必为 $\boldsymbol{\theta}$ 的 MLE.

因此求出似然方程 (或方程组) 的解后, 要验证它为 $\boldsymbol{\theta}$ 的 MLE, 有时并非易事. 但对样本分布族是指数族的场合, 有非常满意的结果, 叙述如下:

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从某总体中抽取的简单随机样本, X_1 的分布为指数族, 即

$$f(x, \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp \left\{ \sum_{i=1}^k \theta_i T_i(x) \right\} h(x), \quad \boldsymbol{\theta} \in \Theta,$$

其中 Θ 为自然参数空间, Θ_0 为 Θ 之内点集, 这时 \mathbf{X} 的联合密度为

$$L(\boldsymbol{\theta}, \mathbf{x}) = C^n(\boldsymbol{\theta}) \exp \left\{ \sum_{i=1}^k \theta_i \sum_{j=1}^n T_i(x_j) \right\} h(\mathbf{x}),$$

此处 $h(\mathbf{x}) = \prod_{i=1}^n h(x_i)$. 对上式取对数得

$$l(\boldsymbol{\theta}, \mathbf{x}) = \log L(\boldsymbol{\theta}, \mathbf{x}) = n \log C(\boldsymbol{\theta}) + \sum_{i=1}^k \theta_i \sum_{j=1}^n T_i(x_j) + \log h(\mathbf{x}).$$

定理 3.3.1 若对任何样本 $\mathbf{X} = (X_1, \dots, X_n)$, 方程组

$$\frac{n}{C(\boldsymbol{\theta})} \frac{\partial C(\boldsymbol{\theta})}{\partial \theta_i} = - \sum_{j=1}^n T_i(X_j), \quad i = 1, 2, \dots, k$$

在 Θ_0 内有解, 则解必唯一且为 $\boldsymbol{\theta}$ 的 MLE.

证 见文献 [1] 第 199 页定理 2.6.3.

因此若样本分布为指数族, 只要似然方程解属于自然参数空间的内点集 (这点很容易验证), 则其解必为 θ 的 MLE. 常见的分布族, 如二项分布族、Poisson 分布族、正态分布族、Gamma 分布族等都是指数族, 定理 3.3.1 的条件皆成立. 因此似然方程的解, 就是有关参数的 MLE.

2. 从定义出发

当似然函数 $L(\theta, \mathbf{x})$ 对参数 θ 不可微, 甚至不连续时, 似然方程一般没有意义, 不能采用上述方法, 必须直接从定义 3.3.2 出发去求参数 θ 的极大似然估计.

下面分别用上述两种方法考察一些例子.

例 3.3.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从两点分布族 $\{b(1, p) : 0 < p < 1\}$ 中抽取的简单样本, 求 p 和 $g(p) = p(1-p)$ 的 MLE.

解 似然函数为

$$L(p, \mathbf{x}) = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i},$$

故有

$$l(p, \mathbf{x}) = \log L(p, \mathbf{x}) = \left(\sum_{i=1}^n x_i \right) \log p + \left(n - \sum_{i=1}^n x_i \right) \log(1-p).$$

对数似然方程为

$$\frac{\partial l(p, \mathbf{x})}{\partial p} = \frac{1}{p} \sum_{i=1}^n x_i - \frac{1}{1-p} \left(n - \sum_{i=1}^n x_i \right) = 0,$$

解得

$$\hat{p}^* = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

由于两点分布族为指数族, 当 $\bar{X} \neq 0, 1$ 时, $\hat{p}^* = \bar{X}$ 属于自然参数空间 $\Theta^* = (0, 1)$ 的内点集, 故 \hat{p}^* 为 p 的 MLE. 它与例 3.2.2 中给得出的 p 的矩估计量相同. 按定义可知 $g(p) = p(1-p)$ 的 MLE 为

$$\hat{g}^*(\mathbf{X}) = \hat{p}^*(1 - \hat{p}^*) = \bar{X}(1 - \bar{X}).$$

注 3.3.2 当 $0 < \bar{X} < 1$ 时, 易知 $\hat{p}^* = \bar{X}$ 为 p 的唯一的 MLE. 而当 $\sum_{i=1}^n X_i = 0$, 或 n 时 $\bar{X} = 0, 1$ 不在 $\Theta = (0, 1)$ 内, 严格意义上讲, 此时 p 的 MLE 不存在. 为了克服这一缺陷, 对 MLE 的定义加以补充: 当 $\hat{p}^* \notin \Theta$ 时, 若存在一列 $\{\hat{p}_n^*\}$ 使得

$\lim_{n \rightarrow \infty} \hat{p}_n^* = \hat{p}^*$, 且 $\hat{p}_n^* \in \Theta$, $n = 1, 2, \dots$, 则 \hat{p}^* 也称为 p 的 MLE. 在给出上述补充定义后, 本题中当 $\bar{X} = 0, 1$ 时, 可认为 $\hat{p}^* = \bar{X}$ 是 p 的 MLE.

例 3.3.3 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从 Poisson 分布族 $\{P(\lambda) : \lambda > 0\}$ 中抽取的简单样本, 求 λ 和 $g(\lambda) = e^{-\lambda}$ 的 MLE.

解 似然函数, 即样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的分布为

$$L(\lambda, \mathbf{x}) = P(X_1 = x_1, \dots, X_n = x_n) = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{x_1! \cdots x_n!}, \quad \lambda > 0.$$

故对数似然函数为

$$l(\lambda, \mathbf{x}) = \left(\sum_{i=1}^n x_i \right) \log \lambda - n\lambda - \sum_{i=1}^n \log x_i!.$$

由对数似然方程

$$\frac{\partial l(\lambda, \mathbf{x})}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^n x_i - n = 0,$$

解得

$$\hat{\lambda}^* = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

由于 Poisson 分布族是指数族, 当 $\bar{X} \neq 0$ 时, $\hat{\lambda}^* = \bar{X}$ 属于自然参数空间 $\Theta^* = (0, \infty)$ 的内点集, 故 $\hat{\lambda}^*$ 为 λ 的 MLE, 它与 λ 的矩估计量相同. 当 $\bar{X} = 0$ 时, 用类似于注 3.3.2 的方法处理.

又由定义可知 $g(\lambda) = e^{-\lambda}$ 的 MLE 为

$$\hat{g}^*(\mathbf{X}) = e^{-\bar{X}}.$$

例 3.3.4 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态分布族 $\{N(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$ 中抽取的简单样本. 记 $\theta = (\mu, \sigma^2)$, 求 μ , σ^2 和 $g(\theta) = \mu/\sigma^2$ 的 MLE.

解 样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的分布为

$$f(\mathbf{x}, \theta) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\}.$$

对数似然函数为

$$l(\theta, \mathbf{x}) = \log f(\mathbf{x}, \theta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

由对数似然方程组

$$\frac{\partial l(\theta, \mathbf{x})}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0,$$

$$\frac{\partial l(\theta, \mathbf{x})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0,$$

解得

$$\hat{\mu}^* = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}, \quad \hat{\sigma}_*^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = S_n^2.$$

由于正态分布族为指数族, 且 $\hat{\mu}^* = \bar{X}$, $\hat{\sigma}_*^2 = S_n^2$ 属于自然参数空间 $\Theta^* = \{(\mu, \sigma^2) : \mu \in R, \sigma^2 > 0\}$ 的内点集, 因此 $\hat{\mu}^*$ 和 $\hat{\sigma}_*^2$ 分别是 μ 和 σ^2 的 MLE, 它们也分别是 μ 和 σ^2 的矩估计量. 前者是 μ 的无偏估计, 后者不是 σ^2 的无偏估计. 可见极大似然估计不一定具有无偏性.

又由定义可知 $g(\theta) = \mu/\sigma^2$ 的 MLE 为

$$\hat{g}^*(\mathbf{X}) = \frac{\bar{X}}{S_n^2}.$$

例 3.3.5 设元件的寿命 X 服从下列指数分布 $\text{Exp}(\lambda)$, 其密度函数为

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

设 X_1, \dots, X_n 分别表示接受试验的 n 个元件寿命. 由于受时间的限制, 试验实际上只进行到有 r ($r \leq n$) 个元件失效时就停止了, 以 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(r)}$ 记这 r 个元件的寿命, 即只观察到了样本 X_1, \dots, X_n 前 r 个次序统计量 $X_{(1)}, X_{(2)}, \dots, X_{(r)}$. 基于这前 r 个次序统计量, 求 λ 和 $g(\lambda) = 1/\lambda$ 的 MLE.

解 为叙述方便, 记 $t_i = x_{(i)}$, $i = 1, 2, \dots, n$, 则有 $t_1 \leq t_2 \leq \dots \leq t_n$. 设 $F(x)$ 为 X 的分布函数, 易知 $F(x) = 1 - e^{-\lambda x}$. 下面利用第 2 章习题 7 的结果和 2.3.2 节所介绍的方法求 $(X_{(1)}, X_{(2)}, \dots, X_{(r)})$ 的联合密度如下:

$$\begin{aligned} p(t_1, \dots, t_r) &= \int_{-t_r < t_{r+1} < \dots < t_n < \infty} \cdots \int_{-t_r < t_{r+1} < \dots < t_n < \infty} n! f(t_1, \lambda) \cdots f(t_n, \lambda) dt_{r+1} \cdots dt_n \\ &= \frac{n!}{(n-r)!} f(t_1, \lambda) \cdots f(t_r, \lambda) [1 - F(t_r)]^{n-r} \\ &= \frac{n!}{(n-r)!} \lambda^r \exp \left\{ -\lambda \left(\sum_{i=1}^r t_i + (n-r)t_r \right) \right\}. \end{aligned}$$

记 $T = \sum_{i=1}^r t_i + (n-r)t_r$, 故 λ 的似然函数为

$$L(\lambda, t_1, \dots, t_r) = c \cdot \lambda^r \exp\{-\lambda T\},$$

其中 $c = n!/(n-r)!$. 对数似然函数为

$$l(\lambda, t_1, \dots, t_r) = \log c + r \log \lambda - \lambda T.$$

对 λ 求导, 得似然方程为

$$\frac{\partial l}{\partial \lambda} = \frac{r}{\lambda} - T = 0,$$

解得

$$\hat{\lambda}^* = \frac{r}{T} = r \left[\sum_{i=1}^r X_{(i)} + (n-r)X_{(r)} \right]^{-1}.$$

似然函数 $L(\lambda, x_{(1)}, \dots, x_{(r)})$ 是指数族的形式, 且 $\hat{\lambda}^*$ 属于自然参数空间 $\Theta^* = \{\lambda : \lambda > 0\} = (0, \infty)$ 的内点集, 故 $\hat{\lambda}^*$ 为 λ 的 MLE. 由定义可知 $g(\lambda) = 1/\lambda$ 的 MLE 为

$$\hat{g}^*(X_{(1)}, \dots, X_{(r)}) = \frac{T}{r} = \frac{1}{r} \left[\sum_{i=1}^r X_{(i)} + (n-r)X_{(r)} \right].$$

本例中所述产品寿命试验进行到第 r 个产品失效时就终止, 这种试验称为定数截尾试验. 另一种方式是先定下一个时间 $T > 0$, 当试验进行 T 时试验就终止, 这种试验称为定时截尾试验.

例 3.3.6 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从均匀分布族 $\{U(0, \theta) : \theta > 0\}$ 中抽取的简单样本.

- (1) 求 θ 的 MLE $\hat{\theta}^*$;
- (2) 说明 $\hat{\theta}^*$ 是否为 θ 的无偏估计. 若不然, 作适当修正获得 θ 的无偏估计 $\hat{\theta}_1^*$;
- (3) 试将 $\hat{\theta}^*$ 与 θ 的矩估计 $\hat{\theta}_1$ 进行比较, 看哪一个有效?
- (4) 证明 θ 的极大似然估计 $\hat{\theta}^*$ 是 θ 的弱相合估计.

解 (1) 样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的联合密度为

$$f(\mathbf{x}, \theta) = \begin{cases} \frac{1}{\theta^n}, & 0 < x_1, \dots, x_n < \theta, \\ 0, & \text{其他.} \end{cases} \quad (3.3.4)$$

因为均匀分布 $U(0, \theta)$ 的支撑集依赖于 θ , 似然函数 $L(\theta, \mathbf{x}) = f(\mathbf{x}, \theta)$ 作为 θ 的函数不是连续函数, 所以不能用对似然函数求微商的办法去求 θ 的 MLE. 只能从 MLE 的定义出发来讨论.

为使 $L(\theta, \mathbf{x})$ 达到极大, 由式 (3.3.4) 可见, 应使分母上的 θ 尽可能地小, 但 θ 又不能太小以致 L 为 0. 这个界限就在

$$\hat{\theta}^* = \max(X_1, \dots, X_n) = X_{(n)}.$$

故当 $\theta > \hat{\theta}^*$ 时 $L > 0$ 且为 θ^{-n} , 当 $\theta < \hat{\theta}^*$ 时 $L = 0$. 因此 $\hat{\theta}^*$ 是唯一使 L 达到最大的 θ 值, 即 θ 的 MLE.

(2) 为求 $E(X_{(n)})$, 就要算出 $T = X_{(n)}$ 的密度函数, 易求 T 的密度函数

$$g(t, \theta) = \begin{cases} \frac{nt^{n-1}}{\theta^n}, & 0 \leq t \leq \theta, \\ 0, & \text{其他,} \end{cases} \quad (3.3.5)$$

故有

$$E(\hat{\theta}^*) = E(T) = \int_0^\theta \frac{nt^n}{\theta^n} dt = \frac{n}{n+1}\theta,$$

故 $\hat{\theta}^* = X_{(n)}$ 不是 θ 的无偏估计. 显见

$$\hat{\theta}_1^* = \frac{n+1}{n} X_{(n)}$$

为 θ 的无偏估计.

(3) θ 的矩估计 $\hat{\theta}_1 = 2\bar{X}$ 是 θ 的无偏估计. 由于

$$D(\hat{\theta}_1^*) = \frac{\theta^2}{n(n+2)}, \quad D(\hat{\theta}_1) = \frac{\theta^2}{3n},$$

所以在 $n \geq 2$ 时 $\hat{\theta}_1^*$ 比 $\hat{\theta}_1$ 有效. 在 $n=1$ 时 $\hat{\theta}_1^* = \hat{\theta}_1$, 即这两个估计是相同的.

(4) 已知 T 的密度函数由式 (3.3.5) 给出, 故对任给的 $\varepsilon > 0$, 有

$$\begin{aligned} P(|\hat{\theta}^* - \theta| \geq \varepsilon) \\ = 1 - P(|\hat{\theta}^* - \theta| < \varepsilon) = 1 - P(\theta - \varepsilon < T < \theta + \varepsilon) \\ = 1 - \int_{\theta-\varepsilon}^\theta \frac{nt^{n-1}}{\theta^n} dt = 1 - \frac{1}{\theta^n} [\theta^n - (\theta - \varepsilon)^n] = \left(1 - \frac{\varepsilon}{\theta}\right)^n. \end{aligned}$$

因此有

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}^* - \theta| \geq \varepsilon) = \lim_{n \rightarrow \infty} \left(1 - \frac{\varepsilon}{\theta}\right)^n = 0,$$

故知 $\hat{\theta}^* = X_{(n)}$ 为 θ 的弱相合估计.

例 3.3.7 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从均匀分布族 $\{U(\theta, \theta+1) : -\infty < \theta < +\infty\}$ 中抽取的简单样本, 求 θ 的 MLE.

解 给定样本 x 时, θ 的似然函数为

$$L(\theta, \mathbf{x}) = \begin{cases} 1, & \theta < x_{(1)} \leq x_{(n)} < \theta + 1 \\ 0, & \text{其他} \end{cases} = \begin{cases} 1, & x_{(n)} - 1 < \theta < x_{(1)}, \\ 0, & \text{其他.} \end{cases}$$

这时, 似然函数只取 1 和 0 两个值, 只要 $x_{(n)} - 1 < \theta < x_{(1)}$ 都可使 L 达到极大. 故 θ 的 MLE 不止一个, 如

$$\hat{\theta}_1^*(\mathbf{x}) = X_{(1)}, \quad \hat{\theta}_2^*(\mathbf{x}) = X_{(n)} - 1$$

都是 θ 的 MLE. 事实上对任给的 $0 \leq \lambda \leq 1$,

$$\hat{\theta}^*(\mathbf{x}) = \lambda \hat{\theta}_1^*(\mathbf{x}) + (1 - \lambda) \hat{\theta}_2^*(\mathbf{x}) = \lambda X_{(1)} + (1 - \lambda) (X_{(n)} - 1)$$

都是 θ 的 MLE, 故知 θ 的 MLE 有无穷多个.

例 3.3.8 设 k 个事件 A_1, A_2, \dots, A_k 构成完备事件群, 事件 A_i 发生的概率为 $0 < p_i < 1$, $i = 1, \dots, k$ 且 $\sum_{i=1}^k p_i = 1$. 将试验独立重复 n 次, 以 X_i 记 A_i 发生的次数, $i = 1, 2, \dots, k$, 则 $\mathbf{X} = (X_1, \dots, X_k)$ 服从多项分布 $M(n, p_1, \dots, p_k)$. 求 p_1, \dots, p_k 的 MLE.

解 记 $\mathbf{p} = (p_1, \dots, p_k)$. 给定样本 \mathbf{x} 时, \mathbf{p} 的似然函数为

$$L(\mathbf{p}, \mathbf{x}) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{x_k}.$$

对数似然函数为

$$l(\mathbf{p}, \mathbf{x}) = \log n! - \sum_{i=1}^k \log x_i! + \sum_{i=1}^{k-1} x_i \log p_i + x_k \log \left(1 - \sum_{i=1}^{k-1} p_i\right). \quad (3.3.6)$$

对 p_i 求偏导数, 得似然方程组

$$\frac{\partial l(\mathbf{p}, \mathbf{x})}{\partial p_i} = \frac{x_i}{p_i} - \frac{x_k}{p_k} = 0, \quad i = 1, 2, \dots, k-1.$$

若令

$$\frac{x_1}{p_1} = \frac{x_2}{p_2} = \cdots = \frac{x_k}{p_k} = \lambda,$$

则有

$$x_i = \lambda p_i, \quad i = 1, 2, \dots, k. \quad (3.3.7)$$

将这 k 个等式两边分别相加得

$$n = \sum_{i=1}^k x_i = \lambda \sum_{i=1}^k p_i = \lambda.$$

由于多项分布族属于指数族, 因此由定理 3.3.1 和式 (3.3.7) 可知 p_i 的 MLE 如下:

$$\hat{p}_i^* = \frac{X_i}{n}, \quad i = 1, 2, \dots, k.$$

在有些问题中, p_1, \dots, p_k 都是另一些参数 $\theta_1, \dots, \theta_r$ ($r \leq k$) 的函数, 这时去掉式 (3.3.6) 中与 $\theta_1, \dots, \theta_r$ 无关的部分后, 得对数似然函数为

$$l(\theta_1, \dots, \theta_r; \mathbf{x}) = \sum_{i=1}^k x_i \log p_i(\theta_1, \dots, \theta_r),$$

对 θ_i 求偏导数得似然方程组

$$\frac{\partial l}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, r. \quad (3.3.8)$$

若似然方程之解 $\hat{\theta}_1^*, \dots, \hat{\theta}_r^*$ 为 $\theta_1, \dots, \theta_r$ 的 MLE, 则 p_1, \dots, p_k 的 MLE 分别是 $\hat{p}_1(\hat{\theta}_1^*, \dots, \hat{\theta}_r^*), \dots, \hat{p}_k(\hat{\theta}_1^*, \dots, \hat{\theta}_r^*)$. 但有时似然方程组 (3.3.8) 无显式解, 就只能用数值方法.

3.3.3 极大似然估计的性质

1. 极大似然估计的无偏性

在前面的例 3.3.4 中已指出极大似然估计不一定是无偏的. 因此极大似然估计可以是无偏的, 但也可以是有偏的, 要视具体情况而定.

2. 极大似然估计与充分统计量

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为自总体 $\{f(x, \theta), \theta \in \Theta\}$ 中抽出的简单随机样本, $T = T(X_1, \dots, X_n)$ 是参数 θ 的充分统计量, 如果 θ 的极大似然估计存在, 则它必为 T 的函数.

由因子分解定理可知样本 \mathbf{X} 的概率函数, 即似然函数可表示为

$$L(\theta, \mathbf{x}) = \prod_{i=1}^n f(x_i, \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x}).$$

若 θ 的 MLE 存在, 记为 $\hat{\theta}^*$, 则

$$L(\hat{\theta}^*, \mathbf{x}) = \sup_{\theta \in \Theta} L(\theta, \mathbf{x}) \iff g(T(\mathbf{x}), \hat{\theta}^*) = \sup_{\theta \in \Theta} g(T(\mathbf{x}), \theta),$$

因此使 $\sup_{\theta} L(\theta, \mathbf{x})$ 达到上确界之点 $\hat{\theta}^*$, 即为使 $\sup_{\theta} g(T(\mathbf{x}), \theta)$ 达到上确界之点, 它必为 $T(\mathbf{x})$ 的函数.

此性质说明 θ 的极大似然估计 $\hat{\theta}^* = \hat{\theta}^*(X_1, \dots, X_n)$ 可表示为充分统计量 $T(\mathbf{X})$ 的函数, 即 $\hat{\theta}^* = \hat{\theta}^*(T(X_1, \dots, X_n))$, 如例 3.3.2–例 3.3.8 中的极大似然估计皆为充分统计量的函数.

3. 极大似然估计的相合性

极大似然估计的相合性问题, 远没有矩估计那么简单。极大似然估计的相合性问题引起许多统计学者的兴趣, 直到现在都不能说已彻底解决了。1946年, Cramer在一些条件下, 证明了似然方程有一根是参数 θ 的弱相合估计。由于似然方程的根不一定是极大似然估计, 这个结果还是没有解决极大似然估计的相合性问题。直到1949年, Wald 才首次证明了极大似然估计的强相合性, 但所要求的条件很复杂。此后有一些学者继续进行研究, 希望在较少的条件下证明相合性。这些结果, 不便在此一一细述。

存在反例(见本章习题 24), 说明极大似然估计可以不相合。

*4 极大似然估计的相合渐近正态性.

只考虑参数 θ 为一维的情形。设 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$ 为一概率函数族, $\Theta = (a, b)$ 为 R_1 上开区间。设 $f(x, \theta)$ 满足下列条件:

(1) 对一切 $\theta \in \Theta$, 有 $f(x, \theta) > 0$ 当 $x \in \mathcal{X}$; $f(x, \theta) = 0$, 当 $x \notin \mathcal{X}$; 且对一切 $\theta \in \Theta$, $x \in \mathcal{X}$ 时, 偏导数

$$\frac{\partial \log f(x, \theta)}{\partial \theta}, \quad \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2}, \quad \frac{\partial^3 \log f(x, \theta)}{\partial \theta^3}$$

存在。

(2) 存在定义于样本空间 \mathcal{X} 上的函数 $F_1(x)$, $F_2(x)$ 和 $H(x)$, 使对一切 $\theta \in \Theta$ 和 $x \in \mathcal{X}$ 有

$$\left| \frac{\partial f(x, \theta)}{\partial \theta} \right| < F_1(x), \quad \left| \frac{\partial^2 f(x, \theta)}{\partial \theta^2} \right| < F_2(x), \quad \left| \frac{\partial^3 f(x, \theta)}{\partial \theta^3} \right| \leq H(x),$$

其中

$$\int_{-\infty}^{\infty} F_i(x) dx < \infty, \quad i = 1, 2; \quad \int_{-\infty}^{\infty} H(x) f(x, \theta) dx < M, \quad \theta \in \Theta,$$

此处 M 与 θ 无关。

(3) 对一切 $\theta \in \Theta$, 有

$$0 < I(\theta) = E \left[\left(\frac{\partial \log f(X, \theta)}{\partial \theta} \right)^2 \right] = \int_{-\infty}^{\infty} \left(\frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta) dx < \infty.$$

关于极大似然估计的相合渐近正态(CAN)性, 有下列结果。

定理 3.3.2 设 $X = (X_1, \dots, X_n)$ 为自满足上述条件 (1)-(3) 的总体中抽取的简单随机样本, 且设对数似然方程

$$\sum_{i=1}^n \frac{\partial \log f(x_i, \theta)}{\partial \theta} = 0$$

有唯一根 $\hat{\theta}^* = \hat{\theta}^*(X_1, \dots, X_n)$, 则 $\hat{\theta}^*$ 为 θ 的 CAN 估计, 即

$$\sqrt{n}(\hat{\theta}^* - \theta) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{I(\theta)}\right), \quad \theta \in \Theta,$$

且 $\hat{\theta}^*$ 为 θ 的弱相合估计.

证 见文献 [2] 第 85 页定理 2.7 的证明.

例 3.3.9 设 X_1, \dots, X_n 是取自正态总体 $N(\mu, \sigma^2)$ 的样本. 证明 μ 和 σ^2 的 MLE 分别具有渐近正态性.

证 (1) 显然正态分布满足定理 3.3.2 的条件 (1)-(3). 因此在 σ^2 已知时 μ 的 MLE 是 $\hat{\mu} = \bar{X}$. 由于

$$f(x, \mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

故有

$$\log f(x, \mu) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}.$$

所以

$$I_1(\theta) = E\left[\left(\frac{\partial \log f(x, \mu)}{\partial \mu}\right)^2\right] = E\left[\left(\frac{x-\mu}{\sigma^2}\right)^2\right] = \frac{1}{\sigma^2},$$

从而由定理 3.3.2 可知

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

即 $\hat{\mu}$ 的渐近分布为 $N(\mu, \sigma^2/n)$.

(2) 在 μ 已知时, 可知 σ^2 的 MLE 是 $\hat{\sigma}^2 = S_\mu^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$. 由于

$$I_2(\theta) = E\left[\left(\frac{\partial \log f(x, \sigma^2)}{\partial \sigma^2}\right)^2\right] = E\left[-\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4}\right]^2 = \frac{1}{2\sigma^4},$$

故由定理 3.3.2 可知

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{\mathcal{L}} N(0, 2\sigma^4),$$

即 $\hat{\sigma}^2$ 的渐近分布为 $N(\sigma^2, 2\sigma^4/n)$.

*3.4 一致最小方差无偏估计

3.4.1 引言及定义

设有一参数分布族 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$, 其中 Θ 为参数空间. 设 $g(\theta)$ 是定义在 Θ 上的实函数, $\mathbf{X} = (X_1, \dots, X_n)$ 为自 \mathcal{F} 中抽取的简单样本, $\hat{g}(\mathbf{X}) =$

$\hat{g}(X_1, \dots, X_n)$ 为 $g(\theta)$ 的一个估计量, 如何评价 $\hat{g}(\mathbf{X})$ 的优劣? 一般用 $\hat{g}(\mathbf{X}) - g(\theta)$ 作为其偏差, 为消除 $\hat{g}(\mathbf{X}) - g(\theta)$ 取值出现 “+,-” 可能抵消的影响, 一般用 $(\hat{g}(\mathbf{X}) - g(\theta))^2$ 来代替. 由于这个量是随机的, 将其求平均, 即计算其均值, 以得到一个整体性的指标 $E_\theta(\hat{g}(\mathbf{X}) - g(\theta))^2$, 这就是估计量 $\hat{g}(\mathbf{X})$ 的均方误差 (在统计决策问题中, 若取损失函数为平方损失, 则 $E_\theta(\hat{g}(\mathbf{X}) - g(\theta))^2$ 就是风险函数).

定义 3.4.1 设 $\hat{g}(\mathbf{X})$ 为 $g(\theta)$ 的估计量, 则称 $E_\theta(\hat{g}(\mathbf{X}) - g(\theta))^2$ 为 $\hat{g}(\mathbf{X})$ 的均方误差 (mean square error, MSE).

设 $\hat{g}_1(\mathbf{X})$ 和 $\hat{g}_2(\mathbf{X})$ 为 $g(\theta)$ 的两个不同的估计量, 若

$$E_\theta(\hat{g}_1(\mathbf{X}) - g(\theta))^2 \leq E_\theta(\hat{g}_2(\mathbf{X}) - g(\theta))^2, \quad \text{一切 } \theta \in \Theta,$$

且不等号至少对某个 $\theta \in \Theta$ 成立, 则称在 MSE 准则下 $\hat{g}_1(\mathbf{X})$ 优于 $\hat{g}_2(\mathbf{X})$.

若存在 $\hat{g}^*(\mathbf{X})$, 使得对 $g(\theta)$ 的任一估计量 $\hat{g}(\mathbf{X})$, 都有

$$E_\theta(\hat{g}^*(\mathbf{X}) - g(\theta))^2 \leq E_\theta(\hat{g}(\mathbf{X}) - g(\theta))^2, \quad \text{一切 } \theta \in \Theta,$$

则称 $\hat{g}^*(\mathbf{X})$ 为 $g(\theta)$ 的一致最小均方误差估计.

可惜的是, 一致最小均方误差估计常不存在. 解决这个问题的办法之一是把最优性准则放宽一些, 使适合这种最优性准则的估计一般能存在. 从直观上想, 在一个大的估计类中, 一致最优估计量不存在, 把估计类缩小, 就有可能存在一致最优的估计量. 因此把估计类缩小为无偏估计类来考虑. 在无偏估计类中, 估计量的均方误差就变为其方差, 即当 $\hat{g}(\mathbf{X})$ 为 $g(\theta)$ 的无偏估计时, $MSE(\hat{g}(\mathbf{X})) = D_\theta(\hat{g}(\mathbf{X}))$, 此处 $D_\theta(\hat{g}(\mathbf{X}))$ 表示 $\hat{g}(\mathbf{X})$ 的方差.

存在这样的情形, 参数 $g(\theta)$ 的无偏估计不存在. 请看下例.

例 3.4.1 设样本 $X \sim$ 二项分布 $b(n, p)$, n 已知而 p 未知. 令 $g(p) = 1/p$, 则参数 $g(p)$ 的无偏估计不存在.

证 采用反证法. 若不然, $g(p)$ 有无偏估计 $\hat{g}(X)$. 由于 X 只取 $0, 1, \dots, n$ 这些值, 令 $\hat{g}(X)$ 的取值用 $\hat{g}(i) = a_i$ 表示, $i = 0, 1, \dots, n$. 由 $\hat{g}(X)$ 的无偏性, 应有

$$E_p(\hat{g}(X)) = \sum_{i=0}^n a_i \binom{n}{i} p^i (1-p)^{n-i} = \frac{1}{p}, \quad 0 < p < 1.$$

于是有

$$\sum_{i=0}^n a_i \binom{n}{i} p^{i+1} (1-p)^{n-i} - 1 = 0, \quad 0 < p < 1.$$

但上式左端是 p 的 $n+1$ 次多项式, 它最多在 $(0, 1)$ 区间有 $n+1$ 个实根, 可无偏性要求对 $(0, 1)$ 中的任一实数 p 上式都成立. 这个矛盾说明 $g(p) = 1/p$ 的无偏估计不存在.

今后把不存在无偏估计的参数除外. 参数的无偏估计若存在, 则称此参数为可估参数; 若参数函数的无偏估计存在, 则称此函数为可估函数 (estimable function). 因此可估函数的无偏估计类是非空的.

假如可估函数的无偏估计类中的无偏估计不止一个, 怎样比较它们的优劣? 故引入下列的定义.

定义 3.4.2 设 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$ 是一个参数分布族, 其中 Θ 为参数空间, $g(\theta)$ 为定义在 Θ 上的可估函数. 设 $\hat{g}^*(\mathbf{X}) = \hat{g}^*(X_1, \dots, X_n)$ 为 $g(\theta)$ 的一个无偏估计, 若对 $g(\theta)$ 的任一无偏估计 $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \dots, X_n)$, 都有

$$D_\theta(\hat{g}^*(\mathbf{X})) \leq D_\theta(\hat{g}(\mathbf{X})), \quad \text{一切 } \theta \in \Theta,$$

则称 $\hat{g}^*(\mathbf{X})$ 是 $g(\theta)$ 的一致最小方差无偏估计 (uniformly minimum variance unbiased estimation, UMVUE).

对给定参数分布族, 如何寻找可估函数的 UMVUE 呢? 本节以下将介绍两种方法: 零无偏估计法和充分完全统计量法, 3.5 节的 Cramer-Rao 不等式也是求 UMVUE 的一种方法.

下列的引理提供了一个改进无偏估计的方法, 它在本节以下寻找 UMVUE 中, 起到简化问题的作用.

引理 3.4.1 设 $T = T(\mathbf{X})$ 是一个充分统计量, 而 $\hat{g}(\mathbf{X})$ 是 $g(\theta)$ 的一个无偏估计, 则

$$h(T) = E(\hat{g}(\mathbf{X})|T)$$

是 $g(\theta)$ 的无偏估计, 并且

$$D_\theta(h(T)) \leq D_\theta(\hat{g}(\mathbf{X})), \quad \text{一切 } \theta \in \Theta, \tag{3.4.1}$$

其中等号当且仅当 $P_\theta(\hat{g}(\mathbf{X}) = h(T)) = 1$, 即 $\hat{g}(\mathbf{X}) = h(T)$, a.s. P_θ 成立.

证 因 $T(\mathbf{X})$ 为充分统计量, 按定义, 给定 T 时 \mathbf{X} 的条件分布与 θ 无关. 因此给定 T 时, $\hat{g}(\mathbf{X})$ 的条件期望 $h(T) = E_\theta(\hat{g}(\mathbf{X})|T)$ 与 θ 无关. 所以 $h(T)$ 是统计量, 可作为 $g(\theta)$ 的估计量, 且有

$$E_\theta(h(T)) = E_\theta[E_\theta(\hat{g}(\mathbf{X})|T)] = E_\theta(\hat{g}(\mathbf{X})) = g(\theta), \quad \text{一切 } \theta \in \Theta.$$

因此 $h(T)$ 是 $g(\theta)$ 的无偏估计. 为证式 (3.4.1), 易知

$$\begin{aligned} D_\theta(\hat{g}(\mathbf{X})) &= E\{\hat{g}(\mathbf{X}) - h(T)\}^2 + D_\theta(h(T)) \\ &= E_\theta[\hat{g}(\mathbf{X}) - h(T)]^2 + D_\theta(h(T)) \\ &\quad + 2E_\theta\{[h(T) - g(\theta)][\hat{g}(\mathbf{X}) - h(T)]\}, \end{aligned}$$

由于 $E_\theta[\hat{g}(\mathbf{X})|T] - h(T) = 0$, 可知

$$\begin{aligned} & E_\theta\{[h(T) - g(\theta)][\hat{g}(\mathbf{X}) - h(T)]\} \\ &= E_\theta\{E_\theta[(h(T) - g(\theta))(\hat{g}(\mathbf{X}) - h(T))|T]\} \\ &= E_\theta\{[h(T) - g(\theta)]E_\theta[(\hat{g}(\mathbf{X}) - h(T))|T]\} \\ &= E_\theta\{[h(T) - g(\theta)][E_\theta(\hat{g}(\mathbf{X})|T) - h(T)]\} = 0. \end{aligned}$$

故有

$$\begin{aligned} D_\theta(\hat{g}(\mathbf{X})) &= E_\theta[g(\mathbf{X}) - h(T)]^2 + D_\theta(h(T)) \\ &\geq D_\theta(h(T)), \quad \text{一切 } \theta \in \Theta, \end{aligned}$$

且等号成立的充要条件是

$$E_\theta[\hat{g}(\mathbf{X}) - h(T)]^2 = 0,$$

此即 $\hat{g}(\mathbf{X}) = h(T)$, a.s. P_θ 成立. 引理证毕.

这个引理提供了一个改进无偏估计的方法, 即一个无偏估计 $\hat{g}(\mathbf{X})$ 对充分统计量 $T(\mathbf{X})$ 的条件期望 $E\{\hat{g}(\mathbf{X})|T\}$ 将能导出一个新的无偏估计, 且它的方差不会超过原估计量 $\hat{g}(\mathbf{X})$ 的方差. 若原估计 $\hat{g}(\mathbf{X})$ 不是 $T(\mathbf{X})$ 的函数, 则新的无偏估计 $E(\hat{g}(\mathbf{X})|T)$ 一定比原估计 $\hat{g}(\mathbf{X})$ 具有更小的方差. 这个引理还表明一致最小方差无偏估计一定是充分统计量的函数, 否则可以通过充分统计量, 按引理 3.4.1 的方法构造出一个具有更小方差的无偏估计来.

例 3.4.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从两点分布族 $\{b(1, p) : 0 < p < 1\}$ 中抽取的简单样本. 显然, X_1 是 p 的一个无偏估计, $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 是 p 的充分统计量, 试利用 $T = T(\mathbf{X})$ 构造一个具有比 X_1 方差更小的无偏估计.

解 由引理 3.4.1 可知, 容易构造 p 的一个无偏估计如下:

$$\begin{aligned} h(t) &= E(X_1|T=t) = 1 \cdot P(X_1=1|T=t) + 0 \cdot P(X_1=0|T=t) \\ &= \frac{P(X_1=1, T=t)}{P(T=t)} = \frac{P(X_1=1, X_2+\dots+X_n=t-1)}{P(T=t)} \\ &= \frac{p \cdot \binom{n-1}{t-1} p^{t-1} (1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{t}{n} = \bar{x}. \end{aligned}$$

显然, 样本均值 $h(T) = \bar{x}$ 的方差为 $p(1-p)/n$, 而 X_1 的方差为 $p(1-p)$, 当 $n \geq 2$ 时 \bar{x} 的方差更小.

经过上述改进后的无偏估计是否为 UMVUE? 引理 3.4.1 并未解决这个问题. 下面介绍的两种方法可以回答这个问题.

*3.4.2 零无偏估计法

本段介绍一个一般性的定理, 用以判断某一估计量是否为 UMVUE.

定理 3.4.1 设 $\hat{g}(\mathbf{X})$ 是 $g(\theta)$ 的一个无偏估计, $D_\theta(\hat{g}(\mathbf{X})) < \infty$, 对任何 $\theta \in \Theta$ 若对任何满足条件 “ $E_\theta l(\mathbf{X}) = 0$, 对一切 $\theta \in \Theta$ ” 的统计量 $l(\mathbf{X})$, 必有

$$\text{Cov}_\theta(\hat{g}(\mathbf{X}), l(\mathbf{X})) = E_\theta[\hat{g}(\mathbf{X}) \cdot l(\mathbf{X})] = 0, \quad \text{一切 } \theta \in \Theta, \quad (3.4.2)$$

则 $\hat{g}(\mathbf{X})$ 是 $g(\theta)$ 的 UMVUE.

注 3.4.1 从形式上看, 条件 “ $E_\theta l(\mathbf{X}) = 0, \theta \in \Theta$ ” 可解释为 “ $l(\mathbf{X})$ 是零的无偏估计”, 由此得到求 UMVUE 的方法之一的名称“零无偏估计法”.

定理 3.4.1 还可进一步加强为: 设 $\hat{g}(\mathbf{X})$ 为 $g(\theta)$ 的一个无偏估计, $D_\theta(\hat{g}(\mathbf{X})) < \infty, \theta \in \Theta$, 则 $\hat{g}(\mathbf{X})$ 是 $g(\theta)$ 的 UMVUE 的充分必要条件是: 对任何满足条件 “ $E_\theta l(\mathbf{X}) = 0, \text{一切 } \theta \in \Theta$ ” 的统计量 $l(\mathbf{X})$, 必有式 (3.4.2) 成立.

可见定理 3.4.1 只是给出了充分条件, 下面将给出其证明. 必要性的证明留给读者作为练习.

证 设 $\hat{g}_1(\mathbf{X})$ 为 $g(\theta)$ 的任一无偏估计. 记 $l(\mathbf{X}) = \hat{g}_1(\mathbf{X}) - \hat{g}(\mathbf{X})$ 为零的无偏估计, 由于式 (3.4.2) 成立, 因而

$$\begin{aligned} D_\theta(\hat{g}_1(\mathbf{X})) &= D_\theta[\hat{g}(\mathbf{X}) + l(\mathbf{X})] \\ &= D_\theta(\hat{g}(\mathbf{X})) + D_\theta(l(\mathbf{X})) + 2\text{Cov}_\theta(\hat{g}(\mathbf{X}), l(\mathbf{X})) \\ &= D_\theta(\hat{g}(\mathbf{X})) + D_\theta(l(\mathbf{X})) \\ &\geq D_\theta(\hat{g}(\mathbf{X})), \quad \text{一切 } \theta \in \Theta. \end{aligned}$$

这就证明了所要的结果.

从定理的内容看, 它是一个验证某个特定的估计量 $\hat{g}(\mathbf{X})$ 为 UMVUE 的方法. 至于这个特定的估计 $\hat{g}(\mathbf{X})$ 从何而来, 定理 3.4.1 不能提供任何帮助, 它不是 UMVUE 的构造性定理. $\hat{g}(\mathbf{X})$ 可以从直观的想法提出, 如通过矩估计或极大似然估计等方法获得, 然后利用此定理验证它是否为 $g(\theta)$ 的 UMVUE. 条件 (3.4.2) 的验证也不容易, 因为零无偏估计很多.

在定理 3.4.1 中若存在充分统计量 $T = T(\mathbf{X})$, 使得 $g(\theta)$ 的无偏估计 $\hat{g}(\mathbf{X})$ 还是充分统计量 $T(\mathbf{X})$ 的函数, 即 $\hat{g}(\mathbf{X}) = h(T(\mathbf{X}))$, 则有下述的推理.

推论 3.4.1 设 $T = T(\mathbf{X})$ 为 θ 的充分统计量, $h(T(\mathbf{X}))$ 是 $g(\theta)$ 的一个无偏估计, $D_\theta(h(T)) < \infty$, 对任何 $\theta \in \Theta$, 且对任何满足条件 “ $E_\theta \delta(T) = 0$, 对一切 $\theta \in \Theta$ ” 的统计量 $\delta(T)$, 必有

$$\text{Cov}_\theta(h(T), \delta(T)) = E_\theta[h(T) \cdot \delta(T)] = 0, \quad \text{一切 } \theta \in \Theta, \quad (3.4.3)$$

则 $h(T)$ 是 $g(\theta)$ 的 UMVUE.

证 与定理 3.4.1 的证明方法完全相同.

下面的几个例子中有关参数的无偏估计都是充分统计量 T 的函数, 因此可直接利用推论 3.4.1 验证其为 UMVUE.

例 3.4.3 求例 3.4.2 中 $g(p) = p$ 的 UMVUE.

解 由例 3.4.2 已知 $T = \sum_{i=1}^n X_i$ 为充分统计量, p 的无偏估计 $h(T) = T/n$, 故只要验证它满足推论 3.4.1 的条件. 显然 $D_\theta(h(T)) = p(1-p)/n < \infty$, $0 < p < 1$. 现设 $\delta(T)$ 为任一零无偏估计, 并记 $a_i = \delta(i)$, $i = 0, 1, 2, \dots, n$, 则因 $T \sim b(n, p)$, 故有

$$E_p(\delta(T)) = \sum_{i=0}^n a_i \binom{n}{i} p^i (1-p)^{n-i} = 0, \quad 0 < p < 1.$$

约去因子 $(1-p)^n$, 并记 $\varphi = p/(1-p)$ (φ 取值于 $(0, \infty)$), 将上式改写为

$$\sum_{i=0}^n a_i \binom{n}{i} \varphi^i = 0, \quad \text{一切 } 0 < \varphi < \infty.$$

上式左边是 φ 的多项式, 要使其为 0, 必有 $a_i \binom{n}{i} = 0$, 即 $a_i = 0$, $i = 0, 1, 2, \dots, n$. 故 $\delta(T)$ 在其定义域中处处为 0, 因而有 $\delta(T) \equiv 0$. 从而有

$$\text{Cov}_p(h(T), \delta(T)) = E_p(h(T) \cdot \delta(T)) = 0,$$

即推论 3.4.1 条件成立, 故 $h(T) = \bar{X} = T/n$ 为 p 的 UMVUE.

例 3.4.4 设 $X = (X_1, \dots, X_n)$ 为从指数分布 $\text{Exp}(\lambda)$ (见例 3.3.5) 中抽取的简单样本, 求总体均值 $g(\lambda) = 1/\lambda$ 的 UMVUE.

解 由于在指数分布族中 $T = \sum_{i=1}^n X_i$ 为 λ 的充分统计量, 由指数分布的性质可知 $T \sim \Gamma(n, \lambda)$, 其密度函数为

$$\varphi(t, \lambda) = \begin{cases} \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}, & \text{当 } t > 0, \\ 0, & \text{当 } t \leq 0, \end{cases}$$

其中 $\lambda > 0$. 取 $h(T) = T/n$, 显然 $E(h(T)) = 1/\lambda$, 即 $h(T)$ 为 $g(\lambda) = 1/\lambda$ 的无偏估计, 且 $D_\lambda(h(T)) = 1/(n\lambda^2) < \infty$. 现设 $\delta(T)$ 为任一零无偏估计, 故有

$$E(\delta(T)) = \int_0^\infty \delta(t) \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} dt = 0,$$

即 $\int_0^\infty \delta(t) t^{n-1} e^{-\lambda t} dt = 0$. 两边对 λ 求导得

$$\int_0^\infty \delta(t)t^n e^{-\lambda t} dt = 0,$$

此式等价于 $E_\lambda(h(T) \cdot \delta(T)) = \text{Cov}_\lambda(h(T) \cdot \delta(T)) = 0$, 即推论 3.4.1 条件成立. 因此 $h(T) = T/n$ 为 $g(\lambda) = 1/\lambda$ 的 UMVUE.

例 3.4.5 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从均匀分布 $U(0, \theta)$ 中抽取的简单样本, 求 θ 的 UMVUE.

解 由例 2.7.7 已知 $T = T(\mathbf{X}) = X_{(n)}$ 是参数 θ 的充分统计量, 又在例 3.3.6 中已证明了 $h(T) = (n+1)T/n$ 是 θ 的无偏估计, 且 $D_\theta(h(T)) = \theta^2/[n(n+2)] < \infty$. 现设 $\delta(T)$ 为任一零无偏估计, T 的密度函数如式 (3.3.5) 所示, 因此有

$$E_\theta(\delta(T)) = \int_0^\theta \delta(t) \cdot \frac{nt^{n-1}}{\theta^n} dt = 0, \quad \text{一切 } \theta > 0,$$

于是有

$$\int_0^\theta \delta(t)t^{n-1} dt = 0, \quad \text{一切 } \theta > 0.$$

将上式两边对 θ 求导得 $\delta(\theta)\theta^{n-1} = 0$, 一切 $\theta > 0$. 故有 $\delta(\theta) \equiv 0$, 一切 $\theta > 0$. 可见 $\text{Cov}_\theta(h(T), \delta(T)) = E_\theta(h(T) \cdot \delta(T)) = 0$, 即推论 3.4.1 条件成立. 因此 $h(T) = (n+1)X_{(n)}/n$ 为 $g(\theta) = \theta$ 的 UMVUE.

* **例 3.4.6** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从正态分布 $N(a, \sigma^2)$ 中抽取的简单随机样本, 求 a 和 σ^2 的 UMVUE.

解 由例 2.7.5 可知 $T = (T_1, T_2)$ 为 $\theta = (a, \sigma^2)$ 的充分统计量, 其中 $T_1 = \bar{X}$, $T_2 = \sum_{i=1}^n (X_i - \bar{X})^2$, 又 T_1 和 T_2 独立且 $T_1 \sim N(a, \sigma^2/n)$, $T_2/\sigma^2 \sim \chi_{n-1}^2$. 因此 (T_1, T_2) 的联合密度为

$$f_\theta(t_1, t_2) = \begin{cases} \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-\frac{n(t_1-a)^2}{2\sigma^2}} \cdot \frac{1}{2^{\frac{n-1}{2}} \Gamma((n-1)/2) \sigma^{n-1}} t_2^{\frac{n-1}{2}-1} e^{-\frac{t_2}{2\sigma^2}}, \\ \quad -\infty < t_1 < +\infty, t_2 > 0, \\ 0, \quad \text{其他.} \end{cases} \quad (3.4.4)$$

先考虑 a 的 UMVUE, 令 $h_1(T) = T_1$, 显然 $h_1(T)$ 为 $g_1(\theta) = a$ 的无偏估计, 且 $D_\theta(h_1(T)) = \sigma^2/n < \infty$. 现设 $\delta(T) = \delta(T_1, T_2)$ 为任一零无偏估计, 则有

$$E_\theta(\delta(T)) = \int_0^\infty \int_{-\infty}^\infty \delta(t_1, t_2) f_\theta(t_1, t_2) dt_1 dt_2 = 0,$$

此处 $-\infty < a < \infty$, $\sigma > 0$. 将上式两边对 a 求导数, 得

$$\int_0^\infty \int_{-\infty}^\infty \delta(t_1, t_2) (t_1 - a) \cdot t_2^{\frac{n-3}{2}} \exp \left\{ -\frac{1}{2\sigma^2} [n(t_1 - a)^2 + t_2] \right\} dt_1 dt_2 = 0,$$

按式 (3.4.4), 上式等价于

$$\text{Cov}_\theta(h_1(T), \delta(T)) = E_\theta(h_1(T) \cdot \delta(T_1, T_2)) = 0, \quad -\infty < a < +\infty, \sigma > 0,$$

故推论 3.4.1 的条件满足. 所以 $h_1(T) = T_1$ 为 $g_1(\theta) = a$ 的 UMVUE.

同理可验证 $T_2/(n-1) = S^2$ 为 $g_2(\theta) = \sigma^2$ 的 UMVUE, 这一验证留给读者作为练习.

比较例 3.4.5 和例 3.4.6, 可以发现下列现象: 在例 3.4.5 中总体均值为 $\theta/2$, 故 $(n+1)X_{(n)}/(2n)$ 和 $\bar{X} = \sum_{i=1}^n X_i/n$ 皆为 $g(\theta) = \theta/2$ 的无偏估计, 但 $(n+1)X_{(n)}/(2n)$ 为 $g(\theta)$ 的 UMVUE, 它比 \bar{X} 好. 而在例 3.4.6 中正好相反. 此例中因 \bar{X} 为 $g_1(\theta) = a$ 的 UMVUE, 故 \bar{X} 比基于统计量 $X_{(n)}$ 的无偏估计好. 这说明某一统计量的优劣不仅取决于统计量的本身, 而且与统计模型有关.

3.4.3 充分完全统计量法

下列定理给出的求 UMVUE 的方法, 即充分完全统计量法, 是由 E.L. Lehmann 和 H. Scheffé 给出的, 完全统计量的概念也是由他们在 1950 年提出的.

定理 3.4.2 (Lehmann-Scheffé 定理) 设 $X \sim \{f(x, \theta), \theta \in \Theta\}$, Θ 为参数空间. 令 X_1, \dots, X_n 为从总体 X 中抽取的简单样本, $g(\theta)$ 为定义于参数空间 Θ 上的可估函数, $T(\mathbf{X})$ 为一个充分完全统计量. 若 $\hat{g}(T(\mathbf{X}))$ 为 $g(\theta)$ 的一个无偏估计, 则 $\hat{g}(T(\mathbf{X}))$ 是 $g(\theta)$ 的唯一的 UMVUE (唯一性是在这样的意义下: 设 \hat{g} 和 \hat{g}_1 是 $g(\theta)$ 的两个估计量, 若 $P_\theta(\hat{g} = \hat{g}_1) = 1$, 对一切 $\theta \in \Theta$, 则视 \hat{g} 和 \hat{g}_1 是同一个估计量).

证 先证唯一性. 设 $\hat{g}_1(T(\mathbf{X}))$ 为 $g(\theta)$ 的任一无偏估计, 令 $\delta(T(\mathbf{X})) = \hat{g}(T(\mathbf{X})) - \hat{g}_1(T(\mathbf{X}))$, 则 $E_\theta \delta(T(\mathbf{X})) = E_\theta \hat{g}(T(\mathbf{X})) - E_\theta \hat{g}_1(T(\mathbf{X})) = 0, \theta \in \Theta$. 由 $T(\mathbf{X})$ 为完全统计量, 可知 $\delta(T(\mathbf{X})) = 0$, a.s. P_θ 成立, 即 $\hat{g}(T(\mathbf{X})) = \hat{g}_1(T(\mathbf{X}))$, a.s. P_θ 成立, 故唯一性成立.

再证一致最小方差性. 设 $\varphi(\mathbf{X})$ 为 $g(\theta)$ 的任一无偏估计. 令 $h(T(\mathbf{X})) = E(\varphi(\mathbf{X})|T)$, 由 $T(\mathbf{X})$ 为充分统计量, 故知 $h(T(\mathbf{X}))$ 与 θ 无关, 是统计量. 由引理 3.4.1 可知

$$E_\theta(h(T(\mathbf{X}))) = g(\theta), \quad \text{一切 } \theta \in \Theta,$$

$$D_\theta(h(T(\mathbf{X}))) \leq D_\theta(\varphi(\mathbf{X})), \quad \text{一切 } \theta \in \Theta.$$

由唯一性得 $\hat{g}(T(\mathbf{X})) = h(T(\mathbf{X}))$, a.e. P_θ 成立, 由上式可知

$$D_\theta(\hat{g}(T(\mathbf{X}))) = D_\theta(h(T(\mathbf{X}))) \leq D_\theta(\varphi(\mathbf{X})), \quad \text{一切 } \theta \in \Theta,$$

所以 $\hat{g}(T(\mathbf{X}))$ 为 $g(\theta)$ 的 UMVUE, 且唯一.

推论 3.4.2 设样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的分布为指数族

$$f(\mathbf{x}, \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp \left\{ \sum_{j=1}^k \theta_j T_j(\mathbf{x}) \right\} h(\mathbf{x}), \quad \boldsymbol{\theta} = (\theta_1, \dots, \theta_k) \in \Theta^*.$$

令 $T(\mathbf{X}) = (T_1(\mathbf{X}), \dots, T_k(\mathbf{X}))$, 若自然参数空间 Θ^* 作为 R_k 的子集有内点, 且 $h(T(\mathbf{X}))$ 为 $g(\boldsymbol{\theta})$ 的无偏估计, 则 $h(T(\mathbf{X}))$ 为 $g(\boldsymbol{\theta})$ 的唯一的 UMVUE.

证 在推论 3.4.2 的条件下, 由指数族的性质可知 $T(\mathbf{X})$ 为充分完全统计量. 故由 Lehmann-Scheffé 定理(简记 L-S 定理), 得知 $h(T(\mathbf{X}))$ 为 $g(\boldsymbol{\theta})$ 唯一的 UMVUE.

例 3.4.7 证明例 3.4.3 中获得的 p 的无偏估计 $h(T) = T/n = \bar{X}$ 为 p 的 UMVUE.

证 由因子分解定理可知 $T = T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为两点分布 $b(1, p)$ 中参数 p 的充分统计量, 由例 2.8.4 可知 T 也是完全统计量, 故 $h(T) = T/n$ 是充分完全统计量 T 的函数, 且 $E_p[h(T)] = p$, 对 $0 < p < 1$. 因此由 L-S 定理可知 $h(T) = \bar{X}$ 为 p 的唯一的 UMVUE.

例 3.4.8 在例 3.4.7 中, 已知 $T = \sum_{i=1}^n X_i$ 服从二项分布 $b(n, p)$, 且 $T(\mathbf{X})$ 为充分完全统计量, 求 $g(p) = p(1-p)$ 的 UMVUE.

解法 1 令 $\varphi(\mathbf{X}) = I_{[X_1=1, X_2=0]}$, 则 $E[\varphi(\mathbf{X})] = P(X_1=1, X_2=0) = p(1-p)$. 因此 $\varphi(\mathbf{X})$ 为 $g(p)$ 的无偏估计. 注意到 $T = \sum_{i=1}^n X_i \sim b(n, p)$ 和 $\sum_{i=3}^n X_i \sim b(n-2, p)$, 故 $g(p)$ 改进的无偏估计为

$$\begin{aligned} h(t) &= E(\varphi(\mathbf{X}) | T=t) = P(X_1=1, X_2=0 | T=t) \\ &= P\left(X_1=1, X_2=0, \sum_{i=3}^n X_i = t-1\right) / P(T=t) \\ &= \binom{n-2}{t-1} / \binom{n}{t} = \frac{t(n-t)}{n(n-1)}. \end{aligned}$$

由引理 3.4.1 可知 $h(T) = T(n-T)/[n(n-1)]$ 为 $g(p)$ 的一个无偏估计, 它还是充分完全统计量 T 的函数, 故由 L-S 定理可知 $h(T)$ 为 $g(p)$ 的 UMVUE.

解法 2 设 $\delta(T)$ 为 $g(p) = p(1-p)$ 的一个无偏估计, 要导出 $\delta(T)$ 的表达式. 按无偏估计的定义 $E[\delta(T)] = p(1-p)$ 及 $T \sim b(n, p)$, 可得

$$\sum_{t=0}^n \binom{n}{t} \delta(t) p^t (1-p)^{n-t} = p(1-p), \quad \text{一切 } 0 < p < 1.$$

令 $\rho = p/(1-p)$, 故有 $p = \rho/(1+\rho)$, $1-p = 1/(1+\rho)$, 将它们代入上式得

$$\sum_{t=0}^n \binom{n}{t} \delta(t) \rho^t = \rho(1+\rho)^{n-2}, \quad \text{一切 } 0 < \rho < \infty.$$

将 $\rho(1+\rho)^{n-2}$ 展开得

$$\rho(1+\rho)^{n-2} = \sum_{l=0}^{n-2} \binom{n-2}{l} \rho^{l+1} = \sum_{t=1}^{n-1} \binom{n-2}{t-1} \rho^t,$$

将其代入前一式右边得

$$\sum_{t=0}^n \binom{n}{t} \delta(t) \rho^t = \sum_{t=1}^{n-1} \binom{n-2}{t-1} \rho^t, \quad \text{一切 } 0 < \rho < \infty.$$

上式两边为 ρ 的多项式, 比较其系数得

$$\begin{aligned} \delta(t) &= 0, \quad \text{当 } t = 0, n; \\ \delta(t) &= \frac{\binom{n-2}{t-1}}{\binom{n}{t}} = \frac{t(n-t)}{n(n-1)}, \quad \text{当 } t = 1, 2, \dots, n-1. \end{aligned}$$

综合上述两式得

$$\delta(T) = \frac{T(n-T)}{n(n-1)}, \quad T = 0, 1, \dots, n$$

为 $g(p) = p(1-p)$ 的无偏估计, 它又是充分完全统计量 $T = \sum_{i=1}^n X_i$ 的函数, 由 L-S 定理可知 $\delta(T)$ 为 $g(p)$ 的 UMVUE.

例 3.4.9 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 Poisson 分布 $P(\lambda)$ 中抽取的简单随机样本, 求 (1) $g_1(\lambda) = \lambda$; (2) $g_2(\lambda) = \lambda^r$, $r > 0$ 为自然数; (3) $g_3(\lambda) = P_\lambda(X_1 = x)$ 的 UMVUE.

解 由 2.7 节和 2.8 节可知 $T = T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为充分完全统计量.

(1) 令 $h_1(T) = T/n$, $E(h_1(T)) = E(\bar{X}) = \lambda$, 故 $h_1(T)$ 是充分完全统计量 T 的函数, 且是 λ 的无偏估计, 故由 L-S 定理可知 $h_1(T)$ 是 λ 的 UMVUE.

(2) 由于 $T \sim P(n\lambda)$, 令 $h_2(T)$ 为 $g_2(\lambda) = \lambda^r$ 的无偏估计, 故有 $E_\lambda[h_2(T)] = g_2(\lambda)$, 即

$$\sum_{t=0}^{\infty} h_2(t) \frac{e^{-n\lambda}(n\lambda)^t}{t!} = \lambda^r.$$

此式等价于

$$\sum_{t=0}^{\infty} h_2(t) \frac{n^t \lambda^t}{t!} = \lambda^r e^{n\lambda}.$$

将上式右边作展开得

$$\lambda^r e^{n\lambda} = \sum_{l=0}^{\infty} \frac{n^l \lambda^{l+r}}{l!} = \sum_{t=r}^{\infty} \frac{n^{t-r} \lambda^t}{(t-r)!}.$$

将其代入前一式右边得

$$\sum_{t=0}^{\infty} h_2(t) \frac{n^t \lambda^t}{t!} = \sum_{t=r}^{\infty} \frac{n^{t-r} \lambda^t}{(t-r)!}.$$

上述等式两边是 λ 的幂级数, 比较其系数得

$$\begin{aligned} h_2(t) &= 0, \quad \text{当 } t = 0, 1, \dots, r-1, \\ h_2(t) &= \frac{t! n^{t-r}}{(t-r)! n^t} = \frac{t(t-1)\cdots(t-r+1)}{n^r}, \quad \text{当 } t = r, r+1, \dots \end{aligned}$$

综合上述两式得

$$h_2(T) = \frac{T(T-1)\cdots(T-r+1)}{n^r}, \quad \text{当 } T = 0, 1, 2, \dots$$

为 $g_2(\lambda) = \lambda^r$ 的无偏估计, $h_2(T)$ 是充分完全统计量 T 的函数, 故由 L-S 定理可知 $h_2(T)$ 为 $g_2(\lambda)$ 的 UMVUE.

(3) 由 $P_\lambda(X_1 = x) = e^{-\lambda} \lambda^x / x!$, 可见它是参数 λ 的函数, 故可用 $g_3(\lambda)$ 表示. 令 $\varphi(X_1) = I_{[X_1=x]}$, 则 $E_\lambda[\varphi(X_1)] = P_\lambda(X_1 = x)$. 因此 $\varphi(X_1)$ 为 $g_3(\lambda)$ 的无偏估计, 注意到 $T = T(\mathbf{X}) = \sum_{i=1}^n X_i \sim P(n\lambda)$ 和 $\sum_{i=2}^n X_i \sim P((n-1)\lambda)$, 故有

$$\begin{aligned} h_3(t) &= E(\varphi(X_1)|T=t) = P_\lambda(X_1 = x|T=t) = \frac{P_\lambda(X_1 = x, T=t)}{P_\lambda(T=t)} \\ &= \frac{P_\lambda(X_1 = x) P_\lambda(X_2 + \cdots + X_n = t-x)}{P_\lambda(X_1 + \cdots + X_n = t)} = \frac{(n-1)^{t-x} t!}{n^t (t-x)! x!} \\ &= \binom{t}{x} \frac{(n-1)^{t-x}}{n^t} = \binom{t}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{t-x}, \quad t \geq x. \end{aligned}$$

由引理 3.4.1 可知 $h_3(T)$ 为 $g_3(\lambda)$ 的无偏估计, 它又是充分完全统计量 T 的函数, 所以由 L-S 定理可知

$$h_3(T) = \binom{T}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{T-x}$$

为 $g_3(\lambda)$ 的 UMVUE.

例 3.4.10 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从指数分布 $\text{Exp}(\lambda)$ 中抽取的简单随机样本, 求 (1) $g(\lambda) = \lambda$; (2) $g_2(\lambda) = 1 - e^{-\lambda x_0}$ 的 UMVUE, 其中 x_0 已知.

解 由因子分解定理和 2.8.2 节介绍的方法容易证明 $T = T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为充分完全统计量. 由例 3.4.4 已知 $T(\mathbf{X}) \sim \Gamma(n, \lambda)$, 即参数为 n 和 λ 的 Gamma 分布.

(1) 显然

$$E\left(\frac{1}{T}\right) = \int_0^\infty \frac{1}{t} \cdot \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} dt = \frac{\lambda}{n-1}.$$

因此 $h(T) = (n-1)/T$ 为 λ 的无偏估计, 由 L-S 定理可知它是 λ 的 UMVUE.

(2) 令 $h_1(T)$ 为 $e^{-\lambda x_0}$ 的无偏估计, 由于 $T \sim \Gamma(n, \lambda)$, 故有

$$\begin{aligned} E[h_1(T)] &= \int_0^\infty h_1(t) \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} dt \\ &= e^{-\lambda x_0} \int_0^\infty h_1(t) \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda(t-x_0)} dt. \end{aligned}$$

易见要使 $E[h_1(T)] = e^{-\lambda x_0}$, 必须取

$$h_1(t) = \frac{(t-x_0)^{n-1}}{t^{n-1}} I_{(x_0, \infty)}(t).$$

因此

$$h(T) = 1 - h_1(T) = 1 - \frac{(T-x_0)^{n-1}}{T^{n-1}} I_{(x_0, \infty)}(T).$$

为 $g_2(\lambda) = 1 - e^{-\lambda x_0}$ 的无偏估计, 它又是充分完全统计量 T 的函数, 所以由 L-S 定理可知 $h(T)$ 为 $g_2(\lambda)$ 的 UMVUE.

例 3.4.11 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从正态分布 $N(a, \sigma^2)$ 中抽取的简单随机样本, 记 $\theta = (a, \sigma^2)$. 求 (1) a 和 σ^2 ; (2) $g_1(\theta) = \sigma^r$, $r > 0$; (3) $g_2(\theta) = a/\sigma^2$ 的 UMVUE.

解 由 2.7 节和 2.8 节可知 $T = (T_1, T_2)$ 为充分完全统计量, 其中 $T_1 = T_1(\mathbf{X}) = \bar{X}$, $T_2 = T_2(\mathbf{X}) = \sum_{i=1}^n (X_i - \bar{X})^2$.

(1) 由于 $h_1(T) = \bar{X} = T_1$ 和 $h_2(T) = T_2/(n-1)$ 分别为 a 和 σ^2 的无偏估计, 它们又是充分完全统计量的函数, 故由 L-S 定理可知它们分别是 a 和 σ^2 的 UMVUE.

(2) 由于 $Y = T_2/\sigma^2 \sim \chi_{n-1}^2$, 故 σ^r 的无偏估计与 T_2 的幂函数有关. 先计算下式:

$$\begin{aligned} \frac{1}{\sigma^r} E\left(T_2^{\frac{r}{2}}\right) &= E\left(\frac{T_2}{\sigma^2}\right)^{r/2} = \int_0^\infty y^{\frac{r}{2}} \cdot \frac{1}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)} y^{\frac{n-1}{2}-1} e^{-\frac{y}{2}} dy \\ &= \frac{2^{r/2} \Gamma\left(\frac{n+r-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \triangleq \frac{1}{K_{n-1,r}}. \end{aligned}$$

由上式可知

$$E\left(K_{n-1,r} \cdot T_2^{\frac{r}{2}}\right) = \sigma^r.$$

因此估计量

$$h_3(T) = K_{n-1,r} T_2^{r/2} = \frac{\Gamma\left(\frac{n-1}{2}\right)}{2^{r/2} \Gamma\left(\frac{n+r-1}{2}\right)} T_2^{r/2}$$

是 σ^r 的无偏估计, 它是充分完全统计量 $T = (T_1, T_2)$ 的函数, 故由 L-S 定理可知 $h_3(T)$ 是 $g(\theta) = \sigma^r$ 的 UMVUE.

(3) 由下列事实: 若 $X \sim \chi_n^2$, 则 $E(1/X) = 1/(n-2)$, 以及 T_1 和 T_2 相互独立可知

$$E\left(\frac{T_1}{T_2}\right) = E(T_1)E\left(\frac{1}{T_2}\right) = \frac{a}{(n-3)\sigma^2}.$$

因此 $h_4(T) = (n-3)T_1/T_2$ 为 $g_2(\theta)$ 的无偏估计, 它又是充分完全统计量 T 的函数, 故由 L-S 定理可知它是 $g_2(\theta)$ 的 UMVUE.

例 3.4.12 用 Lehmann-Scheffe 定理求例 3.4.5 中均匀分布 $U(0, \theta)$ 中的参数 θ 的 UMVUE.

解 由 2.7 节和 2.8 节可知 $T = T(\mathbf{X}) = \max(X_1, \dots, X_n) = X_{(n)}$ 为充分完全统计量, 由例 3.3.6 已知 $h(T) = (n+1)T/n$ 为 θ 的无偏估计, 故由 L-S 定理立得 $h(T)$ 为 θ 的 UMVUE. 此处求解过程要比例 3.4.5 简单得多.

3.5 Cramer-Rao 不等式

3.5.1 引言

Cramer-Rao 不等式 (简称 C-R 不等式) 是判别一个无偏估计量是否为 UMVUE 的方法之一. 这一方法的思想如下: 设 \mathcal{U}_g 是 $g(\theta)$ 的一切无偏估计构成的类. \mathcal{U}_g 中估计量的方差有一个下界, 这个下界称为 C-R 下界. 因此, 如果 $g(\theta)$ 的一个无偏估计 \hat{g} 的方差达到这个下界, 则 \hat{g} 就是 $g(\theta)$ 的一个 UMVUE, 当然样本分布族和 \hat{g} 要满足一定的正则条件. 这个不等式是由 C.R. Rao 和 H. Cramer 在 1945 年和 1946 年分别证明的. 以后一些统计学者将此不等式的条件作了一些改进和精确化, 但结果的基本形式并无重大变化.

这一方法的缺陷是: 由于 C-R 不等式确定的下界常比真下界为小. 在一些场合, 虽然 $g(\theta)$ 的 UMVUE \hat{g} 存在, 但其方差大于 C-R 下界. 在这一情况下, 用 C-R 不等式就无法判定 $g(\theta)$ 的 UMVUE 存在. 因此这一方法的适用范围不广. C-R 不等式除用于判别 $g(\theta)$ 的 UMVUE 之外, 它在数理统计理论上还有其他的用处, 如估计的效率和有效估计的概念以及 Fisher 信息量都与之有关.

C-R 不等式成立需要样本分布族满足一些正则条件, 适合这些条件的分布族称为 C-R 正则分布族, 下面给出其定义.

定义 3.5.1 若单参数概率函数族 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$ 满足下列条件:

- (1) 参数空间 Θ 是直线上的某个开区间;
- (2) 对任何 $x \in \mathcal{X}$ 及 $\theta \in \Theta$, $f(x, \theta) > 0$, 即分布族具有共同支撑;
- (3) 对任何 $x \in \mathcal{X}$ 及 $\theta \in \Theta$, $\frac{\partial f(x, \theta)}{\partial \theta}$ 存在;
- (4) 概率函数 $f(x, \theta)$ 的积分与微分运算可交换, 即

$$\frac{\partial}{\partial \theta} \int f(x, \theta) dx = \int \frac{\partial}{\partial \theta} f(x, \theta) dx,$$

若 $f(x, \theta)$ 为离散随机变量的概率分布, 上述条件改为无穷级数和微分运算可交换;

- (5) 下列数学期望存在, 且

$$0 < I(\theta) = E_\theta \left[\frac{\partial \log f(X, \theta)}{\partial \theta} \right]^2 < \infty,$$

则称该分布族为 C-R 正则分布族, 其中 (1)-(5) 称为 C-R 正则条件. $I(\theta)$ 称为该分布的 Fisher 信息量 (或称为 Fisher 信息函数).

3.5.2 单参数 C-R 不等式

1. C-R 不等式及例

定理 3.5.1 设 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$ 是 C-R 正则分布族, $g(\theta)$ 是定义于参数空间 Θ 上的可微函数. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是由总体 $f(x, \theta) \in \mathcal{F}$ 中抽取的简单随机样本, $\hat{g}(\mathbf{X})$ 是 $g(\theta)$ 的任一无偏估计, 且满足下列条件:

(6) 积分

$$\int \cdots \int \hat{g}(\mathbf{x}) f(\mathbf{x}, \theta) d\mathbf{x}$$

可在积分号下对 θ 求导数, 此处 $d\mathbf{x} = dx_1 \cdots dx_n$, 则有

$$D_\theta[\hat{g}(\mathbf{X})] \geq \frac{(g'(\theta))^2}{nI(\theta)}, \quad \text{一切 } \theta \in \Theta. \quad (3.5.1)$$

特别当 $g(\theta) = \theta$ 时, 式 (3.5.1) 变为

$$D_\theta[\hat{g}(\mathbf{X})] \geq \frac{1}{nI(\theta)}, \quad \text{一切 } \theta \in \Theta. \quad (3.5.2)$$

当 $f(x, \theta)$ 为离散 r.v. X 的概率分布时, 式 (3.5.1) 变为

$$D_\theta[\hat{g}(\mathbf{X})] \geq \frac{[g'(\theta)]^2}{n \sum_i \left\{ \left[\frac{\partial \log f(x_i, \theta)}{\partial \theta} \right]^2 f(x_i, \theta) \right\}}, \quad \text{一切 } \theta \in \Theta. \quad (3.5.3)$$

证 由于 X_1, \dots, X_n 为 i.i.d. 样本, 故有 $f(\mathbf{x}, \theta) = \prod_{i=1}^n f(x_i, \theta)$. 记

$$S(\mathbf{x}, \theta) = \frac{\partial \log f(\mathbf{x}, \theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \log f(x_i, \theta)}{\partial \theta},$$

因此由正则条件 (3) 和 (4) 可知

$$\begin{aligned} E_\theta\{S(\mathbf{X}, \theta)\} &= \sum_{i=1}^n E_\theta\left\{\frac{\partial \log f(X_i, \theta)}{\partial \theta}\right\} = \sum_{i=1}^n \int \frac{1}{f(x_i, \theta)} \frac{\partial f(x_i, \theta)}{\partial \theta} \cdot f(x_i, \theta) dx_i \\ &= \sum_{i=1}^n \int \frac{\partial f(x_i, \theta)}{\partial \theta} dx_i = \sum_{i=1}^n \frac{\partial}{\partial \theta} \int f(x_i, \theta) dx = 0. \end{aligned}$$

由 $\hat{g}(\mathbf{x})$ 为 $g(\theta)$ 的无偏估计和正则条件 (5) 和 (6) 可知

$$\begin{aligned} \text{Cov}_\theta(\hat{g}(\mathbf{X}), S(\mathbf{X}, \theta)) &= E_\theta\{\hat{g}(\mathbf{X}) \cdot S(\mathbf{X}, \theta)\} \\ &= \int \cdots \int \hat{g}(\mathbf{x}) \left[\frac{1}{f(\mathbf{x}, \theta)} \frac{\partial f(\mathbf{x}, \theta)}{\partial \theta} \right] f(\mathbf{x}, \theta) d\mathbf{x} \\ &= \int \cdots \int \hat{g}(\mathbf{x}) \frac{\partial f(\mathbf{x}, \theta)}{\partial \theta} d\mathbf{x} \\ &= \frac{\partial}{\partial \theta} \int \cdots \int \hat{g}(\mathbf{x}) f(\mathbf{x}, \theta) d\mathbf{x} = \frac{\partial g(\theta)}{\partial \theta} = g'(\theta), \end{aligned}$$

$$\begin{aligned} D_\theta(S(\mathbf{X}, \theta)) &= \sum_{i=1}^n D_\theta\left\{\frac{\partial \log f(X_i, \theta)}{\partial \theta}\right\} \\ &= \sum_{i=1}^n E_\theta\left\{\frac{\partial \log f(X_i, \theta)}{\partial \theta}\right\}^2 = nI(\theta). \end{aligned} \tag{3.5.4}$$

由 Cauchy-Schwartz 不等式, 得

$$D_\theta\{\hat{g}(\mathbf{X})\} \cdot D_\theta\{S(\mathbf{X}, \theta)\} \geq [\text{Cov}_\theta(\hat{g}(\mathbf{X}), S(\mathbf{X}, \theta))]^2 = [g'(\theta)]^2,$$

将式 (3.5.4) 代入上式得

$$D_\theta[\hat{g}(\mathbf{X})] \geq \frac{[g'(\theta)]^2}{nI(\theta)}, \quad \text{一切 } \theta \in \Theta.$$

定理得证.

不等式 (3.5.1) 称为 Cramer-Rao 不等式, 简称 C-R 不等式. 因此 C-R 不等式可视为验证某一无偏估计是否为 UMVUE 的方法. 用 C-R 不等式寻找 $g(\theta)$ 的 UMVUE 时, 首先要验证样本分布族是否满足正则条件 (1)-(5) 和 (6), 然后再计算

Fisher 信息量 $I(\theta)$ 和无偏估计 $\hat{g}(\mathbf{X})$ 的方差 $D_\theta(\hat{g}(\mathbf{X}))$, 看其是否达到 C-R 下界. 验证正则条件 (1)-(5) 和条件 (6) 十分麻烦. 但幸运的是对指数族, 上述正则条件 (1)-(5) 皆成立, 而条件 (6) 正好是由定理 2.6.2 给出的指数族的一条重要性质. 因此对于指数族定理 3.5.1 的条件皆成立.

但要注意一点的是: 若 $D_\theta(\hat{g}(\mathbf{X}))$ 达不到 C-R 下界, 并不能得出结论说 $g(\theta)$ 的 UMVUE 就不存在, 而只能说用此法无法判别. 存在这样的例子, $\hat{g}(\mathbf{X})$ 是 $g(\theta)$ 的 UMVUE, 但其方差大于 C-R 下界. 后面将给出这一例子.

例 3.5.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从两点分布 $b(1, p)$ 中抽取的简单样本, 利用 C-R 不等式证明样本均值 $\bar{X} = \sum_{i=1}^n X_i/n$ 为 p 的 UMVUE.

证 设随机变量 $X \sim b(1, p)$, 则其概率分布为 $f(x, p) = p^x(1-p)^{1-x}$, $x = 0, 1$, $0 < p < 1$. 由于两点分布族是指数族, C-R 正则条件成立. Fisher 信息函数为

$$I(p) = E_p \left[\frac{\partial \log f(X, p)}{\partial p} \right]^2 = E_p \left[\frac{X - p}{p(1-p)} \right]^2 = \frac{D_p(X)}{p^2(1-p)^2} = \frac{1}{p(1-p)},$$

因此 C-R 下界为 $1/[nI(p)] = p(1-p)/n$.

而 \bar{X} 为 p 的无偏估计, 其方差 $D_p(\bar{X}) = p(1-p)/n$ 达到 C-R 下界. 故 \bar{X} 为 p 的 UMVUE.

例 3.5.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 Poisson 分布 $P(\lambda)$ 中抽取的简单样本, 用 C-R 不等式验证样本均值 $\bar{X} = \sum_{i=1}^n X_i/n$ 为 λ 的 UMVUE.

证 设随机变量 $X \sim P(\lambda)$, 则其概率分布为 $f(x, \lambda) = e^{-\lambda} \lambda^x / x!$, $x = 0, 1, 2, \dots$, $\lambda > 0$. 由于 Poisson 分布族为指数族, 故 C-R 正则条件成立. Fisher 信息函数为

$$I(\lambda) = E_\lambda \left[\frac{\partial \log f(X, \lambda)}{\partial \lambda} \right]^2 = E_\lambda \left[\frac{X - \lambda}{\lambda} \right]^2 = \frac{1}{\lambda^2} D_\lambda(X) = \frac{1}{\lambda},$$

C-R 下界为 $1/[nI(\lambda)] = \lambda/n$.

$\hat{g}(\mathbf{X}) = \bar{X}$ 为 $g(\lambda) = \lambda$ 的无偏估计, 且方差 $D_\lambda(\bar{X}) = \lambda/n$ 达到 C-R 下界, 故 \bar{X} 为 λ 的 UMVUE.

例 3.5.3 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从指数分布 $\text{Exp}(\lambda)$ 中抽取的简单样本, 用 C-R 不等式验证 $\hat{g}(\mathbf{X}) = \bar{X}$ 为 $g(\lambda) = 1/\lambda$ 的 UMVUE.

证 指数分布 $\text{Exp}(\lambda)$ 的密度函数为 $f(x, \lambda) = \lambda e^{-\lambda x} I_{(0, \infty)}(x)$, $\lambda > 0$. 指数分布族是指数族, 故 C-R 正则条件成立. Fisher 信息函数为

$$I(\lambda) = E_\lambda \left[\frac{\partial \log f(X, \lambda)}{\partial \lambda} \right]^2 = E_\lambda \left(\frac{1}{\lambda} - X \right)^2 = D_\lambda(X) = \frac{1}{\lambda^2},$$

故 C-R 下界为 $(g'(\lambda))^2 / [nI(\lambda)] = 1/(n\lambda^2)$.

而 $\hat{g}(\mathbf{X}) = \bar{X}$ 为 $g(\lambda) = 1/\lambda$ 的无偏估计, 其方差 $D_\lambda(\bar{X}) = 1/(n\lambda^2)$ 达到 C-R 下界, 故 \bar{X} 为 $g(\lambda) = 1/\lambda$ 的 UMVUE.

例 3.5.4 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 $N(a, \sigma^2)$ 中抽取的简单随机样本, 其中 σ^2 已知, 用 C-R 不等式验证 \bar{X} 为 a 的 UMVUE.

证 由于正态分布族是指数族, C-R 正则条件皆成立. 正态分布 $N(a, \sigma^2)$ 的密度函数为

$$f(x, a) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-a)^2\right\}.$$

Fisher 信息函数为

$$I(a) = E_a \left[\frac{\partial \log f(X, a)}{\partial a} \right]^2 = E_a \left[\frac{(X-a)^2}{\sigma^4} \right] = \frac{1}{\sigma^4} D_a(X) = \frac{1}{\sigma^2}.$$

故 C-R 下界为 $1/[nI(a)] = \sigma^2/n$. 而 $D_a(\bar{X}) = \sigma^2/n$ 达到 C-R 下界, 故 \bar{X} 为 a 的 UMVUE.

同样, 若本题中改 a 为已知, 但 σ^2 未知, 容易验证 $S_a^2 = \frac{1}{n} \sum_{i=1}^n (X_i - a)^2$ 为 σ^2 的 UMVUE, 这留作练习.

*2. C-R 不等式等号成立的条件

关于“C-R 不等式等号成立的条件”的主要结论概括如下:

(1) 若样本分布族非指数族, 任何 $g(\theta)$ 的任何无偏估计, 其方差不能处处(即对每一个 $\theta \in \Theta$)达到 C-R 不等式中的下界.

(2) 即使样本分布族为指数族 $f(\mathbf{x}, \theta) = C(\theta) \exp\{Q(\theta)T(\mathbf{x})\}h(\mathbf{x})$, 也不是对任何 $g(\theta)$ 都能找到无偏估计 $\hat{g}(\mathbf{X})$, 使其方差处处达到 C-R 下界. 唯有在 $g(\theta) = E_\theta(aT(\mathbf{X}) + b)$ 时才有, 即 $\hat{g}(\mathbf{X}) = aT(\mathbf{X}) + b$ 的情形才可, 此处 $a \neq 0$ 和 b 与 \mathbf{X} 无关, 但可以是 θ 的函数.

详细的介绍请查看文献 [1] 第 114-118 页. 下面是 C-R 不等式等号成立的一个例子.

例 3.5.5 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为自 Poisson 分布 $P(\lambda)$ 中抽取的简单随机样本, 证明只有 $g(\lambda)$ 是 λ 的线性函数时, 才存在 $g(\lambda)$ 的无偏估计, 其方差能处处达到 C-R 下界.

证 将样本 \mathbf{X} 的概率分布

$$f(\mathbf{x}, \lambda) = \frac{\lambda^{x_1+\dots+x_n} e^{-n\lambda}}{x_1! \cdots x_n!}$$

表成指数族的形式

$$f(\mathbf{x}, \lambda) = \frac{e^{-n\lambda} \exp\{n\bar{x} \log \lambda\}}{x_1! \cdots x_n!} = C(\lambda) \exp\{Q(\lambda)T(\mathbf{x})\}h(\mathbf{x}),$$

其中 $C(\lambda) = e^{-n\lambda}$, $Q(\lambda) = \log \lambda$, $T(\mathbf{x}) = n\bar{x} = \sum_{i=1}^n x_i$, $h(\mathbf{x}) = 1/(x_1! \cdots x_n!)$.

由“C-R 不等式等号成立的条件”结论可知：当

$$g(\lambda) = E(aT(\mathbf{X}) + b) = a \cdot n\lambda + b = c\lambda + b, \quad c, b \text{ 为常数}$$

时，无偏估计 $\hat{g}(\mathbf{X}) = aT(\mathbf{X}) + b$ 的方差处处达到 C-R 下界，且是 $g(\lambda)$ 的 UMVUE. 其他形式的参数都不存在达到 C-R 下界的无偏估计.

特别取 $g(\lambda) = \lambda$, 即 $a = 1/n$, $b = 0$, 则 $\hat{g}(\mathbf{X}) = T/n = \bar{X}$ 的方差处处达到 C-R 下界，因此它是 λ 的 UMVUE. 这与例 3.5.2 给出的结果相同.

3. Fisher 信息函数

将 C-R 不等式 (3.5.1) 中的

$$I(\theta) = E_\theta \left[\frac{\partial \log f(X, \theta)}{\partial \theta} \right]^2$$

称为 Fisher 信息函数 (或称为 Fisher 信息量). 为解释 $I(\theta)$, 不妨令 $g(\theta) = \theta$ 并假定 C-R 不等式下界 $1/(nI(\theta))$ 可达到. 这时 $nI(\theta)$ 越大, $g(\theta)$ 的无偏估计 $\hat{g}(\mathbf{X})$ 的方差越小, 表明 $g(\theta) = \theta$ 可以估计得越精. $nI(\theta)$ 与 n 和 $I(\theta)$ 成正比, n 是样本容量, 这表明若以估计量的方差的倒数作为估计量精度的指标, 则精度与 n 成正比. 比例因子, 即 $I(\theta)$, 反映总体分布的一种性质. 就是说, 总体分布的 $I(\theta)$ 越大, 意味着总体的参数越容易估计, 或者说, 该总体模型本身提供的信息量越多. 故有理由把 $I(\theta)$ 视为衡量总体模型所含信息多少的量——信息量. $I(\theta)$ 也可以解释成单个样品提供的信息量. 由于 X_1, \dots, X_n 是 i.i.d. 的, 它们的地位是平等的, 故每个样品提供同样多的信息 $I(\theta)$, 即整个样本 (X_1, \dots, X_n) 所含信息量为 $nI(\theta)$.

Fisher 信息量 $I(\theta)$ 的重要意义还在于, 在点估计大样本理论的研究中, 它起相当作用, 如在寻求 θ 的极大似然估计 $\hat{\theta}^*$ 的渐近分布 (见定理 3.3.2), $\hat{\theta}^*$ 的渐近正态分布的方差就可以用 Fisher 信息量表示, 即

$$\left\{ nE_\theta \left(\frac{\partial \log f(X, \theta)}{\partial \theta} \right)^2 \right\}^{-1} = \frac{1}{nI(\theta)}.$$

所以在定理 3.3.2 的条件下, 参数 θ 的极大似然估计 $\hat{\theta}^*$ 的渐近分布可表示为 $N(\theta, 1/[nI(\theta)])$. 这表明其渐近方差与样本的 Fisher 信息量成反比. 因此, 当 $I(\theta)$ 越大时, 渐近方差越小, 用 θ 的极大似然估计 $\hat{\theta}^*$ 来估计 θ 就越精. 由此看来, Fisher 将 $I(\theta)$ 称为信息量, 确有一定的根据.

3.5.3 多参数 C-R 不等式简介

以上讨论的都是参数 θ 为一维的情形, 对 θ 为高维的情形也可以建立类似的结果. 为此先引进一些记号. 设 $A = (a_{ij})$ 和 $B = (b_{ij})$ 是同阶的非负定方阵, 若 $A - B$ 是非负定的, 则记为 $A \geq B$. 这时必有 $a_{ii} \geq b_{ii}$, 对一切 i .

现设 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$, 总体概率函数记为 $f(x, \boldsymbol{\theta})$, $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体中抽取的简单随机样本. 设 $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{X}) = (\hat{\theta}_1, \dots, \hat{\theta}_k)$ 是 $\boldsymbol{\theta}$ 的一个无偏估计. 以 $\text{Cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ 记 $\hat{\boldsymbol{\theta}}$ 的协方差阵, 它是一个 k 阶非负定方阵, 其 (i, j) 元为 $E_{\boldsymbol{\theta}}[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)]$, 则在类似于 $\boldsymbol{\theta}$ 为一维的正则条件下, 可以证明

$$\text{Cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \geq (n\mathbf{I}(\boldsymbol{\theta}))^{-1}, \quad (3.5.5)$$

其中 $\mathbf{I}(\boldsymbol{\theta}) = (I_{ij}(\boldsymbol{\theta}))$ 是一个 k 阶正定方阵, 且

$$I_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \left[\left(\frac{\partial \log f(\mathbf{X}, \boldsymbol{\theta})}{\partial \theta_i} \right) \left(\frac{\partial \log f(\mathbf{X}, \boldsymbol{\theta})}{\partial \theta_j} \right) \right], \quad i, j = 1, 2, \dots, k, \quad (3.5.6)$$

则式 (3.5.5) 就是多维的 C-R 不等式. 若记 $(\mathbf{I}(\boldsymbol{\theta}))^{-1} = \mathbf{I}^*(\boldsymbol{\theta}) = (I_{ij}^*(\boldsymbol{\theta}))$, 故由式 (3.5.5) 可得

$$D_{\boldsymbol{\theta}}(\hat{\theta}_i) \geq \frac{I_{ii}^*(\boldsymbol{\theta})}{n}, \quad i = 1, 2, \dots, k. \quad (3.5.7)$$

这给出了 $\boldsymbol{\theta}$ 的每个分量 θ_i 的无偏估计 $\hat{\theta}_i$ 的方差的下限.

例 3.5.6 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从正态总体 $N(a, \sigma^2)$ 中抽取的简单样本, 记 $\boldsymbol{\theta} = (a, \sigma^2)$, 其中 $\theta_1 = a$, $\theta_2 = \sigma^2$. 求 $\boldsymbol{\theta}$ 的 C-R 下界, 并将其与 θ_1 和 θ_2 的无偏估计 \bar{X} 和 S^2 的方差进行比较.

解 正态随机变量的密度函数为

$$f(x, \boldsymbol{\theta}) = (2\pi\theta_2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\theta_2}(x - \theta_1)^2 \right\},$$

可知

$$\frac{\partial \log f(x, \boldsymbol{\theta})}{\partial \theta_1} = \frac{x - \theta_1}{\theta_2}, \quad \frac{\partial \log f(x, \boldsymbol{\theta})}{\partial \theta_2} = \frac{-\theta_2 + (x - \theta_1)^2}{2\theta_2^2}.$$

由此算出

$$I_{11}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}, \quad I_{22}(\boldsymbol{\theta}) = \frac{1}{2\sigma^4}, \quad I_{12}(\boldsymbol{\theta}) = I_{21}(\boldsymbol{\theta}) = 0,$$

故有

$$n\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}, \quad (n\mathbf{I}(\boldsymbol{\theta}))^{-1} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix},$$

若记

$$\hat{\theta}_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\theta}_2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

则由多维 C-R 不等式可知

$$\text{Cov}_{\theta}(\hat{\theta}) = \text{Cov}_{\theta} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \geq \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}.$$

而 $D_{\theta}(\hat{\theta}_1) = \sigma^2/n$ 达到 C-R 下界, 故它是 $\theta_1 = a$ 的 UMVUE. 利用 $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$ 及 $D_{\theta}[(n-1)S^2/\sigma^2] = 2(n-1)$ 可得

$$D_{\theta}(\hat{\theta}_2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}.$$

因此, $S^2 = \hat{\theta}_2$ 的方差大于 C-R 下界. 在例 3.4.11 中已证明了 \bar{X} 和 S^2 分别是 a 和 σ^2 的 UMVUE.

由本例看出, 即使在正态总体方差这样简单的场合, 方差 σ^2 的 UMVUE S^2 也达不到 C-R 下界. 因此在一些问题中 C-R 下界常比真下界为小. 这表明以 C-R 不等式作为寻找 UMVUE 的方法是不够理想的. 多年来有一些学者一直在研究改进这个不等式的问题.

3.5.4 有效估计和估计的效率

定义 3.5.2 设 $\hat{g}(\mathbf{X})$ 为 $g(\theta)$ 的无偏估计, 比值

$$e_{\hat{g}}(\theta) = \frac{[g'(\theta)]^2 / (nI(\theta))}{D_{\theta}[\hat{g}(\mathbf{X})]}$$

称为无偏估计 $\hat{g}(\mathbf{X})$ 的效率 (efficiency). 显然 $0 < e_{\hat{g}}(\theta) \leq 1$, 当 $e_{\hat{g}}(\theta) = 1$ 时, 称 $\hat{g}(\mathbf{X})$ 是 $g(\theta)$ 的有效估计 (effective estimation). 若 $\hat{g}(\mathbf{X})$ 不是 $g(\theta)$ 的有效估计, 但 $\lim_{n \rightarrow \infty} e_{\hat{g}}(\theta) = 1$, 则称 $\hat{g}(\mathbf{X})$ 是 $g(\theta)$ 的渐近有效估计 (asymptotically effective estimation).

这一概念有其不足之处: 有效估计是无偏估计类中最好的估计, 当然希望使用它. 可惜, 有效估计是不多的, 但渐近有效估计却不少. 从有效估计的定义看, 有效估计一定是 UMVUE, 但很多 UMVUE 不是有效估计. 这是因为 C-R 下界偏小, 在很多场合 UMVUE 的方差达不到 C-R 下界. 另外 C-R 不等式成立的前提要求样本分布族满足 C-R 正则条件. 当这些条件不成立时, C-R 不等式可以不对, 这时依据它所提供的 C-R 下界去定义估计的效率或求有效估计就不合理了.

例 3.5.7 由例 3.5.1—例 3.5.4 给出的有关参数的无偏估计, 其方差都能达到 C-R 下界, 因此它们都是相应参数的有效估计.

例 3.5.8 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 $N(a, \sigma^2)$ 中抽取的简单随机样本,

(1) 当 a 未知时, 证明样本方差 S^2 不是 σ^2 的有效估计, 但是渐近有效估计.

(2) 当 a 已知时, 求 σ^2 的有效估计.

解 (1) 由例 3.5.6 已经知道, 当 a 未知时, $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$ 之方差为 $2\sigma^4/(n-1)$ 达不到 C-R 下界 $2\sigma^4/n$, 故它不是 σ^2 的有效估计. 估计的效率为 $e_{S^2}(\sigma^2) = (n-1)/n < 1$, 但是

$$\lim_{n \rightarrow \infty} e_{S^2}(\sigma^2) = \lim_{n \rightarrow \infty} \frac{n-1}{n} = 1,$$

因此 S^2 是 σ^2 的渐近有效估计.

(2) 由于 a 已知, 令 $S_a^2 = \frac{1}{n} \sum_{i=1}^n (X_i - a)^2 / n$, 由于 $nS_a^2/\sigma^2 \sim \chi_n^2$, 故 $D(nS_a^2/\sigma^2) = 2n$, 因此有 $D(S_a^2) = 2\sigma^4/n$, 它达到 C-R 下界, 故 S_a^2 为 σ^2 的有效估计. 当 a 已知时, 利用 Lehmann-Scheffe 定理也容易证明 S_a^2 为 σ^2 的 UMVUE.

例 3.5.9 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从下列含有位置参数的指数分布族中抽取的简单样本,

$$f(x, a) = e^{-(x-a)} I_{(a, \infty)}(x), \quad -\infty < a < +\infty,$$

求 a 的 UMVUE.

解 上述分布族不是 C-R 正则族, 即 C-R 正则条件不成立. 因为密度函数的支撑集 $\{x : f(x, a) > 0\} = \{x : x > a\}$ 与未知参数 a 有关, 因此它至少不满足 C-R 正则条件 (2). 因此不能用 C-R 不等式来求 a 的 UMVUE.

显然, 样本的最小次序统计量 $X_{(1)}$ 是 a 的充分完全统计量 (见习题 2, 第 49 题). $Y = X_{(1)}$ 的密度函数为

$$f(y, a) = ne^{-n(y-a)} I_{[a, \infty)}(y).$$

于是

$$E(X_{(1)}) = n \int_a^\infty y e^{-n(y-a)} dy = a + \frac{1}{n},$$

故由 L-S 定理可知 $\hat{a}(\mathbf{X}) = X_{(1)} - 1/n$ 是 a 的 UMVUE; 但不可以讨论 $\hat{a}(\mathbf{X})$ 是否为有效估计, 因为 C-R 正则条件不成立, 去讨论 C-R 下界就失去了意义.

*3.6 概率密度函数的核估计

概率密度函数的估计方法有多种, 其中使用最广泛的是“直方图”法. 7.2 节将介绍如何利用直方图估计先验密度, 对一般的概率密度函数的估计同样适用, 故

本节略去对这一方法的介绍。概率密度函数估计的其他方法有核估计方法和近邻估计方法等，本节只介绍概率密度函数的核估计方法。概率密度函数的核估计方法与经验分布函数有关。关于经验分布函数及其性质已在 1.3 节中介绍，此处不再重复。

3.6.1 概率密度函数的核估计

1. 概率密度函数的“自然”估计

在介绍密度函数的核估计定义之前，先介绍密度函数的一种“自然”估计，然后将其与密度函数的核估计建立联系，以便读者更好地理解核估计方法的思想。

设随机变量 X 的分布函数和密度函数分别为 $F(x)$ 和 $f(x)$ 。若 $f(x)$ 连续，则

$$f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}.$$

当 $h = h_n$ 充分小时近似有 $f(x) \approx \frac{F(x+h_n) - F(x-h_n)}{2h_n}$ ，将其中 $F(\cdot)$ 用经验分布函数 $F_n(\cdot)$ 代替就得到

$$f_n(x) = \frac{F_n(x+h_n) - F_n(x-h_n)}{2h_n}, \quad (3.6.1)$$

则 $f_n(x)$ 就称为 $f(x)$ 的一个“自然”估计。

2. 概率密度函数核估计的定义

下面我们将导出概率密度函数“自然”估计与核估计的关系。令

$$K(x) = \begin{cases} 1/2, & x \in (-1, 1] \\ 0, & \text{其他} \end{cases} = \frac{1}{2} I_{(-1,1]}(x)$$

为一核函数，此处 $I_A(x)$ 为示性函数。我们知道经验分布函数可以通过下式定义：

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i).$$

其中 X_1, \dots, X_n 为从总体 X 中抽取的 i.i.d. 样本。 $f(x)$ 的核估计与其“自然”估计的联系如下：

$$\begin{aligned} f_n(x) &= \frac{F_n(x+h_n) - F_n(x-h_n)}{2h_n} \\ &= \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} \left\{ I_{(-\infty, x+h_n]}(X_i) - I_{(-\infty, x-h_n]}(X_i) \right\} \\ &= \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} \left\{ I_{(-1,1]} \left(\frac{x-X_i}{h_n} \right) \right\} = \frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{x-X_i}{h_n} \right). \end{aligned}$$

此处 $K(\cdot)$ 为核函数. 将上述思想加以推广得到如下定义.

定义 3.6.1 概率密度函数 $f(x)$ 的下述估计量

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \quad (3.6.2)$$

称为核估计 (kernel estimation), 此处 h_n 称为窗宽, 满足条件: $0 < h_n \rightarrow 0$, 当 $n \rightarrow \infty$, 而核函数 $K(\cdot)$ 通常是一个适当概率密度函数.

由定义可见概率密度函数“自然”估计是核估计的一个特例.

设 $K(\cdot)$ 为定义于 $R = (-\infty, \infty)$ 上的核函数, 通常假定 $K(\cdot)$ 满足下列条件:

$$(1) \sup_{x \in R} \{K(x)\} \leq M < \infty, \quad \lim_{|x| \rightarrow \infty} |x|K(x) = 0.$$

$$(2) K(x) = K(-x), \quad x \in R, \text{ 即 } K(x) \text{ 对称, 且 } \int_{-\infty}^{\infty} x^2 K(x) dx < \infty.$$

(3) $\hat{K}(u)$ 是绝对可积的, $\hat{K}(u)$ 为 $K(\cdot)$ 的特征函数.

适合条件 (1)-(3) 的 $K(\cdot)$ 有下列几个例子:

$$(i) K_1(x) = \begin{cases} 1/2, & -1 < x \leq 1, \\ 0, & \text{其他;} \end{cases}$$

$$(ii) K_2(x) = \begin{cases} 1 - |x|, & |x| \leq 1, \\ 0, & |x| > 1; \end{cases}$$

(iii) $K_3(x) = (2\pi)^{-1/2} \exp\{-x^2/2\}$, $x \in R$, 这是标准正态分布的密度;

(iv) $K_4(x) = [\pi(1+x^2)]^{-1}$, $x \in R$, 这是柯西分布 $C(0, 1)$ 的密度;

$$(v) K_5(x) = \begin{cases} \frac{1}{2\pi} \left[\frac{\sin(x/2)}{x/2} \right]^2, & \text{若 } x \neq 0, \\ \frac{1}{2\pi}, & \text{若 } x = 0; \end{cases}$$

$$(vi) K_6(x) = \begin{cases} \frac{3}{4}\lambda^{-3}(\lambda^2 - x^2), & \text{若 } x^2 \leq \lambda^2, \\ 0 & \text{若 } x^2 > \lambda^2, \lambda > 0. \end{cases}$$

当然还有其他形式的核函数, 它们不必为概率密度函数, 但通常都要满足适当的条件.

3.6.2 概率密度函数导数的核估计

求概率密度函数 $f(x)$ 的 p 阶导函数核估计的最简单的方法, 就是将 $f(x)$ 的核估计 $f_n(x)$ 对 x 求 p ($p = 1, 2, \dots$) 阶导数就得到相应的核估计, 即令

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)$$

为 $f(x)$ 的核估计, 对 $f_n(x)$ 关于 x 求 p 阶导数得

$$f_n^{(p)}(x) = \frac{1}{nh_n^{1+p}} \sum_{i=1}^n K^{(p)}\left(\frac{x-X_i}{h_n}\right),$$

其中 $K^{(p)}(\cdot)$ 为核函数 $K(\cdot)$ 的 p 阶导数, 将 $K^{(p)}(\cdot)$ 看成新的核函数并记为 $K_p(\cdot)$, $p=1, 2, \dots$. 可见, 概率密度函数的导函数 $f^{(p)}(x)$ 的核估计与密度函数 $f(x)$ 的核估计的定义无本质的差别. 现定义如下.

定义 3.6.2 设 $K_j(\cdot)$ ($j=0, 1, 2, \dots$) 为一列核函数, 则称

$$f_n^{(j)}(x) = \frac{1}{nh_n^{1+j}} \sum_{i=1}^n K_j\left(\frac{x-X_i}{h_n}\right), \quad j=0, 1, \dots \quad (3.6.3)$$

为 $f^{(j)}(x)$ 的核估计. 此处 $0 < h_n \rightarrow 0$, 当 $n \rightarrow \infty$.

在上述定义中, 特别当 $j=0$ 时 $f^{(0)}(x) = f(x)$, $f_n^{(0)}(x) = f_n(x)$, 即

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K_0\left(\frac{x-X_i}{h_n}\right)$$

为 $f(x)$ 的核估计. 也就是说概率密度函数及其导函数的核估计的表达式可统一由式 (3.6.3) 给出.

3.6.3 概率密度函数核估计的大样本性质

1. 若干定义

首先给出比较不同估计量的大样本性质的优良性准则.

定义 3.6.3 设 \mathcal{F} 为一元概率密度族, X_1, \dots, X_n, \dots 为从 \mathcal{F} 抽取的 i.i.d. 随机变量序列, 它们具有共同的密度函数 $f(x)$, 令 $f_n(x)$ 为密度函数 $f(x)$ 的核估计. 若对样本空间 \mathcal{X} 中的每一个 x 和一切 $f \in \mathcal{F}$ 有

- (1) 若 $\lim_{n \rightarrow \infty} E[f_n(x)] = f(x)$, 则称 $f_n(x)$ 为 $f(x)$ 的渐近无偏估计.
- (2) 若 $\lim_{n \rightarrow \infty} E[f_n(x) - f(x)]^2 = 0$, 则称 $f_n(x)$ 为 $f(x)$ 的均方相合估计.
- (3) 当 $n \rightarrow \infty$ 时, 若 $f_n(x) \xrightarrow{P} f(x)$, 则称 $f_n(x)$ 为 $f(x)$ 的弱相合估计.
- (4) 当 $n \rightarrow \infty$ 时, 若 $f_n(x) \xrightarrow{\text{a.s.}} f(x)$, 则称 $f_n(x)$ 为 $f(x)$ 的强相合估计.

2. 大样本性质

下面仅给出概率密度函数核估计的大样本性质, 有关这些结果的证明可在文献 [18] 和 [19] 中找到, 故在此略去.

设核函数 $K(\cdot)$ 满足下列条件 (I):

- (a) $K(\cdot)$ 有界;
 (b) $\int_{-\infty}^{\infty} |K(u)| du < \infty$;
 (c) $|u| |K(u)| \rightarrow 0$, 当 $|u| \rightarrow \infty$;

关于 $f_n(x)$ 的渐近无偏性有下列结果.

定理 3.6.1 若 $K(\cdot)$ 满足条件 (I) 中的 (a)-(b), 且 $f(\cdot)$ 在 x 处连续, $\lim_{n \rightarrow \infty} h_n = 0$, 则有

$$\lim_{n \rightarrow \infty} E[f_n(x)] = f(x).$$

又若 $f(x)$ 一致连续, 则上述结果关于 x 一致成立.

关于 $f_n(x)$ 的均方相合性和弱相合性有下列结果.

定理 3.6.2 若 $K(\cdot)$ 满足条件 (I) 中的 (a)-(c), $f(\cdot)$ 在 x 处连续, 且 $\lim_{n \rightarrow \infty} h_n = 0$, $\lim_{n \rightarrow \infty} nh_n = \infty$, 则有

- (1) $E|f_n(x) - f(x)|^2 \rightarrow 0$, 当 $n \rightarrow \infty$.
 (2) $f_n(x) \xrightarrow{P} f(x)$, 当 $n \rightarrow \infty$.

关于 $f_n(x)$ 的强相合性有下列结果.

定理 3.6.3 若 $K(\cdot)$ 满足条件 (I) 中的 (a)-(c), $f(\cdot)$ 在 x 连续, 则有

$$f_n(x) \xrightarrow{\text{a.s.}} f(x), \quad \text{当 } n \rightarrow \infty.$$

关于 $f_n(x)$ 的渐近正态性有下列结果.

定理 3.6.4 若 $K(\cdot)$ 满足条件 (I) 中的 (a)-(c), 且 $\lim_{n \rightarrow \infty} h_n = 0$, $\lim_{n \rightarrow \infty} nh_n = \infty$, 则有

$$\frac{f_n(x) - E[f_n(x)]}{\sqrt{\text{Var}(f_n(x))}} \xrightarrow{\mathcal{L}} N(0, 1), \quad \text{当 } n \rightarrow \infty.$$

注 3.6.1 概率密度函数核估计的大样本性质对密度函数导函数的核估计同样也成立, 由定义 3.6.2 可知只要相应的核函数满足类似于条件 (I) 中 (a)-(c) 即可.

除上述大样本性质外, 各种相合性还具有一定的收敛速度; 概率密度函数的估计方法除核估计方法外还有最近邻估计方法等. 对这些内容感兴趣的读者可查看文献 [20] 第二章和文献 [18].

例 3.6.1 设 X_1, \dots, X_n i.i.d $\sim f(x)$. 概率密度函数 $f(x)$ 的核估计 $f_n(x)$ 的定义由式 (3.6.2) 给出, 在定理 3.6.1 的条件下证明 $f_n(x)$ 的渐近无偏性.

证 由 $f_n(x)$ 定义可知

$$\begin{aligned} E[f_n(x)] &= \frac{1}{h_n} E\left[K\left(\frac{x - X_1}{h_n}\right)\right] = \frac{1}{h_n} \int_{-\infty}^{\infty} K\left(\frac{x - y}{h_n}\right) f(y) dy \\ &= \int_{-\infty}^{\infty} K(u) f(x - h_n u) du. \end{aligned} \tag{3.6.4}$$

由于概率密度函数在数轴 R 上有界, 故有 $|K(u)f(x - h_n u)| \leq c|K(u)|$, 由定理 3.6.1 条件(I) 中 (b) 可知 $\int |K(u)|du < \infty$. 故积分 (3.6.4) 关于 $x \in R$ 上一致收敛, 由参数变量积分的性质可知积分 (3.6.4) 在 R 上关于变量 x 连续. 因此有

$$\begin{aligned}\lim_{n \rightarrow \infty} (E[f_n(x)] - f(x)) &= \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} K(u)[f(x - h_n u) - f(x)]du \\ &= \int_{-\infty}^{\infty} K(u) \lim_{h_n \rightarrow 0} [f(x - h_n u) - f(x)]du = 0,\end{aligned}$$

即证明了 $\lim_{n \rightarrow \infty} E[f_n(x)] = f(x)$, $f_n(x)$ 的渐近无偏性成立.

习 题 3

- 设 X_1, \dots, X_n i.i.d. $\sim N(a, \sigma^2)$, 样本方差 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$ 是 σ^2 的无偏估计. 证明 S^2 是 σ^2 的强相合估计和均方相合估计.
- 证明 $\hat{\theta}_n$ 是 θ 的均方相合估计的充要条件为: $\hat{\theta}_n$ 是渐近无偏的且 $\lim_{n \rightarrow \infty} D_\theta(\hat{\theta}_n) = 0$, 对任何 $\theta \in \Theta$.
- 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从某总体中抽取的简单样本, 满足 $E(X_1) = \mu < \infty$, $E(X_1^2) < \infty$, 证明 $T(\mathbf{X}) = 2 \sum_{i=1}^n iX_i / [n(n+1)]$ 是 μ 的弱相合估计.
- 设 $X_n \xrightarrow{\text{a.s.}} X$, $g(x)$ 为 x 的连续函数, 则 $g(X_n) \xrightarrow{\text{a.s.}} g(X)$.
- 设总体 X 服从均匀分布 $U(0, 2\theta)$, 令 X_1, \dots, X_n 为从总体 X 中抽取的简单样本,
 - 证明 $\hat{\theta}_1 = \bar{X}$ 和 $\hat{\theta}_2 = (n+1)X_{(n)} / (2n)$ 为 θ 的无偏估计.
 - 证明 $\hat{\theta}_1$ 为 θ 的强相合估计, $\theta_2^* = X_{(n)} / 2$ 为 θ 的弱相合估计.
 - 求 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 的方差, 问哪一个更有效?
- 设总体 X 服从二项分布 $b(k, p)$, k 是正整数, $0 < p < 1$, 两者都是未知参数, X_1, \dots, X_n 为从中抽取的简单样本, 求 k 和 p 的矩估计量.
- 设 X_1, \dots, X_n 为抽自对数级数分布

$$P(X = k) = -\frac{1}{\ln(1-p)} \cdot \frac{p^k}{k}, \quad 0 < p < 1, \quad k = 1, 2, \dots$$

的随机样本, 求参数 p 的矩估计量.

- 设样本 X_1, \dots, X_n 抽自正态分布 $N(0, \sigma^2)$, 求 σ 的矩估计量:
 - 利用 $E|X_1| = \sqrt{\frac{2}{\pi}}\sigma$;
 - 利用 $\sigma = \sqrt{D(X_1)}$.
- 设 X_1, \dots, X_n 是来自正态总体 $N(a, \sigma^2)$ 的简单样本, 求 $P(X > 1)$ 的矩估计量.
- 设 X_1, \dots, X_n i.i.d. \sim 伽马分布 $\Gamma(r, \lambda)$, 其中 r 已知, 求 λ 矩估计量, 并讨论它的无偏性.

11. 若 $X = e^\xi$, 而 $\xi \sim N(a, \sigma^2)$, 则 r.v. X 的分布称为对数正态分布. 求出 X 的密度函数. 设 X_1, \dots, X_n 是从总体 X 抽取的简单随机样本, 求 a 和 σ^2 的矩估计和极大似然估计.
12. 设 X_1, \dots, X_n 是从总体 X 中抽取的简单样本, X 的分布为下列之一, 试分别用矩估计法和极大似然法求 θ 的估计量:

$$(1) f(x, \theta) = \begin{cases} \theta(\theta+1)x^{\theta-1}(1-x), & 0 < x < 1, \theta > 0, \\ 0, & \text{其他;} \end{cases}$$

$$(2) f(x, \theta) = (\theta+1)x^\theta, \quad 0 < x < 1, \theta > -1;$$

$$(3) f(x, \theta) = \begin{cases} \frac{\theta^{-r}}{\Gamma(r)}x^{r-1}\exp\left\{-\frac{x}{\theta}\right\}, & x > 0, \quad r > 0 \text{ 已知.} \\ 0, & \text{其他,} \end{cases}$$

13. 设总体 X 的密度为 $\frac{1}{2\sigma}\exp\{-|x-a|/\sigma\}$, 其中 $\sigma > 0$ 和 $-\infty < a < \infty$ 为未知参数. 设 X_1, \dots, X_n 为抽自此总体的简单随机样本, 求 a 和 σ 的矩估计和极大似然估计.

14. 设随机变量 X 的密度函数为

$$f(x, \theta) = \begin{cases} \frac{x}{\theta^2}\exp\left\{-\frac{x^2}{2\theta^2}\right\}, & x > 0, \quad \theta > 0. \\ 0, & x \leq 0, \end{cases}$$

设 X_1, \dots, X_n 为从总体 X 中抽取的随机样本, 求 θ 的 MLE.

15. 设 X_1, \dots, X_n 是来自双参数指数分布

$$f(x, \mu, \sigma) = \frac{1}{\sigma}\exp\left\{-\frac{x-\mu}{\sigma}\right\}, \quad x \geq \mu$$

的简单随机样本, 其中 $-\infty < \mu < +\infty, \sigma > 0$. 记 $\theta = (\mu, \sigma)$, 求 μ, σ 和 $P_\theta(X_1 \geq t)$ ($t > \mu$) 的矩估计和极大似然估计.

16. 设总体 X 服从 Weibull 分布, 密度函数为

$$f(x, \lambda) = \begin{cases} \lambda\alpha \cdot x^{\alpha-1}e^{-\lambda x^\alpha}, & x > 0, \quad \lambda > 0, \alpha > 0. \\ 0, & \text{其他,} \end{cases}$$

设 X_1, \dots, X_n 为此总体中抽取的简单样本. 若 α 已知, 求 λ 的 MLE.

17. 设 X_1, \dots, X_n 为从总体 X 中抽取的随机样本, X 服从均匀分布 $U(\theta - 1/2, \theta + 1/2)$, 求 θ 的 MLE.
18. 设 X_1, \dots, X_n 是来自均匀分布 $U(\theta, 2\theta)$ 的简单随机样本, 其中 $0 < \theta < +\infty$, 求 θ 的 MLE, 它是 θ 的无偏估计吗? 如果不是, 试对它略作修改, 得到 θ 的一个无偏估计.
19. 设 X_1, \dots, X_m 和 Y_1, \dots, Y_n 分别来自总体 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$ 的两组独立样本, 求 μ_1, μ_2 和 σ^2 的 MLE.
20. 为了估计湖中有多少条鱼, 从中捞出 1000 条, 标上记号后放回湖中, 然后再捞出 150 条鱼, 发现其中有 10 条鱼有记号. 问湖中有多少条鱼, 才能使 150 条鱼中出现 10 条带记号的鱼的概率最大?

21. 一个罐子中装有黑白两种球, 今有放回地抽取一个大小为 n 的样本, 其中有 k 个白球, 求罐中黑白球之比的极大似然估计.
22. 设 X_1, \dots, X_n 为抽自下列指数分布的简单样本:

$$f(x, \theta) = \begin{cases} e^{-(x-\mu)}, & x > \mu, \\ 0, & \text{其他}, \end{cases}$$

其中 $-\infty < \mu < +\infty$.

- (1) 试求 μ 的极大似然估计 $\hat{\mu}^*$, $\hat{\mu}^*$ 是 μ 的无偏估计吗? 如果不是, 试对它作修改, 以得到 μ 的无偏估计 $\hat{\mu}^{**}$;
- (2) 试求 μ 的矩估计 $\hat{\mu}$, 并证明它是 μ 的无偏估计;
- (3) 试问 $\hat{\mu}^{**}$ 和 $\hat{\mu}$ 哪一个有效?
23. 设 X_1, \dots, X_n i.i.d. \sim 均匀分布 $U(0, \theta)$, 求 θ 的 MLE $\hat{\theta}_n$, 证明 $\hat{\theta}_n$ 是 θ 的强相合估计和 r ($r > 0$) 阶矩相合估计.(提示: 证强相合估计要用到 Borel-Contelli 引理: 证明对任给 $\varepsilon > 0$, 有 $\sum_{n=1}^{\infty} P(|\hat{\theta}_n - \theta| \geq \varepsilon) < \infty$.)
24. 设 X_1, \dots, X_n i.i.d. $\sim N(\sigma, \sigma^2)$, $\sigma > 0$, 试求 σ^2 的 MLE $\hat{\sigma}^2$, 并证明 $\hat{\sigma}$ 为 σ 的弱相合估计.

25. 设总体 X 为下列 0-1 分布族

$$P_{\theta}(X = 1) = 1 - P_{\theta}(X = 0) = \begin{cases} \theta, & \theta \text{ 为有理数,} \\ 1 - \theta, & \theta \text{ 为无理数.} \end{cases}$$

设 X_1, \dots, X_n 为从总体 X 中抽取的简单样本, 证明 θ 的 MLE 不是相合估计.

- *26. 设 X_1, \dots, X_n 是来自伽马分布 $\Gamma(\alpha, \lambda)$ 的简单样本, 试证明

(1) 当 $\lambda = 1$ 时, 记 α 的 MLE 为 $\hat{\alpha}$, 则

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{\mathcal{L}} N\left(0, \left[E\left(\frac{\partial \log g(x, \alpha)}{\partial \alpha}\right)^2\right]^{-1}\right),$$

此处 $g(x, \alpha)$ 是参数分别为 α , $\lambda = 1$ 的伽马分布的密度函数;

(2) 当 α 已知时, λ 的 MLE $\hat{\lambda}$ 近似服从 $N(\lambda, \lambda^2/(n\alpha))$, 即 $\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{\mathcal{L}} N(0, \lambda^2/\alpha)$.

27. 考察均匀分布族 $\{U(0, \theta) : \theta > 0\}$, 则不管样本容量 n 有多大, $g(\theta) = 1/\theta$ 不是可估参数. 试以 $n = 1$ 为例, 证明这个结论.
28. 设 X_1, \dots, X_n i.i.d. $\sim N(\theta, 1)$, 证明 $g(\theta) = |\theta|$ 没有无偏估计, 即 $g(\theta)$ 不是可估参数 (提示: 利用 $g(\theta)$ 在 $\theta = 0$ 处不可导).
29. 试证明定理 3.4.1 逆成立: 若 \hat{g} 为 $g(\theta)$ 的 UMVUE, $D_{\theta}(\hat{g}) < \infty$ 对任何 $\theta \in \Theta$, 则当 $E_{\theta}(\hat{l}) = 0$ 且 $D_{\theta}(\hat{l}) < \infty$ 对一切 $\theta \in \Theta$ 成立时, 必有 $\text{Cov}_{\theta}(\hat{g}, \hat{l}) = 0$ 对一切 $\theta \in \Theta$ (提示: 用反证法. 若 $\text{Cov}_{\theta}(\hat{g}, \hat{l}) \neq 0$ 对某个 $\theta \in \Theta$, 则存在常数 δ , $|\delta|$ 充分小, 使 $D_{\theta}(\hat{g} + \delta\hat{l}) < D_{\theta}(\hat{g})$, 因为 $\hat{g} + \delta\hat{l}$ 为无偏估计, 得到矛盾).
30. 设 X_1, \dots, X_n i.i.d. $\sim N(a, \sigma^2)$, 用零无偏估计法证明 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ 为 σ^2 的 UMVUE.

31. 设 X_1, \dots, X_n i.i.d. $\sim N(0, \sigma^2)$, $\sigma^2 > 0$ 为未知参数, 试求
 (1) σ^2 的充分完全统计量;
 (2) σ 和 $3\sigma^4$ 的 UMVUE.
32. 设 X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma^2)$, $-\infty < \mu < +\infty$, $\sigma^2 > 0$ 都是未知参数, 试求参数函数
 (1) $3\mu + 4\sigma^2$;
 (2) $\mu^2/(4\sigma^2)$ 的 UMVUE.
33. 设 X_1, \dots, X_n i.i.d. 服从两点分布 $b(1, p)$, $0 < p < 1$ 是未知参数, 试求
 (1) p^s 的 UMVUE;
 (2) $p^s + (1-p)^{n-s}$ 的 UMVUE ($0 < s < n$ 为整数).
34. 设 X_1, \dots, X_m i.i.d. $\sim N(a, \sigma^2)$, Y_1, \dots, Y_n i.i.d. $\sim N(a, 2\sigma^2)$, 且两组样本独立, 求 a 和 σ^2 的 UMVUE.
35. 设 X_1, \dots, X_n i.i.d. \sim 几何分布族

$$P(X_1 = i) = \theta(1 - \theta)^{i-1}, \quad i = 1, 2, \dots, 0 < \theta < 1.$$

(1) 试证明 $T = \sum_{i=1}^n X_i$ 是 θ 的充分完全统计量, 且服从帕斯卡(负二项)分布

$$P_\theta(T = t) = \binom{t-1}{n-1} \theta^n (1-\theta)^{t-n}, \quad t = n, n+1, n+2, \dots$$

(2) 计算 $E_\theta T$, 并由此求 θ^{-1} 的 UMVUE.

(3) 试证明

$$\psi(X_1) = \begin{cases} 1, & X_1 = 1, \\ 0, & X_1 = 2, 3, \dots \end{cases}$$

是 θ 的无偏估计, 计算 $E_\theta(\psi(X_1)|T=t)$, 并由此求得 θ 的 UMVUE.

36. 设有分布族 $\{\theta e^{-\theta x} I_{(0, \infty)}(x), \theta > 0\}$, X_1, \dots, X_n 为从中抽取的简单样本, 求 $e^{-\theta\tau}$ 的 UMVUE (注意: $e^{-\theta\tau} = P_\theta(X_1 > \tau)$), 此处 $\tau > 0$ 为给定的数.

- *37. 设 X_1, \dots, X_n 是来自均匀分布总体 $U(\theta_1, \theta_2)$ 的简单随机样本.

- (1) 试求 θ_1 和 θ_2 的 MLE $\hat{\theta}_1$ 和 $\hat{\theta}_2$;
- (2) $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别是 θ_1 和 θ_2 的无偏估计吗? 如果不是, 请加以修正, 以获得 θ_1 和 θ_2 的无偏估计 (提示: 确定 a 和 b , 使得 $a\hat{\theta}_1 + b\hat{\theta}_2$ 是 θ_1 (或 θ_2) 的无偏估计);
- (3) 若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 为充分完全统计量, 由 (2) 得到的无偏估计是否分别为 θ_1 和 θ_2 的 UMVUE?
- (4) 试分别求出中点 $(\theta_1 + \theta_2)/2$ 和变程 $\theta_2 - \theta_1$ 的 UMVUE.
38. 设 X_1, \dots, X_n i.i.d. $\sim N(\theta, 1)$, 求 θ^2 的 UMVUE. 证明此 UMVUE 达不到 C-R 不等式的下界, 即它不是有效估计.

39. 证明均匀分布族 $\mathcal{F} = \{U(0, \theta) : 0 < \theta < \infty\}$ 不是 C-R 正则分布族.

40. 设 X_1, \dots, X_n 为自下列总体中抽取的简单样本,

$$f(x, \theta) = \begin{cases} \theta^{-1} e^{-x/\theta}, & x > 0, \\ 0, & \text{其他,} \end{cases}$$

其中 $\theta > 0$ 为未知参数, 求 θ 的无偏估计的方差下限和 θ 的 UMVUE, 并比较 θ 的 UMVUE 方差与 C-R 下界.

41. 设 X_1, \dots, X_n 是来自伽马分布族 $\{\Gamma(\alpha, \lambda) : \alpha \text{ 已知, } \lambda > 0\}$ 的简单随机样本. 试证: \bar{X}/α 是 $g(\lambda) = 1/\lambda$ 的有效估计.

42. 设 X_1, \dots, X_n i.i.d. $\sim N(0, \sigma^2)$, $\sigma^2 > 0$ 为未知参数,

(1) 求 σ^2 的无偏估计方差的 C-R 下界;

(2) 求 σ^2 的一致最小方差无偏估计及它的效率.

43. 设 X_1, \dots, X_n i.i.d. $\sim N(a, \sigma^2)$, a 已知, 证明 $\frac{1}{n} \sqrt{\frac{\pi}{2}} \sum_{i=1}^n |X_i - a|$ 为 σ 的无偏估计量, 且有效率为 $1/(\pi - 2)$.

44. 设 X_1, \dots, X_n i.i.d. $\sim N(0, \sigma^2)$, 证明

$$\hat{\sigma} = \frac{\Gamma(n/2)}{\sqrt{2}\Gamma((n+1)/2)} \left(\sum_{i=1}^n X_i^2 \right)^{1/2}$$

是 σ 的 UMVUE, 试求其效率.

*45. 设 X_1, \dots, X_n 自下列总体中抽取的简单样本,

$$f(x, \theta) = \begin{cases} 1, & \theta - \frac{1}{2} \leqslant x \leqslant \theta + \frac{1}{2}, \\ 0, & \text{其他,} \end{cases} \quad -\infty < \theta < +\infty.$$

证明样本均值 \bar{X} 及 $\frac{1}{2} \left(\max_{1 \leqslant i \leqslant n} X_i + \min_{1 \leqslant i \leqslant n} X_i \right)$ 都是 θ 的无偏估计, 问何者更有效?

46. 设 X_1, X_2, X_3 i.i.d. 服从均匀分布 $U(0, \theta)$, 试证 $\frac{4}{3} \max_{1 \leqslant i \leqslant 3} X_i$ 及 $4 \min_{1 \leqslant i \leqslant 3} X_i$ 都是 θ 的无偏估计量, 哪个更有效?

47. 设 $\hat{\theta}_1$ 及 $\hat{\theta}_2$ 是参数 θ 的两个独立的无偏估计量, 且 $\hat{\theta}_1$ 的方差是 $\hat{\theta}_2$ 的方差的 2 倍, 试确定常数 c_1 及 c_2 使得 $c_1 \hat{\theta}_1 + c_2 \hat{\theta}_2$ 为参数 θ 的无偏估计量并且在所有这样的线性估计中方差最小.

48. 设总体 X 的数学期望为 a , \hat{a}_1 及 \hat{a}_2 分别为 a 的两个无偏估计量, 它们的方差分别为 σ_1^2 和 σ_2^2 , 相关系数为 ρ , 试确定常数 $c_1 > 0$, $c_2 > 0$, $c_1 + c_2 = 1$, 使得 $c_1 \hat{a}_1 + c_2 \hat{a}_2$ 有最小方差.

49. 设 T_1, \dots, T_m 为 $g(\theta)$ 的 m 个独立线性无偏估计, $D(T_i) = \sigma^2 < \infty$, $i = 1, 2, \dots, m$. 则 $\bar{T} = \sum_{i=1}^m T_i/m$ 为 T_1, \dots, T_m 的线性组合类中 $g(\theta)$ 的最小方差无偏估计 (所谓线性无偏估计量, 即 T_i 为 X_1, \dots, X_n 的线性函数且 $E(T_i) = g(\theta)$).

50. 设概率密度函数 $f(x)$ 的核估计 $f_n(x)$ 的定义由式 (3.6.2) 给出, 在定理 3.6.2 的条件下证明 $f_n(x)$ 的均方相合性, 即证明 $\lim_{n \rightarrow \infty} E|f_n(x) - f(x)|^2 = 0$.
51. 设概率密度函数 $f(x)$ 的核估计 $f_n(x)$ 的定义由式 (3.6.2) 给出, 在定理 3.6.2 的条件下证明 $f_n(x)$ 的弱相合性, 即证明对 $\forall \varepsilon > 0$ 有 $\lim_{n \rightarrow \infty} P(|f_n(x) - f(x)| \geq \varepsilon) = 0$.

第4章 区间估计

4.1 区间估计的基本概念

4.1.1 参数的区间估计问题

设有一个参数分布族 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$, 其中 Θ 是参数空间. $\mathbf{X} = (X_1, \dots, X_n)$ 为取自分布族中某总体 $f(x, \theta)$ 的样本, $g(\theta)$ 为定义在 Θ 上的一个已知函数, 要利用样本 \mathbf{X} 对 $g(\theta)$ 的值作出估计, 就是参数估计问题. 参数估计有两类: 点估计和区间估计. 关于点估计的问题已在第3章中讨论过了. 在那里是用样本函数 $\hat{g}(\mathbf{X})$ 去估计 $g(\theta)$ 的, 称为点估计. 这种估计的缺点是: 单从 $\hat{g}(\mathbf{X})$ 所给出的估计值上, 无法看出它的精度有多大. 当然可以定义某种指标, 如估计的均方误差之类去刻画它的精度, 但也还是间接的. 更直接的方法是指出一个误差限 $d(\mathbf{X})$, 把估计写成 $\hat{g}(\mathbf{X}) \pm d(\mathbf{X})$ 的形式. 在应用部门中常见到这种写法. 这实际上就是一种区间估计, 即估计 $g(\theta)$ 的取值在区间 $[\hat{g}(\mathbf{X}) - d(\mathbf{X}), \hat{g}(\mathbf{X}) + d(\mathbf{X})]$ 之内. 将其一般化, 给出区间估计的下列定义.

定义 4.1.1 设有一个参数分布族 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$, $g(\theta)$ 是定义在参数空间 Θ 上的一个已知函数, $\mathbf{X} = (X_1, \dots, X_n)$ 是从分布族某总体 $f(x, \theta)$ 中抽取的样本, 令 $\hat{g}_1(\mathbf{X})$ 和 $\hat{g}_2(\mathbf{X})$ 为定义在样本空间 \mathcal{X} 上, 取值在 Θ 上的两个统计量, 且 $\hat{g}_1(\mathbf{X}) \leq \hat{g}_2(\mathbf{X})$, 则称随机区间 $[\hat{g}_1(\mathbf{X}), \hat{g}_2(\mathbf{X})]$ 为 $g(\theta)$ 的一个区间估计 (interval estimation).

根据这个定义, 从形式上看, 任何一个满足条件 $\hat{g}_1 \leq \hat{g}_2$ 的统计量 \hat{g}_1, \hat{g}_2 都可构成 $g(\theta)$ 的一个区间估计 $[\hat{g}_1, \hat{g}_2]$. 既然一个未知参数的区间估计有很多种, 如何从中挑选一个好的区间估计呢? 这就涉及评价一个区间估计优劣的标准问题. 评价一个区间估计优劣的标准有两个要素: 可靠度与精度 (也称精确度). 可靠度是指待估参数 $g(\theta)$ 被包含在 $[\hat{g}_1, \hat{g}_2]$ 内可能性的大小. 可能性越大, 可靠度越高. 精度可由随机区间的平均长度来度量. 长度越短, 精度越高.

不言而喻, 希望所作的区间估计既有高的可靠度, 又有高的精度. 但这二者往往是彼此矛盾的, 不可能同时都很高. 当样本大小固定时, 若精度提高了, 可靠度就降低; 反之, 若可靠度提高, 则精度就降低. 举例来说明: 有人要求对科大少年班学生入学的年龄作一个区间估计. 甲估计少年班学生入学年龄为 14-16 岁, 精度很高, 但可靠度小了. 因为根据历年记载少年班学生的入学年龄最小的是 12 岁,

在估计区间之外. 乙估计少年班学生的入学年龄为 10-18 岁, 这个可靠度很高, 因为根据历年记载还没有超过这个范围, 更何况少年班入学年龄上限是 15 岁. 但乙估计的精度较差. 可见高可靠度和高精度往往不可兼得.

如何构造可靠度和精度尽可能高的区间估计呢? 通常采用的方法是在保证一定可靠度的前提下选择精度尽可能高的区间估计. 这就是著名统计学家 Neyman 提出的一种妥协方案.

当然, 如果在应用中人们要求可靠度和精度都很高, 则必须加大样本容量, 也就是说要多做一些试验, 才可能实现.

4.1.2 置信区间

为书写简单计, 本节以下假定被估计的 $g(\theta)$ 就是 θ 自身, 这与一般情况没有原则区别.

1. 置信度

设 \mathbf{X} 为样本, $[\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X})]$ 是 θ 的一个区间估计. 由于 θ 是未知的, 且样本是随机的, 不能保证在任何情况下 (即对任何具体的样本值), 区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 必定包含 θ , 而只能以一定的概率保证它. 希望随机区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 包含 θ 的概率 $P_\theta(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2)$ 越大越好. 这个概率就是前面所说的可靠度, 数理统计学上称这个概率为置信度或置信水平.

一般说来, 这个概率与 θ 有关, 假如一个区间估计对某个 $\theta_1 \in \Theta$ 其置信度大, 而对另一个 $\theta_2 \in \Theta$ 其置信度小, 那么这种区间估计的适应性要差一些, 不能认为是一个好的区间估计. 若对参数空间 Θ 中的任一 θ , 其置信度都很大, 则此区间估计就是一个好的区间估计. 因此有如下定义.

定义 4.1.2 设随机区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 为参数 θ 的一个区间估计, 则 $[\hat{\theta}_1, \hat{\theta}_2]$ 包含 θ 的概率 $P_\theta(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2)$ 称为此区间估计的置信水平 (confidence level). 置信水平在参数空间 Θ 上的下确界

$$\inf_{\theta \in \Theta} P_\theta (\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2)$$

称为该区间估计的置信系数 (confidence coefficient).

显然, 一个区间估计的置信水平越大越好. 为了计算置信水平或置信系数, 需要利用有关统计量的精确分布或渐近分布. 可见, 抽样分布在评价和构造区间估计中发挥着重要的作用.

2. 精度

精度的概念在前面已说过. 精度的标准不止一个. 这里介绍其中最常见的一个标准, 即随机区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 的平均长度 $E_\theta(\hat{\theta}_2 - \hat{\theta}_1)$. 平均长度越短, 精度越高, 这也

是符合实际的一项要求. 为说明精度和置信度及其关系, 请看下例.

例 4.1.1 设样本 $\mathbf{X} = (X_1, \dots, X_n)$ 来自正态总体 $N(\mu, \sigma^2)$, 其中 $\mu \in R$, $\sigma^2 > 0$. μ 和 σ^2 的估计量分别是样本均值 \bar{X} 和样本方差 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$, 用 $[\bar{X} - kS/\sqrt{n}, \bar{X} + kS/\sqrt{n}]$ 作为总体均值 μ 的区间估计. 考虑其置信度和精度.

解 记 $\theta = (\mu, \sigma^2)$, 上述区间估计的置信度为

$$\begin{aligned} P_\theta \left(\bar{X} - kS/\sqrt{n} \leq \mu \leq \bar{X} + kS/\sqrt{n} \right) \\ = P_\theta \left(\left| \sqrt{n}(\bar{X} - \mu)/S \right| \leq k \right) \\ = P(|T| \leq k), \end{aligned}$$

其中 $T = \sqrt{n}(\bar{X} - \mu)/S \sim t_{n-1}$, 其分布与 θ 无关, 因而区间估计的置信系数为 $P(|T| \leq k)$. 显然 k 越大, 区间的置信系数越大, 区间就越可靠.

由于 $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, 所以区间的平均长度为

$$l_k = \frac{2kE(S)}{\sqrt{n}} = \frac{2\sqrt{2}k\sigma}{\sqrt{n(n-1)} \Gamma((n-1)/2)} \Gamma(n/2).$$

显然, k 越大, 区间也越长, 精确就越差.

由此例可以看到, 在样本容量 n 给定后, 为了提高置信度, 需要增加 k 值, 从而放大了区间, 降低了精度. 反过来, 为了提高精度, 需要减小 k 值, 从而缩短了区间, 降低了置信度. 置信度与精度互相制约着. 如前所述, 面对这一矛盾, 著名统计学家 Neyman 建议采取如下方案: 在保证置信系数达到指定要求的前提下, 尽可能提高精度. 这一建议导致引入如下置信区间的概念, 由于是 Neyman 建议的, 通常也称置信区间为 Neyman 置信区间.

定义 4.1.3 设 $[\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X})]$ 是参数 θ 的一个区间估计, 若对于给定的 $0 < \alpha < 1$, 有

$$P_\theta \left(\hat{\theta}_1(\mathbf{X}) \leq \theta \leq \hat{\theta}_2(\mathbf{X}) \right) \geq 1 - \alpha, \quad \theta \in \Theta,$$

则称 $[\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X})]$ 是 θ 的置信水平 (confidence level) 为 $1 - \alpha$ 的置信区间 (confidence interval). 而 $\inf_{\theta \in \Theta} P_\theta(\hat{\theta}_1(\mathbf{X}) \leq \theta \leq \hat{\theta}_2(\mathbf{X}))$ 称为 $[\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X})]$ 的置信系数.

置信区间有如下频率解释: 设 $\alpha = 0.05$, 则 $1 - \alpha = 0.95$, 若把置信区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 反复使用多次, 如使用 100 次, 平均大约有 95 次随机区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 包含真参数 θ , 平均大约有 5 次随机区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 不包含 θ . 当使用次数充分大时, 频率接近于置信系数.

4.1.3 置信限

在一些实际问题中, 人们感兴趣的有时仅仅是未知参数的置信上限或置信下限. 例如, 一种新材料的强度, 关心它最低不少于多少; 一个工厂的废品率, 关心它最高不超过多少等. 这也是一种区间估计, 称之为置信下限或置信上限, 定义如下.

定义 4.1.4 设 $\hat{\theta}_U(\mathbf{X})$ 和 $\hat{\theta}_L(\mathbf{X})$ 是定义在样本空间 \mathcal{X} 上, 在参数空间 Θ 上取值的两个统计量, 若对给定的 $0 < \alpha < 1$, 有

$$P_\theta \left(\theta \leq \hat{\theta}_U(\mathbf{X}) \right) \geq 1 - \alpha, \quad \text{一切 } \theta \in \Theta,$$

$$P_\theta \left(\hat{\theta}_L(\mathbf{X}) \leq \theta \right) \geq 1 - \alpha, \quad \text{一切 } \theta \in \Theta,$$

则分别称 $\hat{\theta}_U(\mathbf{X})$ 和 $\hat{\theta}_L(\mathbf{X})$ 是 θ 的置信水平为 $1 - \alpha$ 的 (单侧) 置信上限 (upper confidence limit) 和置信下限 (lower confidence limit). 上式左端概率在参数空间 Θ 上的下确界分别称为置信上、下限的置信系数.

显然, 对置信上限 $\hat{\theta}_U$ 而言, 若 $E(\hat{\theta}_U)$ 越小, 其精度越高; 对置信下限 $\hat{\theta}_L$ 而言, 若 $E(\hat{\theta}_L)$ 越大, 则置信下限的精度越高.

容易看出, 单侧置信上、下限都是置信区间的特例. 因此寻求置信区间的方法可以毫不困难地用来求单侧置信上、下限. 单侧置信限与双侧置信限 (置信区间) 之间存在着一个简单的联系, 下述的引理说明, 在有了单侧置信上、下限后, 不难得求得置信区间.

引理 4.1.1 设 $\hat{\theta}_L(\mathbf{X})$ 和 $\hat{\theta}_U(\mathbf{X})$ 分别是参数 θ 的置信水平为 $1 - \alpha_1$ 和 $1 - \alpha_2$ 的单侧置信下限和置信上限, 且对任何样本 \mathbf{X} 都有 $\hat{\theta}_L(\mathbf{X}) \leq \hat{\theta}_U(\mathbf{X})$, 则 $[\hat{\theta}_L(\mathbf{X}), \hat{\theta}_U(\mathbf{X})]$ 是 θ 的置信水平为 $1 - (\alpha_1 + \alpha_2)$ 的双侧置信区间.

证 在引理的假设下, 下列三个事件:

$$\left\{ \hat{\theta}_L(\mathbf{X}) \leq \theta \leq \hat{\theta}_U(\mathbf{X}) \right\}, \quad \left\{ \theta < \hat{\theta}_L(\mathbf{X}) \right\}, \quad \left\{ \theta > \hat{\theta}_U(\mathbf{X}) \right\}$$

是互不相容的, “三个事件之并” 为“必然事件”. 再考虑到

$$P_\theta \left(\theta < \hat{\theta}_L(\mathbf{X}) \right) = 1 - P_\theta \left(\hat{\theta}_L(\mathbf{X}) \leq \theta \right) < \alpha_1,$$

$$P_\theta \left(\theta > \hat{\theta}_U(\mathbf{X}) \right) = 1 - P_\theta \left(\theta \leq \hat{\theta}_U(\mathbf{X}) \right) < \alpha_2,$$

因此有

$$\begin{aligned} P_\theta \left(\hat{\theta}_L(\mathbf{X}) \leq \theta \leq \hat{\theta}_U(\mathbf{X}) \right) &= 1 - P_\theta \left(\theta < \hat{\theta}_L(\mathbf{X}) \right) - P_\theta \left(\theta > \hat{\theta}_U(\mathbf{X}) \right) \\ &\geq 1 - (\alpha_1 + \alpha_2). \end{aligned}$$

引理得证.

4.1.4 置信域

以上讨论的置信区间和置信上、下限都是假定参数 θ 是一维的, 如果将其推广到参数 θ 是 k 维 ($k \geq 2$) 的情形, 就得如下定义的置信域.

定义 4.1.5 设有一个参数分布族 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$, Θ 是参数空间, 其中 $\theta = (\theta_1, \dots, \theta_k) \in \Theta \subset R_k$, $k \geq 2$. $\mathbf{X} = (X_1, \dots, X_n)$ 是来自分布族中某总体 $f(x, \theta)$ 的样本. 若统计量 $S(\mathbf{X})$ 满足

- (1) 对任一样本 \mathbf{X} , $S(\mathbf{X})$ 是 θ 的一个子集;
- (2) 对给定的 $0 < \alpha < 1$, $P_\theta(\theta \in S(\mathbf{X})) \geq 1 - \alpha$, 一切 $\theta \in \Theta$,

则称 $S(\mathbf{X})$ 是 θ 的置信水平为 $1 - \alpha$ 的置信域 (confidence region) 或置信集, 而 $\inf_{\theta \in \Theta} P_\theta(\theta \in S(\mathbf{X}))$ 称为置信系数.

在多维场合, 置信域 $S(\mathbf{X})$ 的形状可以是各种各样的, 但实用上只限于一些规则的几何图形, 如其各面与坐标平面平行的长方体、球、椭球等. 特别当置信集是长方体 (其面与坐标平面平行) 时, 它又称为联合区间估计.

4.1.5 构造区间估计的方法

目前应用最广泛的区间估计的形式是 Neyman 的置信区间. 4.2 节和 4.3 节将介绍这一方法, 这一方法的关键是基于点估计去构造枢轴变量, 因此也称为枢轴变量法. 另外一种构造区间估计的重要方法是利用假设检验来构造置信区间, 它与枢轴变量法同属于一个理论体系, 即 Neyman 的关于置信区间和假设检验的理论. 利用假设检验构造置信区间的方法将在 5.5 节介绍.

本章的最后两节将介绍区间估计的其他两种方法, 即 Fisher 的信仰推断方法及容忍区间和容忍限.

用 Bayes 方法求区间估计的内容将放在本书的第 7 章介绍.

4.2 枢轴变量法——正态总体参数的置信区间

4.2.1 引言

枢轴变量法的基本要点, 就是在参数的点估计基础上, 去构造它的置信区间. 由于点估计是由样本决定的, 最有可能接近真参数 θ 之值. 因此, 围绕点估计值的区间, 包含真参数值的可能性也要大一些. 下面举一个例子, 说明是如何基于点估计去构造置信区间的.

例 4.2.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的简单随机样本, 此处 σ^2 已知, 求 μ 的置信系数为 $1 - \alpha$ 的置信区间.

解 显然, μ 的一个良好的点估计是 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, 其分布为 $\bar{X} \sim N(\mu, \sigma^2/n)$, 将其标准化得

$$U = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1),$$

其分布与 μ 无关. 由正态分布的对称性, 可得

$$P_\mu \left(\left| \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \right| \leq u_{\alpha/2} \right) = 1 - \alpha,$$

此处 $u_{\alpha/2}$ 为标准正态分布的上侧 $\alpha/2$ 分位数. 经不等式等价变形, 可知

$$P_\mu \left(\bar{X} - \frac{\sigma}{\sqrt{n}} u_{\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \right) = 1 - \alpha.$$

因此 $\left[\bar{X} - \frac{\sigma}{\sqrt{n}} u_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \right]$ 为 μ 的置信系数 $1 - \alpha$ 的置信区间.

由本例可知构造置信区间的步骤如下:

- (1) 找待估参数 μ 的一个良好点估计. 此例中这个点估计是 $T(\mathbf{X}) = \bar{X}$.
- (2) 构造一个 $T(\mathbf{X})$ 和 μ 的函数 $\varphi(T, \mu)$, 使其满足
 - (i) 其表达式与待估参数 μ 有关,
 - (ii) 其分布与待估参数 μ 无关,

则称随机变量 $\varphi(T, \mu)$ 为枢轴变量. 本例中这一变量即为 $U = \sqrt{n}(\bar{X} - \mu)/\sigma$, 它的表达式与 μ 有关, 但其分布 $N(0, 1)$ 与 μ 无关. 因此 U 为枢轴变量.

- (3) 对给定的 $0 < \alpha < 1$, 决定两个常数 a 和 b (本例中 a, b 为 $\pm u_{\alpha/2}$), 使得

$$P_\mu (a \leq \varphi(T, \mu) \leq b) = 1 - \alpha.$$

解括号中的不等式得到 $\hat{\mu}_1(\mathbf{X}) \leq \mu \leq \hat{\mu}_2(\mathbf{X})$, 则有

$$P_\mu (\hat{\mu}_1(\mathbf{X}) \leq \mu \leq \hat{\mu}_2(\mathbf{X})) = 1 - \alpha.$$

这表明 $[\hat{\mu}_1(\mathbf{X}), \hat{\mu}_2(\mathbf{X})]$ 是 μ 的置信水平为 $1 - \alpha$ 的置信区间.

类似的步骤可获得 μ 的置信水平为 $1 - \alpha$ 的单侧置信上、下限.

例 4.2.1 中的 μ 的置信区间就是通过上述三个步骤获得的, 其中最关键的步骤是 (2), 即构造枢轴变量 $\varphi(T, \mu)$, 这个变量一定和 μ 的一个良好的点估计有关. 这种构造置信区间的方法称为枢轴变量法.

4.2.2 单个正态总体参数的置信区间

正态分布 $N(\mu, \sigma^2)$ 是常用的分布。寻求它的两个参数 μ 和 σ^2 的置信区间是实际中常遇到的问题，下面将分几种情况分别加以讨论。这里假设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态总体 $N(\mu, \sigma^2)$ 抽取的简单随机样本。记

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

即 \bar{X} 和 S^2 分别为样本均值和样本方差。

1. σ^2 已知，求 μ 的置信区间

这就是例 4.2.1 讨论过的问题。 μ 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} u_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \right], \quad (4.2.1)$$

其单侧置信上、下限分别为 $\bar{X} + \sigma u_\alpha / \sqrt{n}$ 和 $\bar{X} - \sigma u_\alpha / \sqrt{n}$ 。

由式 (4.2.1) 可知置信区间的长度为 $l_n = 2\sigma u_{\alpha/2} / \sqrt{n}$ 。由此可以看出

(1) 样本容量 n 越大，该区间越短，精度就越高。

(2) σ 越大，则 l_n 越大，精度越低。这是因为方差越大，随机影响也就越大，精度就会低下来。

(3) 置信系数 $1 - \alpha$ 越大，则 α 越小，从而 $u_{\alpha/2}$ 就越大， l_n 越长，精度就越低。

由此可见，在 σ 和 α 固定的情形下，要提高精度，只有增加样本容量。例如，置信系数 $1 - \alpha$ 固定，要使上述置信区间的长度 $l_n \leq l_0$ ， l_0 为给定的常数，则 $n \geq \left[(2\sigma u_{\alpha/2} / l_0)^2 \right] + 1$ ，其中 $[x]$ 表示实数 x 的整数部分。

例 4.2.2 设某车间生产零件的长度 $X \sim N(\mu, 0.09)$ ，若得到一组样本观察值为

$$12.6, 13.4, 12.8, 13.2.$$

求零件平均长度 μ 的 95% 的置信区间。

解 由样本观察值算得 $\bar{X} = 13$ ， $n = 4$ ， $\sigma = 0.3$ ，查表求得 $u_{0.025} = 1.96$ ， $\sigma u_{\alpha/2} / \sqrt{n} = 0.3 \times 1.96 / 2 = 0.294$ ，由式 (4.2.1) 可知， μ 的 95% 的置信区间为

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} u_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \right] = [12.71, 13.29].$$

2. σ^2 未知，求 μ 的置信区间

在这种情况下， μ 的良好的点估计仍为 \bar{X} ，由于 σ^2 未知，例 4.2.1 中的随机变量 $U = \sqrt{n}(\bar{X} - \mu) / \sigma$ 在此不能作为枢轴变量，将其中的 σ 用 S (S^2 是样本方差)

代替, 得到

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S}.$$

由推论 2.4.2 可知 $T \sim t_{n-1}$. 可见 T 的表达式与 μ 有关, 而其分布与 μ 无关, 故取 T 为枢轴变量. 由于 t 分布关于原点对称, 令

$$P(|T| \leq c) = P\left(-c \leq \frac{\sqrt{n}(\bar{X} - \mu)}{S} \leq c\right) = 1 - \alpha,$$

则 $c = t_{n-1}(\alpha/2)$. 将括号中的不等式经过等价变形得 μ 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\bar{X} - \frac{S}{\sqrt{n}}t_{n-1}(\alpha/2), \bar{X} + \frac{S}{\sqrt{n}}t_{n-1}(\alpha/2)\right], \quad (4.2.2)$$

其单侧置信上、下限为 $\bar{X} + St_{n-1}(\alpha)/\sqrt{n}$ 和 $\bar{X} - St_{n-1}(\alpha)/\sqrt{n}$.

例 4.2.3 为测得某种溶液中的甲醛浓度, 取样得 4 个独立测定值的平均值 $\bar{X} = 8.34\%$, 样本标准差 $S = 0.03\%$, 并设测量值近似服从正态分布, 求总体均值 μ 的 95% 的置信区间.

解 因为 $1 - \alpha = 0.95$, $n = 4$, 查表得 $t_{n-1}(\alpha/2) = t_3(0.025) = 3.182$, 故有 $St_{n-1}(\alpha/2)/\sqrt{n} = 0.03 \times 3.182/2 = 0.0477$, $\bar{X} = 8.34$, 由式 (4.2.2) 可知, μ 的置信系数为 95% 的置信区间为

$$\left[\bar{X} - \frac{S}{\sqrt{n}}t_{n-1}(\alpha/2), \bar{X} + \frac{S}{\sqrt{n}}t_{n-1}(\alpha/2)\right] = [8.292\%, 8.388\%].$$

3. μ 已知, 求 σ^2 的置信区间

当 μ 已知时, σ^2 的一个良好的无偏估计为 $S_\mu^2 = \sum_{i=1}^n (X_i - \mu)^2/n$, 且 $nS_\mu^2/\sigma^2 \sim \chi_n^2$, 则取 $T = nS_\mu^2/\sigma^2$ 为枢轴变量, 其表达式与 σ^2 有关, 但其分布与 σ^2 无关, 找 c_1 和 c_2 使得

$$P_{\sigma^2}\left(c_1 \leq \frac{nS_\mu^2}{\sigma^2} \leq c_2\right) = 1 - \alpha.$$

满足上式要求的 c_1 和 c_2 有无穷多对, 其中有一对 c_1 和 c_2 , 使区间的长度最短, 但这样一对 c_1 和 c_2 不易求得且表达式复杂, 应用不方便. 一般令 c_1 和 c_2 满足下列要求:

$$P_{\sigma^2}\left(\frac{nS_\mu^2}{\sigma^2} < c_1\right) = \frac{\alpha}{2}, \quad P_{\sigma^2}\left(\frac{nS_\mu^2}{\sigma^2} > c_2\right) = \frac{\alpha}{2}.$$

由 χ^2 分布的上侧分位数表可知 $c_1 = \chi_n^2(1 - \alpha/2)$, $c_2 = \chi_n^2(\alpha/2)$, 即有

$$P_{\sigma^2}\left(\chi_n^2\left(1 - \frac{\alpha}{2}\right) \leq \frac{nS_\mu^2}{\sigma^2} \leq \chi_n^2\left(\frac{\alpha}{2}\right)\right) = 1 - \alpha.$$

最后利用不等式的等价变形, 得到 σ^2 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\frac{nS_\mu^2}{\chi_n^2(\alpha/2)}, \frac{nS_\mu^2}{\chi_n^2(1-\alpha/2)} \right], \quad (4.2.3)$$

其单侧置信下、上限为 $nS_\mu^2/\chi_n^2(\alpha)$ 和 $nS_\mu^2/\chi_n^2(1-\alpha)$.

例 4.2.4 为了解一台测量长度的仪器的精度, 对一根长 30mm 的标准金属棒进行了 6 次测量, 结果(单位: mm)是

$$30.1, 29.9, 29.8, 30.3, 30.2, 29.6.$$

假如测量值服从正态分布 $N(30, \sigma^2)$, 要求 σ^2 的置信水平为 0.95 的置信区间.

解 此处 $n = 6$, $\mu = 30$, 易得出 $\sum_{i=1}^6 (X_i - \mu)^2 = 0.35$, $\alpha = 0.05$, 查表得 $\chi_6^2(0.025) = 14.4494$, $\chi_6^2(0.975) = 1.2375$, 由式 (4.2.3) 可算得

$$\begin{aligned}\hat{\sigma}_1^2 &= \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2(\alpha/2)} = \frac{0.35}{14.4494} = 0.0242, \\ \hat{\sigma}_2^2 &= \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2(1-\alpha/2)} = \frac{0.35}{1.2375} = 0.2828.\end{aligned}$$

因此 σ^2 的置信水平为 95% 的置信区间为 $[\hat{\sigma}_1^2, \hat{\sigma}_2^2] = [0.0242, 0.2828]$.

4. μ 未知, 求 σ^2 的置信区间

记 $\theta = (\mu, \sigma^2)$. 此时 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$ 是 σ^2 的良好估计, 它是无偏的, 且由定理 2.2.3 可知 $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$. 取 $T = (n-1)S^2/\sigma^2$ 为枢轴变量, 其表达式与 σ^2 有关, 而其分布与 σ^2 无关. 找 d_1 和 d_2 , 使得

$$P_\theta \left(d_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq d_2 \right) = 1 - \alpha.$$

类似于 3 中确定 c_1 和 c_2 的理由和方法, 取 $d_1 = \chi_{n-1}^2(1-\alpha/2)$, $d_2 = \chi_{n-1}^2(\alpha/2)$, 故有

$$P_\theta \left(\chi_{n-1}^2(1-\alpha/2) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{n-1}^2(\alpha/2) \right) = 1 - \alpha.$$

最后再利用不等式的等价变形, 得出 σ^2 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\frac{(n-1)S^2}{\chi_{n-1}^2(\alpha/2)}, \frac{(n-1)S^2}{\chi_{n-1}^2(1-\alpha/2)} \right]. \quad (4.2.4)$$

若把这个随机区间的两个端点开平方, 得到 σ 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\left(\frac{(n-1)S^2}{\chi_{n-1}^2(\alpha/2)} \right)^{1/2}, \left(\frac{(n-1)S^2}{\chi_{n-1}^2(1-\alpha/2)} \right)^{1/2} \right].$$

类似可得 σ^2 的置信系数为 $1 - \alpha$ 的单侧置信下、上限为 $(n-1)S^2/\chi_{n-1}^2(\alpha)$ 和 $(n-1)S^2/\chi_{n-1}^2(1-\alpha)$.

例 4.2.5 求例 4.2.3 中总体方差 σ^2 及 σ 的置信系数为 95% 的置信区间.

解 如同例 4.2.3, $n-1=3$, $\alpha/2=0.025$, $1-\alpha/2=0.975$. 查表求得 $\chi_3^2(0.025)=9.348$, $\chi_3^2(0.975)=0.216$, $S^2=0.0009$, 由式 (4.2.4) 可知, σ^2 的置信系数为 95% 的置信区间为

$$\left[\frac{(n-1)S^2}{\chi_{n-1}^2(\alpha/2)}, \frac{(n-1)S^2}{\chi_{n-1}^2(1-\alpha/2)} \right] = [0.00029, 0.0125],$$

σ 的置信系数为 95% 的置信区间为

$$\left[\left(\frac{(n-1)S^2}{\chi_{n-1}^2(\alpha/2)} \right)^{1/2}, \left(\frac{(n-1)S^2}{\chi_{n-1}^2(1-\alpha/2)} \right)^{1/2} \right] = [0.017, 0.112].$$

4.2.3 两个正态总体参数的置信区间

设 X_1, \dots, X_m 是自正态总体 $N(a, \sigma_1^2)$ 抽取的简单随机样本, Y_1, \dots, Y_n 是自正态总体 $N(b, \sigma_2^2)$ 抽取的简单随机样本, 且 X_1, \dots, X_m 和 Y_1, \dots, Y_n 独立. 设 \bar{X}, \bar{Y} 和 S_1^2, S_2^2 分别为这两组样本的样本均值和样本方差, 其中 $S_1^2 = \sum_{i=1}^m (X_i - \bar{X})^2 / (m-1)$, $S_2^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 / (n-1)$. 下面分两种情况讨论两个正态总体均值差和方差比的置信区间问题.

1. 均值差 $b-a$ 的置信区间

分下列四种情况.

(1) 当 $m=n$ 时, 令 $Z_i = Y_i - X_i$, $i=1, 2, \dots, n$, 且记 $\tilde{\mu} = b-a$, $\tilde{\sigma}^2 = \sigma_1^2 + \sigma_2^2$, 则有

$$Z_i \sim N(\tilde{\mu}, \tilde{\sigma}^2), \quad i=1, 2, \dots, n.$$

这就转化为单个正态总体当 σ^2 未知, 求其均值 $\tilde{\mu}$ 的置信区间问题. 显见 $\bar{Z} = \bar{Y} - \bar{X}$ 是 $\tilde{\mu}$ 的一个良好的无偏估计, 枢轴变量

$$T_Z = \frac{\sqrt{n}(\bar{Z} - \tilde{\mu})}{S_Z} \sim t_{n-1},$$

此处 $S_Z^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2 / (n-1)$, T_Z 的表达式与 $\tilde{\mu} = b - a$ 有关, 但其分布与 $\tilde{\mu}$ 无关, 因此取 T_Z 为枢轴变量. 由前面已讨论过的情形的结果, 可知 $\tilde{\mu} = b - a$ 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\bar{Z} - \frac{S_Z}{\sqrt{n}} t_{n-1}(\alpha/2), \bar{Z} + \frac{S_Z}{\sqrt{n}} t_{n-1}(\alpha/2) \right], \quad (4.2.5)$$

其单侧置信下、上限分别为 $\bar{Z} - S_Z t_{n-1}(\alpha)/\sqrt{n}$ 和 $\bar{Z} + S_Z t_{n-1}(\alpha)/\sqrt{n}$.

(2) 当 σ_1^2 和 σ_2^2 已知时, 易知 $\bar{Y} - \bar{X}$ 为 $b - a$ 的一个良好的无偏估计, 且 $\bar{Y} - \bar{X} \sim N(b - a, \sigma_1^2/m + \sigma_2^2/n)$, 将 r.v. $\bar{Y} - \bar{X}$ 标准化, 可知

$$U = \frac{\bar{Y} - \bar{X} - (b - a)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0, 1).$$

U 的表达式与 $b - a$ 有关, 但其分布与 $b - a$ 无关, 取 U 为枢轴变量, 故有

$$P_{a,b} \left(\left| \frac{\bar{Y} - \bar{X} - (b - a)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \right| \leq u_{\alpha/2} \right) = 1 - \alpha.$$

再由括号中不等式的等价变形得到 $b - a$ 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\bar{Y} - \bar{X} - u_{\alpha/2} \sqrt{\sigma_1^2/m + \sigma_2^2/n}, \bar{Y} - \bar{X} + u_{\alpha/2} \sqrt{\sigma_1^2/m + \sigma_2^2/n} \right]. \quad (4.2.6)$$

其单侧置信上、下限为 $(\bar{Y} - \bar{X}) \pm u_\alpha \sqrt{\sigma_1^2/m + \sigma_2^2/n}$, 其中 “+” 号是置信上限.

(3) 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知时, 令

$$\begin{aligned} S_\omega^2 &= \frac{1}{m+n-2} [(m-1)S_1^2 + (n-1)S_2^2] \\ &= \frac{1}{m+n-2} \left[\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2 \right], \end{aligned}$$

显然 $\bar{Y} - \bar{X}$ 是 $b - a$ 的无偏估计, 由推论 2.4.3 可知

$$T_\omega = \frac{\bar{Y} - \bar{X} - (b - a)}{S_\omega} \sqrt{\frac{mn}{m+n}} \sim t_{n+m-2},$$

T_ω 的表达式与 $b - a$ 有关, 但其分布与 $b - a$ 无关, 取 T_ω 为枢轴变量, 故有

$$P \left(\left| \frac{\bar{Y} - \bar{X} - (b - a)}{S_\omega} \right| \sqrt{\frac{mn}{m+n}} \leq t_{n+m-2}(\alpha/2) \right) = 1 - \alpha,$$

由括号中不等式的等价变形, 可得 $b - a$ 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\bar{Y} - \bar{X} - S_\omega t_{n+m-2} \left(\frac{\alpha}{2} \right) \sqrt{\frac{1}{m} + \frac{1}{n}}, \bar{Y} - \bar{X} + S_\omega t_{n+m-2} \left(\frac{\alpha}{2} \right) \sqrt{\frac{1}{m} + \frac{1}{n}} \right], \quad (4.2.7)$$

其单侧置信上、下限为 $(\bar{Y} - \bar{X}) \pm S_{\omega} t_{m+n-2}(\alpha) \sqrt{\frac{1}{m} + \frac{1}{n}}$, 其中 “+” 号是置信上限.

(4) 当 $\sigma_1^2 \neq \sigma_2^2$ 皆未知时, 求 $b-a$ 的置信区间问题. 这是著名的 Behrens–Fisher 问题. 它是 Behrens 在 1929 年从实际应用提出的问题, 它的几种特殊情况如上所述, 已获圆满解决, 但一般情况至今仍有文献在讨论. Fisher 首先研究了这个问题, 并对一般情况给出近似解法. 随后许多著名统计学家, 如 Scheffe 和 Welch 等也研究过这个问题. 但至今还得不出简单、精确的解法, 只提出一些近似的解法. 下面给出两种近似结果.

(i) 当 m 与 n 都充分大时可用大样本方法, 由于

$$\frac{\bar{Y} - \bar{X} - (b-a)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0, 1), \quad (4.2.8)$$

且当 $m \rightarrow \infty$ 时 $S_1^2 \xrightarrow{P} \sigma_1^2$, 当 $n \rightarrow \infty$ 时 $S_2^2 \xrightarrow{P} \sigma_2^2$, 将式 (4.2.8) 中的 σ_1^2 和 σ_2^2 分别用 S_1^2 和 S_2^2 代入, 利用引理 2.5.1 可知, 当 $m, n \rightarrow \infty$ 时, 有

$$\tilde{U} = \frac{\bar{Y} - \bar{X} - (b-a)}{\sqrt{S_1^2/m + S_2^2/n}} = \frac{\bar{Y} - \bar{X} - (b-a)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \cdot \frac{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}{\sqrt{S_1^2/m + S_2^2/n}} \xrightarrow{D} N(0, 1).$$

因此, 取 \tilde{U} 为枢轴变量. 当 m, n 充分大时, $b-a$ 的置信系数近似为 $1-\alpha$ 的置信区间是

$$\left[\bar{Y} - \bar{X} - u_{\alpha/2} \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}, \bar{Y} - \bar{X} + u_{\alpha/2} \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}} \right]. \quad (4.2.9)$$

其单侧置信上、下限为 $(\bar{Y} - \bar{X}) \pm u_{\alpha} \sqrt{S_1^2/m + S_2^2/n}$, 其中 “+” 号是置信上限.

*(ii) 一般情形, 即 m 和 n 都不是充分大的情形. 令

$$S_*^2 = \frac{S_1^2}{m} + \frac{S_2^2}{n}.$$

取枢轴变量为

$$T_* = \frac{\bar{Y} - \bar{X} - (b-a)}{S_*}.$$

在一般情况下, T_* 已不服从 t 分布. 但与具有适当自由度 r 的 t 分布很接近, 其中 r 由下列公式确定:

$$r = S_*^4 / \left[\frac{S_1^4}{m^2(m-1)} + \frac{S_2^4}{n^2(n-1)} \right].$$

r 一般不为整数, 可取与其最接近的整数代替之. 于是, 近似地有

$$T_* = \frac{\bar{Y} - \bar{X} - (b-a)}{S_*} \sim t_r.$$

运用与前面类似步骤可得 $b - a$ 的置信系数近似为 $1 - \alpha$ 的置信区间为

$$\left[\bar{Y} - \bar{X} - S_* t_r(\alpha/2), \bar{Y} - \bar{X} + S_* t_r(\alpha/2) \right]. \quad (4.2.10)$$

其单侧置信上、下限为 $(\bar{Y} - \bar{X}) \pm S_* t_r(\alpha)$, 其中“+”号是置信上限.

这是由 Welch 在 1938 年给出的 Behrens-Fisher 问题的一个近似解法.

例 4.2.6 某公司利用两条自动化流水线灌装矿泉水. 现从生产线上抽取样本 X_1, \dots, X_{12} 和 Y_1, \dots, Y_{17} , 它们是每瓶矿泉水的体积 (单位: ml). 算得样本均值 $\bar{X} = 501.1$ 和 $\bar{Y} = 499.7$; 样本方差 $S_1^2 = 2.4$, $S_2^2 = 4.7$. 假设这两条流水线所装的矿泉水的体积分别服从正态分布 $N(a, \sigma^2)$ 和 $N(b, \sigma^2)$, 试求 $b - a$ 置信系数为 0.95 的置信区间.

解 $\bar{Y} - \bar{X} = -1.4$, $S_\omega^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{n+m-2} = \frac{11 \times 2.4 + 16 \times 4.7}{12+17-2} = 3.763$, 于是 $S_\omega = 1.94$, 查表求得 $t_{m+n-2}(\alpha/2) = t_{27}(0.025) = 2.05$, 算得

$$S_\omega t_{m+n-2}(\alpha/2) \sqrt{\frac{1}{n} + \frac{1}{m}} = 1.94 \times 2.05 \sqrt{\frac{1}{12} + \frac{1}{17}} = 1.50.$$

因此 $b - a$ 的 95% 的置信区间按式 (4.2.7) 算得 $[-2.9, 0.1]$.

例 4.2.7 欲比较甲、乙两种棉花品种的优劣. 现假设用它们纺出的棉纱强度服从 $N(a, \sigma_1^2)$ 和 $N(b, \sigma_2^2)$, $\sigma_1^2 \neq \sigma_2^2$. 试验者从这两批棉纱中分别抽取样本 X_1, \dots, X_{120} 和 Y_1, \dots, Y_{60} , 其均值分别为 $\bar{X} = 3.32$, $\bar{Y} = 3.76$, $S_1^2 = 2.18$, $S_2^2 = 5.76$. 试给出 $a - b$ 的 95% 的置信区间.

解 由于 $\sigma_1^2 \neq \sigma_2^2$, 故这是 Behrens-Fisher 问题, 用大样本的近似方法求置信区间. 由数据算得 $\bar{X} - \bar{Y} = -0.44$, $\sqrt{S_1^2/m + S_2^2/n} = \sqrt{2.18/120 + 5.76/60} = 0.338$, 查表求得 $u_{0.025} = 1.96$, 算得

$$u_{\alpha/2} \sqrt{S_1^2/m + S_2^2/n} = 1.96 \times 0.338 = 0.662,$$

按式 (4.2.9) 可得 $a - b$ 的置信系数近似为 0.95 的置信区间为 $[-1.102, 0.222]$.

2. 方差比 σ_1^2/σ_2^2 的置信区间

(1) 若 a 和 b 已知, 记 $S_a^2 = \sum_{i=1}^m (X_i - a)^2/m$, $S_b^2 = \sum_{i=1}^n (Y_i - b)^2/n$, 显见 $mS_a^2/\sigma_1^2 \sim \chi_m^2$, $nS_b^2/\sigma_2^2 \sim \chi_n^2$, 且 S_a^2 为 σ_1^2 的无偏估计, S_b^2 为 σ_2^2 的无偏估计, 且二者独立, 故由推论 2.4.4 可知

$$F = \frac{S_a^2/\sigma_1^2}{S_b^2/\sigma_2^2} = \frac{S_a^2}{S_b^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{m,n}.$$

F 的表达式与 σ_1^2/σ_2^2 有关, 但其分布与 σ_1^2/σ_2^2 无关, 故取 F 为枢轴变量. 找 c_1 和 c_2 , 使得

$$P\left(c_1 \leq \frac{S_a^2}{S_b^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leq c_2\right) = 1 - \alpha.$$

满足上述要求的 c_1 和 c_2 有无穷多对, 其中存在一对使区间估计精度最高, 但这样一对 c_1 和 c_2 不但不易求得而且表达式复杂, 应用起来不方便. 下面方法确定的 c_1 和 c_2 虽然不能使置信区间的精度最高, 但 c_1 和 c_2 容易求得且表达式简单, 使用方便. 令

$$P\left(\frac{S_a^2}{S_b^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} < c_1\right) = \frac{\alpha}{2}, \quad P\left(\frac{S_a^2}{S_b^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} > c_2\right) = \frac{\alpha}{2},$$

查自由度为 m, n 的 F 分布上侧分位数表易得 $c_2 = F_{m,n}(\alpha/2)$, $c_1 = F_{m,n}(1-\alpha/2)$, 因此有

$$P\left(F_{m,n}(1-\alpha/2) \leq \frac{S_a^2}{S_b^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leq F_{m,n}(\alpha/2)\right) = 1 - \alpha.$$

再利用不等式的等价变形, 得到 σ_1^2/σ_2^2 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\frac{S_a^2}{S_b^2} \cdot \frac{1}{F_{m,n}(\alpha/2)}, \frac{S_a^2}{S_b^2} \cdot \frac{1}{F_{m,n}(1-\alpha/2)}\right]. \quad (4.2.11)$$

注意到由于 α 较小, $F_{m,n}(1-\alpha/2)$ 在 F 分布表上查不到, 利用 F 分布的如下性质 (第2章习题32):

$$F_{m,n}(1-\alpha/2) = \frac{1}{F_{n,m}(\alpha/2)}, \quad (4.2.12)$$

可以将其通过查 $F_{n,m}(\alpha/2)$ 算得.

由引理4.1.1可知, σ_1^2/σ_2^2 的置信系数为 $1 - \alpha$ 的单侧置信下、上限为

$$\frac{S_a^2}{S_b^2} \cdot \frac{1}{F_{m,n}(\alpha)} \quad \text{和} \quad \frac{S_a^2}{S_b^2} \cdot \frac{1}{F_{m,n}(1-\alpha)} = \frac{S_a^2}{S_b^2} \cdot F_{n,m}(\alpha).$$

(2) 若 a 和 b 未知, 令 S_1^2 和 S_2^2 由本小节开头的公式给出, 显然 $(m-1)S_1^2/\sigma_1^2 \sim \chi_{m-1}^2$, $(n-1)S_2^2/\sigma_2^2 \sim \chi_{n-1}^2$, 且 S_1^2 和 S_2^2 分别为 σ_1^2 和 σ_2^2 的无偏估计, 故由推论2.4.4可知

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{m-1, n-1}.$$

F 的表达式与 σ_1^2/σ_2^2 有关, 但其分布与 σ_1^2/σ_2^2 无关, 故取 F 为枢轴变量. 找 d_1 和 d_2 , 使得

$$P\left(d_1 \leq \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leq d_2\right) = 1 - \alpha.$$

类似于(1)中确定 c_1 和 c_2 的理由和方法, 取 $d_1 = F_{m-1,n-1}(1 - \alpha/2)$, $d_2 = F_{m-1,n-1}(\alpha/2)$, 故有

$$P\left(F_{m-1,n-1}(1 - \alpha/2) \leq \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leq F_{m-1,n-1}(\alpha/2)\right) = 1 - \alpha.$$

最后利用不等式的等价变形, 得到 σ_1^2/σ_2^2 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{m-1,n-1}(\alpha/2)}, \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{m-1,n-1}(1 - \alpha/2)}\right]. \quad (4.2.13)$$

由于当 α 较小时 (如 $\alpha = 0.01$, $\alpha = 0.05$), $F_{m-1,n-1}(1 - \alpha/2)$ 在 F 分布表中无法查到, 可用式(4.2.12)所述的方法解决.

由引理 4.1.1 可知, σ_1^2/σ_2^2 的置信系数为 $1 - \alpha$ 的单侧置信下、上限为

$$\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{m-1,n-1}(\alpha)} \quad \text{和} \quad \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{m-1,n-1}(1 - \alpha)} = \frac{S_1^2}{S_2^2} \cdot F_{n-1,m-1}(\alpha).$$

例 4.2.8 求例 4.2.7 中方差比 σ_1^2/σ_2^2 的置信系数为 90% 的置信区间.

解 $S_1^2/S_2^2 = 2.18/5.76 = 0.378$, 查表求得 $F_{m-1,n-1}(\alpha/2) = F_{119,59}(0.05) = 1.47$, $F_{m-1,n-1}(1 - \alpha/2) = 1/F_{n-1,m-1}(\alpha/2) = 1/F_{59,119}(0.05) = 1/1.43$, 由式(4.2.13)得 σ_1^2/σ_2^2 的置信系数为 90% 的置信区间为

$$\left[\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{m-1,n-1}(\alpha/2)}, \frac{S_1^2}{S_2^2} \cdot F_{n-1,m-1}(\alpha/2)\right] = [0.257, 0.541].$$

4.3 枢轴变量法——非正态总体参数的置信区间

利用枢轴变量法构造置信区间的方法在 4.2 节已介绍, 本节将用这一方法讨论几个非正态总体参数的置信区间问题. 若枢轴变量的精确分布容易求得, 可用小样本方法获得精确的置信区间; 若枢轴变量的精确分布不易求得, 或若其精确分布虽可以求得, 但表达式很复杂, 使用不方便, 则可用枢轴变量的极限分布来构造有关参数近似的置信区间.

4.3.1 小样本方法

1. 指数分布参数的置信区间

设 X_1, \dots, X_n 为从指数分布 $\text{Exp}(\lambda)$ 中抽取的简单随机样本, 其密度函数为 $f(x, \lambda) = \lambda e^{-\lambda x} I_{(0, \infty)}(x)$, $\lambda > 0$, 要求 λ 的置信系数为 $1 - \alpha$ 的置信区间.

因为 $\bar{X} = \sum_{i=1}^n X_i/n$ 是 $1/\lambda$ 的一个无偏估计 (且是 UMVUE), 设想 λ 的置信区间可通过 \bar{X} 表示. 枢轴变量可取为 $T = 2\lambda n \bar{X}$, 由推论 2.4.5 可知

$$2\lambda n \bar{X} = 2\lambda \sum_{i=1}^n X_i \sim \chi_{2n}^2. \quad (4.3.1)$$

确定 a, b 使得

$$P(a \leq 2\lambda n \bar{X} \leq b) = 1 - \alpha.$$

使上式成立的 a, b 有很多对, 其中存在一对, 使置信区间的长度最短, 但那对 a, b 表达式复杂, 不易求得, 应用上也不方便. 通常采用下列方法, 令

$$P(2\lambda n \bar{X} < a) = \frac{\alpha}{2}, \quad P(2\lambda n \bar{X} > b) = \frac{\alpha}{2}.$$

由式 (4.3.1) 可知, $a = \chi_{2n}^2(1 - \alpha/2)$, $b = \chi_{2n}^2(\alpha/2)$, 这样找到的 a, b 虽不能使置信区间的精度最高, 但表达式简单, 可通过查 χ^2 分布的上侧 α 分位数表求得, 应用上很方便. 因此有

$$P\left(\chi_{2n}^2(1 - \alpha/2) \leq 2\lambda n \bar{X} \leq \chi_{2n}^2(\alpha/2)\right) = 1 - \alpha.$$

利用不等式的等价变形, 可得 λ 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\frac{\chi_{2n}^2(1 - \alpha/2)}{2n \bar{X}}, \frac{\chi_{2n}^2(\alpha/2)}{2n \bar{X}} \right], \quad (4.3.2)$$

其单侧置信下、上限为 $\chi_{2n}^2(1 - \alpha)/(2n \bar{X})$ 和 $\chi_{2n}^2(\alpha)/(2n \bar{X})$.

例 4.3.1 设某电子产品的寿命服从指数分布 $\text{Exp}(\lambda)$, 现从此分布的一批产品中抽取容量为 9 的样本, 测得寿命为 (单位: 千小时)

$$15, 45, 50, 53, 60, 65, 70, 83, 90.$$

求平均寿命 $1/\lambda$ 的置信系数为 90% 的置信区间和置信上、下限.

解 $n = 9$, 由样本算得 $\bar{X} = 59$, $2n \bar{X} = 1062$. 查表求得

$$\chi_{18}^2(0.05) = 28.869, \quad \chi_{18}^2(0.95) = 9.390,$$

$$\chi_{18}^2(0.10) = 25.989, \quad \chi_{18}^2(0.90) = 10.865,$$

则 $g(\lambda) = 1/\lambda$ 的置信系数为 90% 的置信区间由下式确定:

$$P_\lambda\left(\chi_{18}^2(0.95) \leq 2n \lambda \bar{X} \leq \chi_{18}^2(0.05)\right) = 0.90.$$

解括号内不等式得 $g(\lambda) = 1/\lambda$ 的 90% 的置信区间为

$$\left[\frac{2n\bar{X}}{\chi^2_{18}(0.05)}, \frac{2n\bar{X}}{\chi^2_{18}(0.95)} \right] = [36.787, 113.099].$$

类似方法求得 $g(\lambda) = 1/\lambda$ 的置信系数为 90% 的置信上、下限 \hat{g}_U 和 \hat{g}_L 分别为

$$\hat{g}_U = \frac{2n\bar{X}}{\chi^2_{18}(0.90)} = 97.745 \text{ 千小时},$$

$$\hat{g}_L = \frac{2n\bar{X}}{\chi^2_{18}(0.10)} = 40.863 \text{ 千小时}.$$

2. 均匀分布参数的置信区间

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为自均匀分布总体 $U(0, \theta)$ 中抽取的简单随机样本, 求 θ 的置信系数为 $1 - \alpha$ 的置信区间.

设 $T(\mathbf{X}) = X_{(n)} = \max\{X_1, \dots, X_n\}$, $T(\mathbf{X})$ 为充分统计量, 由例 3.3.6 可知 $(n+1)T/n$ 是 θ 的无偏估计 (也是 UMVUE 的). 设想枢轴变量一定与 T 有关. 由于 $Y_i = X_i/\theta \sim U(0, 1)$, $i = 1, 2, \dots, n$, 故 $Y_{(n)} = T/\theta$ 的密度函数为 $f(y) = ny^{n-1}I_{(0,1)}(y)$. 取 $Z = 1/Y_{(n)} = \theta/T$ 为枢轴变量, 其表达式与 θ 有关, 其密度函数为

$$\frac{\theta}{T} \sim g(z) = nz^{-(n+1)}I_{(1,\infty)}(z). \quad (4.3.3)$$

$g(z)$ 与 θ 无关. 确定 d_1, d_2 , $1 \leq d_1 < d_2 \leq \infty$ 使得

$$\begin{aligned} P_\theta \left(d_1 \leq \frac{\theta}{T} \leq d_2 \right) &= P(d_1 T \leq \theta \leq d_2 T) \\ &= \int_{d_1}^{d_2} nz^{-n-1} dz = \frac{1}{d_1^n} - \frac{1}{d_2^n} = 1 - \alpha. \end{aligned} \quad (4.3.4)$$

于是 $[d_1 T, d_2 T]$ 为 θ 的置信系数为 $1 - \alpha$ 的置信区间. 由式 (4.3.4) 中被积函数 $g(z)$ 单调降的性质, 要使区间估计的精度最高, 宜取 $d_1 = 1$, 再解方程 $1 - 1/d_2^n = 1 - \alpha$ 得到 $d_2 = 1/\sqrt[n]{\alpha}$.

因此 θ 的置信系数为 $1 - \alpha$ 的置信区间为

$$[d_1 T, d_2 T] = \left[T, \frac{T}{\sqrt[n]{\alpha}} \right]. \quad (4.3.5)$$

4.3.2 大样本方法

1. Cauchy 分布位置参数的置信区间

设 X_1, \dots, X_n 为从 Cauchy 分布 $C(\theta, 1)$ 中抽取的简单随机样本, 此分布的密

度函数为

$$f(x, \theta) = \frac{1}{\pi[1 + (x - \theta)^2]}, \quad -\infty < x < +\infty, -\infty < \theta < +\infty,$$

要求位置参数 θ 的置信区间.

以 m_n 记 X_1, \dots, X_n 的样本中位数, θ 是总体的中位数, 因此 m_n 作为 θ 的点估计是合适的. 由于 Cauchy 分布关于 θ 对称, 故 $m_n - \theta$ 的分布与从 $\theta = 0$ 的 Cauchy 分布中抽取的大小为 n 的样本中位数的分布相同. 因此 $m_n - \theta$ 的分布与 θ 无关, 可作为枢轴变量, 其分布密度记为 $f_n(x)$. 当 n 为奇数时 $f_n(x)$ 的表达式利用式 (2.3.1) 可直接写出来. 当 n 为偶数时要复杂些, 但原则上求其表达式并无困难. 找到 $f_n(x)$ 后, 要确定 c , 使得

$$P_\theta(|m_n - \theta| \leq c) = \int_{-c}^c f_n(x) dx = 1 - \alpha, \quad (4.3.6)$$

由此得出 θ 的置信系数为 $1 - \alpha$ 的置信区间为 $[m_n - c, m_n + c]$.

但是, 因为 $f_n(x)$ 的表达式很复杂, 要由式 (4.3.6) 决定 c 不容易. 但根据式 (2.5.3), 并注意到 $f(\xi_{1/2}) = 1/\pi$, 可知当 $n \rightarrow \infty$ 时有

$$\frac{2\sqrt{n}(m_n - \theta)}{\pi} \xrightarrow{\mathcal{L}} N(0, 1), \quad (4.3.7)$$

因此有

$$P\left(\frac{2\sqrt{n}|m_n - \theta|}{\pi} \leq u_{\alpha/2}\right) \approx 1 - \alpha,$$

此处 $u_{\alpha/2}$ 是标准正态分布的上侧 $\alpha/2$ 分位数. 利用不等式的等价变形可知, 当 n 充分大时, θ 的置信系数近似为 $1 - \alpha$ 的置信区间为

$$\left[m_n - \frac{\pi u_{\alpha/2}}{2\sqrt{n}}, m_n + \frac{\pi u_{\alpha/2}}{2\sqrt{n}}\right], \quad (4.3.8)$$

其单侧置信上、下限为 $m_n \pm \pi u_{\alpha}/(2\sqrt{n})$, 其中 “+” 号是置信上限.

2. 二项分布总体参数的置信区间

设 $\boldsymbol{X} = (X_1, \dots, X_n)$ 为自两点分布 $b(1, p)$ 中抽取的简单随机样本, 要求 p 的置信区间.

令 $S_n = \sum_{i=1}^n X_i$, 可知 $S_n \sim b(n, p)$, 即二项分布. 当 n 充分大时, 利用中心极限定理, 有

$$\frac{S_n - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} N(0, 1), \quad n \rightarrow \infty. \quad (4.3.9)$$

这表明当 n 充分大时, 随机变量 $T = \sqrt{n}(\bar{X} - p) / \sqrt{p(1-p)}$ 的极限分布是 $N(0, 1)$, 与未知参数 p 无关. 取这样的 T 为枢轴变量. 因此 n 充分大时有

$$P(|T| \leq u_{\alpha/2}) = P\left(\left|\frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}}\right| \leq u_{\alpha/2}\right) \approx 1 - \alpha. \quad (4.3.10)$$

解上式括号中的不等式 $|T| \leq u_{\alpha/2}$, 则由式 (4.3.10) 得到

$$P(|T| \leq u_{\alpha/2}) = P(c_1 \leq p \leq c_2) \approx 1 - \alpha,$$

其中 $c_1 = c_1(\mathbf{X})$, $c_2 = c_2(\mathbf{X})$ 与 p 无关. 故 $[c_1(\mathbf{X}), c_2(\mathbf{X})]$ 可作为 p 的置信系数近似为 $1 - \alpha$ 的置信区间. 记 $\hat{p} = \bar{X}$, $\gamma = u_{\alpha/2}$, 则式 (4.3.10) 括号中不等式等价于

$$(\hat{p} - p)^2 \leq \frac{\gamma^2 p(1-p)}{n},$$

将不等式两边展开, 整理合并同类项, 得到 p 的二次三项式

$$p^2(n + \gamma^2) - p(2n\hat{p} + \gamma^2) + n\hat{p}^2 \leq 0.$$

上式左端 p 的二次三项式的辨别式大于 0, 且平方项系数为正, 故满足上述不等式的 p 介于其两个不相等的正根之间, 记这两个根为 c_1 , c_2 , $c_1 < c_2$, 它们是

$$c_1, c_2 = \frac{n}{n + \gamma^2} \left[\hat{p} + \frac{\gamma^2}{2n} \pm \gamma \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{\gamma^2}{4n^2}} \right], \quad (4.3.11)$$

其中 c_1 相应于 “-” 号. 因此 $[c_1(\mathbf{X}), c_2(\mathbf{X})]$ 就是 p 的置信系数近似为 $1 - \alpha$ 的置信区间.

在实用中, 可采用下列更简单的方法: 由 $\hat{p} = S_n/n \xrightarrow{P} p$ (即当 $n \rightarrow \infty$ 时依概率收敛到 p) 及式 (4.3.9) 可知, 当 $n \rightarrow \infty$ 时, 有

$$\sqrt{\frac{p(1-p)}{\hat{p}(1-\hat{p})}} \xrightarrow{P} 1, \quad \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} N(0, 1).$$

故由引理 2.5.1 可知

$$\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\hat{p}(1-\hat{p})}} = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1-p)}} \cdot \frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{\mathcal{L}} N(0, 1), \quad (4.3.12)$$

故可取 $T = \sqrt{n}(\hat{p} - p) / \sqrt{\hat{p}(1-\hat{p})}$ 作为枢轴变量, 其极限分布与 p 无关. 令

$$P\left(\left|\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\hat{p}(1-\hat{p})}}\right| \leq u_{\alpha/2}\right) \approx 1 - \alpha.$$

利用不等式等价变形, 得到 p 的置信系数近似为 $1 - \alpha$ 的置信区间为

$$\left[\hat{p} - u_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}, \hat{p} + u_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}\right], \quad (4.3.13)$$

其单侧置信上、下限为 $\hat{p} \pm u_{\alpha} \sqrt{\hat{p}(1-\hat{p})/n}$, 其中 “+” 号是置信上限.

3. Poisson 分布参数的置信区间

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为抽自 Poisson 总体 $P(\lambda)$ 的简单随机样本, 要求 λ 的置信区间.

记 $S_n = \sum_{i=1}^n X_i$, 则 S_n 为服从参数为 $n\lambda$ 的 Poisson 分布, 即

$$P(S_n = k) = \frac{e^{-n\lambda}(n\lambda)^k}{k!}, \quad k = 0, 1, 2, \dots$$

当 n 充分大时, 由中心极限定理可知

$$\frac{S_n - n\lambda}{\sqrt{n\lambda}} = \frac{\sqrt{n}(\bar{X} - \lambda)}{\sqrt{\lambda}} \xrightarrow{\mathcal{D}} N(0, 1), \quad \text{当 } n \rightarrow \infty, \quad (4.3.14)$$

将随机变量 $T = \sqrt{n}(\bar{X} - \lambda)/\sqrt{\lambda}$ 作为枢轴变量, 其极限分布与未知参数 λ 无关. 记 $\hat{\lambda} = \bar{X}$, 令

$$P(|T| \leq u_{\alpha/2}) = P\left(\left|\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\lambda}}\right| \leq u_{\alpha/2}\right) \approx 1 - \alpha.$$

仿照前一段的方法得到 λ 的置信区间 $[d_1, d_2]$, 其中 $d_1 = d_1(\mathbf{X})$, $d_2 = d_2(\mathbf{X})$ 为二次方程

$$(\hat{\lambda} - \lambda)^2 = \frac{\lambda u_{\alpha/2}^2}{n}$$

的两根, 即有

$$d_1, d_2 = \hat{\lambda} + \frac{u_{\alpha/2}^2}{2n} \pm u_{\alpha/2} \sqrt{\frac{u_{\alpha/2}^2}{4n^2} + \frac{\hat{\lambda}}{n}}, \quad d_1 \text{ 相应于 “-” 号.} \quad (4.3.15)$$

因此 λ 的置信系数近似为 $1 - \alpha$ 的置信区间为 $[d_1(\mathbf{X}), d_2(\mathbf{X})]$.

实用上, 可采用下列更简单的方法: 由 $\hat{\lambda} = \bar{X} \xrightarrow{P} \lambda$ (当 $n \rightarrow \infty$ 时), 因此当 $n \rightarrow \infty$ 时有

$$\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\lambda}} \xrightarrow{\mathcal{D}} N(0, 1), \quad \frac{\sqrt{\lambda}}{\sqrt{\hat{\lambda}}} \xrightarrow{P} 1.$$

故由引理 2.5.1 可知

$$\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\hat{\lambda}}} = \frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\lambda}} \cdot \frac{\sqrt{\lambda}}{\sqrt{\hat{\lambda}}} \xrightarrow{\mathcal{D}} N(0, 1). \quad (4.3.16)$$

令 $T = \sqrt{n}(\hat{\lambda} - \lambda)/\sqrt{\hat{\lambda}}$ 为枢轴变量, 其极限分布与未知参数 λ 无关. 给定置信系数 $1 - \alpha$, 则有

$$P\left(\left|\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\hat{\lambda}}}\right| \leq u_{\alpha/2}\right) \approx 1 - \alpha.$$

由不等式的等价变形, 得到 λ 的置信系数近似为 $1 - \alpha$ 的置信区间为

$$\left[\hat{\lambda} - u_{\alpha/2} \sqrt{\hat{\lambda}/n}, \hat{\lambda} + u_{\alpha/2} \sqrt{\hat{\lambda}/n} \right], \quad (4.3.17)$$

其单侧置信上、下限为 $\hat{\lambda} \pm u_\alpha \sqrt{\hat{\lambda}/n}$, 其中 “+” 号是置信上限.

*4. 一般情形

用渐近分布寻求参数 θ 的近似置信区间在很多场合是切实可行的. 例如, 在 3.3 节中, 定理 3.3.2 在一般条件下证明了 θ 的极大似然估计 $\hat{\theta}_n^* = \hat{\theta}_n^*(X_1, \dots, X_n)$ 有渐近正态分布 $\sqrt{n}(\hat{\theta}_n^* - \theta) \xrightarrow{\mathcal{L}} N(0, 1/I(\theta))$, 即 $\hat{\theta}_n^*$ 渐近服从 $N(\theta, \sigma^2(\theta)/n)$. 此处 $\sigma^2(\theta) = 1/I(\theta)$, 其中

$$I(\theta) = E_\theta \left[\left(\frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2 \right]$$

是 Fisher 信息量, $f(x, \theta)$ 是总体的密度函数. 当 n 很大时, 常用 $\hat{\theta}_n^*$ 去代替 $\sigma^2(\theta)$ 中的 θ , 大数定律保证 $\sigma^2(\hat{\theta}_n^*)$ 依概率收敛到 $\sigma^2(\theta)$. 由引理 2.5.1 可证当 $n \rightarrow \infty$ 时, 有

$$\frac{\sqrt{n}(\hat{\theta}_n^* - \theta)}{\sigma(\hat{\theta}_n^*)} = \frac{\sqrt{n}(\hat{\theta}_n^* - \theta)}{\sigma(\theta)} \cdot \frac{\sigma(\theta)}{\sigma(\hat{\theta}_n^*)} \xrightarrow{\mathcal{L}} N(0, 1). \quad (4.3.18)$$

令 $T = \sqrt{n}(\hat{\theta}_n^* - \theta)/\sigma(\hat{\theta}_n^*)$ 作为枢轴变量, 它的极限分布与 θ 无关, 则有

$$P \left(\left| \frac{\sqrt{n}(\hat{\theta}_n^* - \theta)}{\sigma(\hat{\theta}_n^*)} \right| \leqslant u_{\alpha/2} \right) \approx 1 - \alpha,$$

解括号中的不等式 可得 θ 的置信系数近似为 $1 - \alpha$ 的置信区间

$$\left[\hat{\theta}_n^* - \frac{u_{\alpha/2}}{\sqrt{n}} \sigma(\hat{\theta}_n^*), \hat{\theta}_n^* + \frac{u_{\alpha/2}}{\sqrt{n}} \sigma(\hat{\theta}_n^*) \right]. \quad (4.3.19)$$

其单侧置信上、下限为 $\hat{\theta}_n^* \pm u_\alpha \sigma(\hat{\theta}_n^*)/\sqrt{n}$, 其中 “+” 号是置信上限.

*5. 非参数情形

设某总体 F 有均值 θ_F 和方差 σ_F^2 , 且 θ_F 和 σ_F^2 皆未知. 从这一总体中抽取简单随机样本 X_1, \dots, X_n , 要求 θ_F 的置信区间.

因为对总体分布没有作任何假定, 这是非参数型的分布族. 要找出适合小样本情形的枢轴变量是不可能的. 但是, 若 n 充分大, 则由中心极限定理可知 $\sqrt{n}(\bar{X} - \theta_F)/\sigma_F \xrightarrow{\mathcal{L}} N(0, 1)$, 当 $n \rightarrow \infty$ 时. 但此处 σ_F 未知, 仍不能以 $\sqrt{n}(\bar{X} - \theta_F)/\sigma_F$ 作为枢轴变量. 因为 n 充分大, 样本标准差 S 是 σ_F 的一个相合估计, 故可近似地用 S 代替 σ_F , 利用引理 2.5.1 可证, 当 $n \rightarrow \infty$ 时, 有

$$\frac{\sqrt{n}(\bar{X} - \theta_F)}{S} = \frac{\sqrt{n}(\bar{X} - \theta_F)}{\sigma_F} \cdot \frac{\sigma_F}{S} \xrightarrow{\mathcal{L}} N(0, 1). \quad (4.3.20)$$

此时可将 $T = \sqrt{n}(\bar{X} - \theta_F)/S$ 作为枢轴变量, 它的极限分布与 θ_F 无关, 令

$$P\left(\left|\frac{\sqrt{n}(\bar{X} - \theta_F)}{S}\right| \leq u_{\alpha/2}\right) \approx 1 - \alpha,$$

解上式括号中的不等式, 得到 θ_F 的置信系数近似为 $1 - \alpha$ 的置信区间为

$$\left[\bar{X} - \frac{S}{\sqrt{n}}u_{\alpha/2}, \bar{X} + \frac{S}{\sqrt{n}}u_{\alpha/2}\right], \quad (4.3.21)$$

其单侧置信上、下限为 $\bar{X} \pm S u_{\alpha}/\sqrt{n}$, 其中 “+” 号是置信上限.

*4.4 Fisher 的信仰推断法

4.4.1 引言

本章前几节讨论了构造 Neyman 的置信区间的方法, 这是目前应用最广泛的区间估计的形式. 它的好处在于理论上是基于柯尔莫哥洛夫公理体系下的概率论, 在实用上允许给出频率的解释. 在本章后面这两节将介绍区间估计的两种其他方法. 差不多在 Neyman 发表其置信区间理论的同时 R.A. Fisher 提出了一种求区间估计的方法, 这个方法原则上可用于任何统计推断问题, 因而它不仅是一个方法上的问题, 而且还代表了对待统计推断问题的一种根本不同的观点. 这就是本节将要介绍的 Fisher 信仰推断法.

Fisher 认为在有了样本 x 后, 就有了参数 θ 的一个分布. 这个分布表示由于样本所提供的信息, θ 落在了各个范围内的“可信程度”. 下面举例来说明. 设 X_1, \dots, X_n 为从 $N(\theta, 1)$ 中抽取的简单随机样本, 要求 θ 的区间估计. 假设在抽样之前对 θ 一无所知. 因此对事件 “ $-1 \leq \theta \leq 1$ ” 和 “ $150 \leq \theta \leq 152$ ” 发生的可能性谁大谁小, 无法断言. 但在抽取样本后, 由于样本中所包含 θ 的信息, 使得对上述两事件发生可能性将有某种估计, 即赋予不同的“信任程度”或称之为“信仰(信任)概率”(fiducial probability). 例如, 由样本算出 $\bar{X} = 0$, 而样本容量 n 比较大, 则将认为 “ $-1 \leq \theta \leq 1$ ” 发生的可能性比 “ $150 \leq \theta \leq 152$ ” 发生的可能性大得多. 这是因为 θ 是总体均值, \bar{X} 为样本均值, 当 n 较大时, \bar{X} 和 θ 很接近. 所以由 $\bar{X} = 0$ 可断定 θ 取值在 0 附近. 因此如果能对这种信仰分布(概率)作一数量上的刻画, 则就可利用它作出 θ 的区间估计. 这种区间估计称为信仰区间, 信仰区间的可靠程度称为信仰系数. 下面将通过例子说明信仰区间的求法.

4.4.2 信仰区间的求法

例 4.4.1 设样本 $\mathbf{X} = (X_1, \dots, X_n)$ 为自 $N(\theta, 1)$ 中抽取的简单随机样本, 其中 $-\infty < \theta < +\infty$, 求 θ 的信仰区间.

解 由于 \bar{X} 是 θ 的充分统计量, 所以可基于 \bar{X} 来考虑该问题. 因为 $\bar{X} \sim N(\theta, 1/n)$, 若记随机误差 $e \sim N(0, 1)$, 则等价地有

$$\bar{X} = \theta + \frac{1}{\sqrt{n}} e. \quad (4.4.1)$$

移项可得

$$\theta = \bar{X} - \frac{1}{\sqrt{n}} e. \quad (4.4.2)$$

由式 (4.4.1) 到式 (4.4.2), 从通常概率的观点来看, 不过是写法的不同, 没有什么实质性的新内容, 但 Fisher 却赋予式 (4.4.2) 以完全新的解释: 有了样本 X 后, 并继而得到 \bar{X} (即 \bar{X} 看成一个已知数), 把 θ 看成一个随机变量, 式 (4.4.2) 给出了 θ 的分布, $\theta \sim N(\bar{X}, 1/n)$, Fisher 称这个分布为 θ 的信仰分布.

为什么可以把 θ 看成随机变量呢? 按 Fisher 的意思, 在进行抽样得到 X , 继而得到 \bar{X} 之前, θ 是一个未知数, 对它茫然无知, 就对它的了解而言, 它什么都可能取, 取什么值的可能性有多大, 一点也说不上. 但在抽取样本 X , 继而得到 \bar{X} 后, 虽然这时 θ 仍有可能取各种值, 但对取这些值的可能性如何, 通过样本 X 提供的信息 (或者说由 \bar{X} 提供的信息, 因为 \bar{X} 为 θ 的充分统计量), 就有些知识了. 例如, 考虑到 θ 是总体均值, 就会觉得 θ 落在样本均值 \bar{X} 附近的可能性, 比远离 \bar{X} 的可能性 (似然性) 大 (可以把这个说法与“似然性”概念联系起来 —— 从这个意义上讲“信仰推断法”是 Fisher“极大似然方法”的一个推广). 这只是一个笼统的、非定量的说法, 下面的信仰概率式 (4.4.3) 则对 θ 取各种值的可能性给予了定量的刻画 (在得到 \bar{X} 的背景下).

记 \tilde{P} 表示信仰概率, P 为通常意义上的概率, 利用信仰分布 (4.4.2), 对任意的 $a, b (a < b)$, 算出事件 $\{a \leq \theta \leq b\}$ 的信仰概率为

$$\tilde{P}(a \leq \theta \leq b) = \int_a^b \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n(y-\bar{x})^2}{2}} dy. \quad (4.4.3)$$

由式 (4.4.2) 可知 $\sqrt{n}(\theta - \bar{X}) = -e \sim N(0, 1)$, e 是关于 0 对称分布, 故 $-e$ 与 e 同分布. 令

$$\tilde{P}(|\sqrt{n}(\theta - \bar{X})| \leq c) = P(|e| \leq c) = 1 - \alpha,$$

则 $c = u_{\alpha/2}$. 等价地有

$$\tilde{P}\left(\bar{X} - \frac{u_{\alpha/2}}{\sqrt{n}} \leq \theta \leq \bar{X} + \frac{u_{\alpha/2}}{\sqrt{n}}\right) = 1 - \alpha,$$

其中 $1 - \alpha$ 称为信仰系数 (fiducial coefficient). $[\bar{X} - u_{\alpha/2}/\sqrt{n}, \bar{X} + u_{\alpha/2}/\sqrt{n}]$ 称为 θ 的信仰系数为 $1 - \alpha$ 的信仰区间 (fiducial interval). 由 Neyman 置信区间理论

可知, 这个区间也是 θ 的置信系数为 $1 - \alpha$ 的置信区间. 因此在本例中两种方法得到同一种结果, 虽则其解释根本不同. 但有时信仰系数为 $1 - \alpha$ 的区间估计并不是置信系数为 $1 - \alpha$ 的区间估计, 如本节最后将要介绍的 Behrens-Fisher 问题所示.

例 4.4.2 设样本 $\mathbf{X} = (X_1, \dots, X_n)$ 为自正态分布总体 $N(\theta, \sigma^2)$ 中抽取的简单随机样本, 其中 $-\infty < \theta < +\infty$, $\sigma^2 > 0$, 求 θ 和 σ^2 的信仰区间.

解 设 \bar{X} 为样本均值. 记

$$Q^2 = (n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2, \quad (4.4.4)$$

其中 S^2 为样本方差, 由于 (\bar{X}, Q^2) 是 (θ, σ^2) 的充分统计量, 所以可以基于 (\bar{X}, Q^2) 来考虑问题. 因 $\bar{X} \sim N(\theta, \sigma^2/n)$, $Q^2/\sigma^2 \sim \chi_{n-1}^2$, 并且 \bar{X} 和 Q^2 相互独立. 若记 $e_1 \sim N(0, 1)$, $e_2 \sim \chi_{n-1}^2$ 且 e_1, e_2 相互独立, 则等价地有

$$\begin{cases} \bar{X} = \theta + \frac{\sigma}{\sqrt{n}} \cdot e_1, \\ Q^2 = \sigma^2 \cdot e_2. \end{cases} \quad (4.4.5)$$

将式 (4.4.5) 第 2 式中 $\sigma = \sqrt{Q^2/e_2}$ 代入到第 1 式, 移项得到

$$\begin{cases} \theta = \bar{X} - \frac{1}{\sqrt{n}} \frac{Q}{\sqrt{e_2}} \cdot e_1, \\ \sigma^2 = \frac{Q^2}{e_2}. \end{cases} \quad (4.4.6)$$

在有了样本 \mathbf{X} , 从而有了 (\bar{X}, Q^2) 后, 可诱导出 (θ, σ^2) 的联合信仰分布. 由于 $e_1 \sim N(0, 1)$, $e_2 \sim \chi_{n-1}^2$ 且 e_1 和 e_2 相互独立, 则 (θ, σ^2) 联合信仰分布为

$$p(\theta, \tau) = c \cdot \exp \left\{ -\frac{n(\theta - \bar{X})^2 + Q^2}{2\tau} \right\} \cdot (\tau)^{-\frac{n+2}{2}},$$

$$-\infty < \mu < +\infty, \tau > 0, \quad (4.4.7)$$

其中 $\tau = \sigma^2$, c 为正则化常数, 使得上述密度函数积分为 1. 由此联合信仰分布可分别导出 θ 和 σ^2 的边缘分布. 当然, 也可以直接由式 (4.4.6) 分别导出 θ 和 σ^2 的边缘分布. 例如, 在式 (4.4.6) 第 1 式中将 $Q = \sqrt{n-1} S$ 带入, 得到

$$\frac{\sqrt{n}(\theta - \bar{X})}{S} = -\frac{e_1}{\sqrt{e_2/(n-1)}} \sim t_{n-1}.$$

因此 θ 的边缘信仰分布为自由度 $n-1$ 的 t 分布. 若令

$$\tilde{P}\left(\left|\frac{\sqrt{n}(\theta - \bar{X})}{S}\right| \leq c\right) = P\left(\left|\frac{e_1}{\sqrt{e_2/(n-1)}}\right| \leq c\right) = 1 - \alpha,$$

则 $c = t_{n-1}(\alpha/2)$. 上式等价于

$$\tilde{P}\left(\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2) \leq \theta \leq \bar{X} + \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2)\right) = 1 - \alpha.$$

因此 θ 的信仰系数为 $1 - \alpha$ 的置信区间为

$$\left[\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2), \bar{X} + \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2)\right].$$

这与 4.2 节中求得的 θ 的置信系数为 $1 - \alpha$ 的置信区间完全相同.

由式 (4.4.6) 第 2 式可得

$$\frac{\sigma^2}{Q^2} = \frac{1}{e_2}, \quad e_2 \sim \chi_{n-1}^2.$$

因此 σ^2 的边缘信仰分布是一个逆 χ^2 分布. 若令

$$\tilde{P}\left(c \leq \frac{\sigma^2}{Q^2} \leq d\right) = P\left(\frac{1}{d} \leq e_2 \leq \frac{1}{c}\right) = 1 - \alpha,$$

由于 $Q^2/\sigma^2 = e_2 \sim \chi_{n-1}^2$, 故可取 $1/d = \chi_{n-1}^2(1 - \alpha/2)$, $1/c = \chi_{n-1}^2(\alpha/2)$, 即有

$$\tilde{P}\left(\frac{Q^2}{\chi_{n-1}^2(\alpha/2)} \leq \sigma^2 \leq \frac{Q^2}{\chi_{n-1}^2(1 - \alpha/2)}\right) = 1 - \alpha.$$

因此 σ^2 的信仰系数为 $1 - \alpha$ 的信仰区间为

$$\left[\frac{Q^2}{\chi_{n-1}^2(\alpha/2)}, \frac{Q^2}{\chi_{n-1}^2(1 - \alpha/2)}\right].$$

这与 4.2 节中求得的 σ^2 的置信系数为 $1 - \alpha$ 的置信区间也完全相同.

在以上的例子中, 信仰分布的求得基本上是个形式上的转换, 由其所确定的信仰区间, 与 Neyman 理论所确定的置信区间也完全一样. 因此, 在 Fisher 方法发表的初期, 人们认为它与 Neyman 方法是同一件事的两种不同的说法. 然而, 往后的发展使人们清楚地看到, 这二者是不同的, 具体结果也不同. 这一点可以从下面要介绍的 Behrens-Fisher 问题中看出来.

在 Neyman 的理论中, θ 是一个虽未知, 但是为非随机的常数, 说不上有什么分布而言. 这个理论不需要在传统概率论之外引进什么新概念; 它在求具体参数的区间估计时所涉及的计算和推理, 都是在传统的概率论中早已确立的框架内.

Fisher 的信仰推断理论则不然, 它虽然也要借用传统概率的结果, 但有两个根本问题要解决:

(1) 信仰分布究竟是什么. 在基于频率解释的概率论中, 概率有一种可诉诸实验进行验证的频率解释. 在 Fisher 的理论中, 信仰分布为人们对 θ 取各种值的信任程度的刻画. 但这种信任程度因人而异, 且无法诉诸实验去验证.

(2) 怎样确定信仰分布, 即要制订一些合理的规则, 使用这些规则就可以在已知样本具体值及样本分布时, 唯一地定出参数的信仰分布.

这个问题至今未解决. Fisher 在充分统计量存在时, 指出了如何确定信仰分布, 如前面讨论过的两个例子. 但是, 充分统计量不是经常存在的. 即使在充分统计量存在的情形, Fisher 也没有指出一种足以唯一确定信仰分布的普遍方法.

由于这些原因, Fisher 的思想在统计学界引起相当大的兴趣和争论. 大约有这样几种看法: 有的学者, 如 Neyman, 对此持完全否定的态度; 有的人承认这理论不完善, 但不妨探索一下这些问题可否解决; 另一种持实用观点的人认为, 不管其基础如何, Fisher 的方法在某些问题中提供了可用的解法, 在这种情况下, 不必因其基础问题未解决而拒绝使用.

这里还要说明的是 Fisher 在提出信仰分布的概念时, 就特别将它与 Bayes 统计中的后验分布划清了界限, 认为前者不需对参数作任何先验的假定.

4.4.3 用 Fisher 方法解 Behrens-Fisher 问题

4.2 节已提过, Behrens-Fisher 问题是这样一个问题: 设 $\mathbf{X} = (X_1, \dots, X_m)$ 是从正态总体 $N(\mu_1, \sigma_1^2)$ 中抽取的简单随机样本, $\mathbf{Y} = (Y_1, \dots, Y_n)$ 是从正态总体 $N(\mu_2, \sigma_2^2)$ 中抽取的简单随机样本, 其中 $-\infty < \mu_1, \mu_2 < +\infty, \sigma_1^2, \sigma_2^2 > 0$ 为未知参数, 且合样本 \mathbf{X}, \mathbf{Y} 独立. 要求 $\theta = \mu_2 - \mu_1$ 的精确的区间估计. 这个问题是 Behrens 于 1929 年提出的, Fisher 用他提出的信仰推断法给出了一个解法, 所以人们习惯称这个问题为 Behrens-Fisher 问题.

这个问题在实用上有很重要的意义. 当 $\sigma_1^2 = \sigma_2^2$ 但未知时, 区间估计问题已在 4.2 节得到解决. 在许多情况下, 两总体方差相等的假定未必成立. 不少学者在 Neyman 理论的范围内提出了一些解法, 它们在其意义下是不精确的, 即其置信系数只是近似而非精确地等于指定的 $1 - \alpha$. 1943 年, Scheffé 提出过一种解法, 他在上述意义下是精确的, 可是这个解法有一个很大的缺点: 其解依赖样本排列的次序. 然而样本的次序应与问题的解无关, 无论如何都是一个合理的要求. 除了 Scheffé 的精确解法, Welch 在 1938 年给出了 Behrens-Fisher 问题的一个近似解法, Welch 的近似解法较 Scheffé 的精确解法简单, 并且他们给出的区间估计相差也不大, 所以在实用中通常使用 Welch 的近似解法. 4.2 节已对 Welch 的近似解法作过简单的介绍.

现在来叙述 Fisher 基于信仰分布的解法. 令 \bar{X}, \bar{Y}, S_1^2 和 S_2^2 分别表示样本 \mathbf{X} 和 \mathbf{Y} 的样本均值和样本方差, 记 $Q_1^2 = (m-1)S_1^2, Q_2^2 = (n-1)S_2^2$.

由于 $\bar{X} \sim N(\mu_1, \sigma_1^2/m)$, $\bar{Y} \sim N(\mu_2, \sigma_2^2/n)$, $Q_1^2/\sigma_1^2 \sim \chi_{m-1}^2$, $Q_2^2/\sigma_2^2 \sim \chi_{n-1}^2$; 若记随机误差 $e_i \sim N(0, 1)$, $i = 1, 2$; $\xi_1 \sim \chi_{m-1}^2$, $\xi_2 \sim \chi_{n-1}^2$; e_1 和 e_2 , ξ_1 和 ξ_2 相互独立, 则等价地有

$$\begin{cases} \bar{X} = \mu_1 + \frac{\sigma_1}{\sqrt{m}} \cdot e_1, \\ Q_1^2 = \sigma_1^2 \cdot \xi_1, \\ \bar{Y} = \mu_2 + \frac{\sigma_2}{\sqrt{n}} \cdot e_2, \\ Q_2^2 = \sigma_2^2 \cdot \xi_2. \end{cases} \quad (4.4.8)$$

从式 (4.4.8) 的第 2, 4 两个式子中解得 $\sigma_1 = Q_1/\sqrt{\xi_1}$, $\sigma_2 = Q_2/\sqrt{\xi_2}$, 将它们分别代入到式 (4.4.8) 的第 1, 3 两个式子中, 整理移项得到

$$\begin{cases} \mu_1 = \bar{X} - \frac{Q_1}{\sqrt{m}} \cdot \frac{e_1}{\sqrt{\xi_1}}, \\ \sigma_1^2 = Q_1^2/\xi_1, \\ \mu_2 = \bar{Y} - \frac{Q_2}{\sqrt{n}} \cdot \frac{e_2}{\sqrt{\xi_2}}, \\ \sigma_2^2 = Q_2^2/\xi_2. \end{cases} \quad (4.4.9)$$

将 $Q_1 = \sqrt{m-1}S_1$ 和 $Q_2 = \sqrt{n-1}S_2$ 分别代入到式 (4.4.9) 的第 1, 3 两式, 整理得

$$\begin{cases} \mu_1 = \bar{X} - \frac{S_1}{\sqrt{m}}t_1 = \bar{X} - S_1^*t_1, \\ \mu_2 = \bar{Y} - \frac{S_2}{\sqrt{n}}t_2 = \bar{Y} - S_2^*t_2, \end{cases} \quad (4.4.10)$$

其中 $t_1 = \frac{e_1}{\sqrt{\xi_1/(m-1)}} \sim t_{m-1}$, $t_2 = \frac{e_2}{\sqrt{\xi_2/(n-1)}} \sim t_{n-1}$, $S_1^* = S_1/\sqrt{m}$, $S_2^* = S_2/\sqrt{n}$. 记 $Z = \bar{Y} - \bar{X}$, $\theta = \mu_2 - \mu_1$, 将式 (4.4.10) 中的两式相减, 得到

$$Z - \theta = S_2^*t_2 - S_1^*t_1, \quad \theta = Z - (S_2^*t_2 - S_1^*t_1). \quad (4.4.11)$$

一旦有了样本, 算出 $Z = z$ 和 $S_i^* = s_i^*$, $i = 1, 2$ 时, 式 (4.4.11) 就表示了 θ 的信仰分布. 式 (4.4.11) 第 1 式右边是两个相互独立的 t 变量的线性组合, 将其改写成如下形式:

$$S_2^*t_2 - S_1^*t_1 = \gamma(t_2 \cos \varphi - t_1 \sin \varphi),$$

其中 $\gamma = \sqrt{S_1^{*2} + S_2^{*2}}$, $\cos \varphi = S_2^*/\gamma$, $0 < \varphi < \pi/2$. 记 $W = t_2 \cos \varphi - t_1 \sin \varphi$, 可以证明 W 的分布关于原点对称, 它的分布函数只依赖于 m , n 和 φ . 以 $F_{m,n,\varphi}$ 记这一分布函数, 找 $y_{m,n,\varphi,\alpha} > 0$, 使得

$$F_{m,n,\varphi}(y_{m,n,\varphi,\alpha}) - F_{m,n,\varphi}(-y_{m,n,\varphi,\alpha}) = 1 - \alpha,$$

则由 θ 的信仰分布 (4.4.11) 可得

$$\tilde{P}(|\theta - Z| \leqslant \gamma y_{m,n,\varphi,\alpha}) = P(|W| \leqslant y_{m,n,\varphi,\alpha}) = 1 - \alpha.$$

此处 \tilde{P} 为信仰概率, P 是通常意义下的概率. 上式的等价的形式为

$$\tilde{P}(Z - \gamma y_{m,n,\varphi,\alpha} < \theta < Z + \gamma y_{m,n,\varphi,\alpha}) = 1 - \alpha. \quad (4.4.12)$$

因此, $\theta = \mu_2 - \mu_1$ 的信仰系数为 $1 - \alpha$ 的信仰区间为

$$[Z - \gamma y_{m,n,\varphi,\alpha}, Z + \gamma y_{m,n,\varphi,\alpha}].$$

注意在有了样本并给定 α 后, Z , γ , $y_{m,n,\varphi,\alpha}$ 都可以定出, Fisher 和 Yates 曾给出 $y_{m,n,\varphi,\alpha}$ 的表.

Neyman 通过计算证明, 若把式 (4.4.12) 视为一个置信区间, 其置信系数 (而非信仰系数) 并非 $1 - \alpha$. Neyman 指出: 区间 (4.4.12) 包含被估计的 $\theta = \mu_2 - \mu_1$ 的概率 (是指通常的概率而非信仰概率!) 依赖于比值 $\rho = \sigma_1/\sigma_2$, 它就 $m = 12$, $n = 6$ 和 $\alpha = 0.05$ 的情况, 算出 $\rho = 0.1$, 1.0 和 10 时, 这个概率分别为 0.966 , 0.960 和 0.934 , 而不是名义上的 0.95 . Neyman 的这个批评的意义在于, 明确了在本问题中, “置信系数”和“信仰系数”的确不是“形异实同”的东西. 因而 Fisher 的方法与 Neyman 的方法不是一回事. 使人们感兴趣的是 Neyman 算出的置信系数的值与 Fisher 信仰系数 0.95 相去不远. 人们因此觉得 Fisher 提供的解是可以放心使用的.

Fisher 将信仰推断这一方法用于解决著名的 Behrens-Fisher 问题之后, 信仰推断方法受到人们很大的关注. 这说明信仰推断方法是解决区间估计问题的一个有效方法. 另外, 信仰推断方法直观, 容易被实际工作者所接受. 但是随着研究的深入, 人们发现了信仰推断方法的一些内在问题, 正如在上一段末尾所指出的信仰推断有两个根本问题未得到解决, 还算不上是一个完善的统计推断的理论和方法.

*4.5 容忍区间与容忍限

本节要讨论的问题, 其提法与区间估计并无共同之处, 但其解与区间估计有形式上的相似性, 都是用一个由样本确定的区间或上、下限的形式表达, 因此把它放在本章末尾, 作一简单介绍.

4.5.1 问题的提法及定义

设有分布族 $\{F_\theta, \theta \in \Theta\}$, Θ 是参数空间, $\mathbf{X} = (X_1, \dots, X_n)$ 为从分布族某总体中抽取的简单样本. 本节不是考虑参数 θ 的置信区间, 而是考虑随机变量 X 的

“置信区间”，称之为容忍区间，即希望求 $T_1 = T_1(\mathbf{X})$ 和 $T_2 = T_2(\mathbf{X})$ ，使得对给定的 $0 < \beta < 1$ ，

$$P_{\theta}^* \left\{ X \in [T_1(\mathbf{X}), T_2(\mathbf{X})] \right\} \geq 1 - \beta, \quad (4.5.1)$$

其中 P_{θ}^* 应理解为在给定 \mathbf{X} 的条件下关于 r.v. X 的分布计算的条件概率。请看以下两个例子。

例 4.5.1 设某轴承厂有一部自动化机器，生产直径为 0.25 cm 的轴承，允许误差为 0.001 cm。生产中要求 99% 的产品达到以上规定，即要求轴承的直径落在 $[0.249, 0.251]$ 。今对一批轴承抽取 n 件，测得其直径为 X_1, \dots, X_n ，问这一批轴承是否合格？

设随机变量 X 表示轴承的直径，其分布函数 $F_{\theta}(x)$ 。解决这一问题的方法，就是由样本 $\mathbf{X} = (X_1, \dots, X_n)$ 确定两个统计量 $T_1 = T_1(\mathbf{X})$ 和 $T_2 = T_2(\mathbf{X})$ ，使得 $P_{\theta}^*(X \in [T_1, T_2]) = F_{\theta}(T_2) - F_{\theta}(T_1) \geq 0.99$ ，若 $[T_1, T_2] \subset [0.249, 0.251]$ ，则说明此批轴承合格。这就归结为求容忍区间的问题。

例 4.5.2 钢厂生产某种钢材，要求其钢材的强度不少于 ξ_0 （如 $\xi_0 = 120$ 单位强度）。若生产中钢材强度有 99% 符合上述要求，认为此批钢材合格。今对一批钢材测试了 n 根，强度为 X_1, \dots, X_n ，问这批钢材是否合格？

设钢材的强度为随机变量 X ，其分布函数为 $F_{\theta}(x)$ 。解决这一问题的一种方法是由样本 $\mathbf{X} = (X_1, \dots, X_n)$ 确定一个统计量 $T_L = T_L(\mathbf{X})$ ，使得 $P_{\theta}^*(X \in [T_L, \infty)) = 1 - F_{\theta}(T_L) \geq 0.99$ ，若 $T_L \geq \xi_0$ ，则说明这批钢材合格，这就归结为求容忍下限的问题。

将上述两个例子提出的问题统一在一个模型下，描述如下：设有分布族 $\{F_{\theta}, \theta \in \Theta\}$ 。令 $\mathbf{X} = (X_1, \dots, X_n)$ 为从此分布族中抽取的简单样本，要找到两个统计量 $T_1 = T_1(\mathbf{X})$ 和 $T_2 = T_2(\mathbf{X})$ ，使得对 $0 < \beta < 1$ 有

$$P_{\theta}^*(T_1 \leq X \leq T_2) = F_{\theta}(T_2) - F_{\theta}(T_1) \geq 1 - \beta, \quad (4.5.2)$$

或找一个统计量 $T_L = T_L(\mathbf{X})$ ，使得对 $0 < \beta < 1$ 有

$$P_{\theta}^*(T_L \leq X < +\infty) = 1 - F_{\theta}(T_L) \geq 1 - \beta, \quad (4.5.3)$$

或找一个统计量 $T_U = T_U(\mathbf{X})$ ，对 $0 < \beta < 1$ 有

$$P_{\theta}^*(-\infty < X \leq T_U) = F_{\theta}(T_U) \geq 1 - \beta. \quad (4.5.4)$$

在式 (4.5.2) 中，由于 T_1 和 T_2 为随机变量，故 $F_{\theta}(T_1)$ 和 $F_{\theta}(T_2)$ 也是随机变量，所以 “ $F(T_2) - F(T_1) \geq 1 - \beta$ ” 是一个随机事件。显然不能保证这一事件绝对发生，于

是只能降低要求：给定 r (通常 $0 < r < 1$)，要求 “ $F(T_2) - F(T_1) \geq 1 - \beta$ ” 这个事件至少以概率 $1 - r$ 成立，即

$$P_\theta \left\{ F(T_2) - F(T_1) \geq 1 - \beta \right\} \geq 1 - r.$$

对式 (4.5.3) 和式 (4.5.4) 中的 T_L, T_U 也可以提出类似要求，这就引导到容忍区间和容忍限的概念。下面给出定义。

定义 4.5.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $X \sim \{F_\theta, \theta \in \Theta\}$ 中抽取的简单样本。又设 $T_1 = T_1(\mathbf{X})$ 和 $T_2 = T_2(\mathbf{X})$ 是两个统计量，且 $T_1 \leq T_2$ ，若对任意给定的 β, γ (通常取较小的数，如 $\beta = 0.05, \gamma = 0.01$)， $0 < \beta, \gamma < 1$ 有

$$\begin{aligned} & P_\theta \left\{ P_\theta^*(T_1 \leq X \leq T_2) \geq 1 - \beta \right\} \\ &= P_\theta \left\{ F_\theta(T_2) - F_\theta(T_1) \geq 1 - \beta \right\} \geq 1 - \gamma, \quad \text{一切 } \theta \in \Theta, \end{aligned} \quad (4.5.5)$$

则称 $[T_1, T_2]$ 是 F_θ 的一个水平为 $(1 - \beta, 1 - \gamma)$ 的容忍区间 (tolerance interval)。

设 $T_L = T_L(\mathbf{X})$ 和 $T_U = T_U(\mathbf{X})$ 是两个统计量，若对任意给定的 $\beta, \gamma, 0 < \beta, \gamma < 1$ ，一切 $\theta \in \Theta$ ，分别有

$$\begin{aligned} P_\theta \left\{ P_\theta^*(T_L \leq X) \geq 1 - \beta \right\} &= P_\theta \left\{ 1 - F_\theta(T_L) \geq 1 - \beta \right\} \\ &= P_\theta \left\{ F_\theta(T_L) \leq \beta \right\} \geq 1 - \gamma, \end{aligned} \quad (4.5.6)$$

$$P_\theta \left\{ P_\theta^*(X \leq T_U) \geq 1 - \beta \right\} = P_\theta \left\{ F_\theta(T_U) \geq 1 - \beta \right\} \geq 1 - \gamma, \quad (4.5.7)$$

则称 T_L 和 T_U 分别是 F_θ 的一个水平为 $(1 - \beta, 1 - \gamma)$ 容忍下限 (tolerance lower limit) 和容忍上限 (tolerance upper limit)。

注意上述定义中的两个 P_θ 的含义是不同的，里面的 P_θ^* 的意义如式 (4.5.1) 中所述，外面的 P_θ 是按样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的联合分布来计算的。

容忍区间和容忍限之间有下列关系。

引理 4.5.1 若 $T_2 = T_2(\mathbf{X})$ 和 $T_1 = T_1(\mathbf{X})$ 分别是分布 F_θ 的水平为 $(1 - \beta/2, 1 - \gamma/2)$ 的容忍上、下限，且总有 $T_2 \geq T_1$ ，则 $[T_1, T_2]$ 为 F_θ 的水平为 $(1 - \beta, 1 - \gamma)$ 容忍区间。

证 令 A 表示事件 “ $F(T_1) \leq \beta/2$ ”； B 表示事件 “ $F(T_2) \geq 1 - \beta/2$ ”； C 表示事件 “ $F(T_2) - F(T_1) \geq 1 - \beta$ ”，则由定义 4.5.1 可知

$$P_\theta(A) \geq 1 - \gamma/2, \quad P_\theta(B) \geq 1 - \gamma/2. \quad (4.5.8)$$

希望证明 $P_\theta(C) \geq 1 - \gamma$ 。由以上定义可知，若 A, B 同时成立，则必有 $F(T_2) -$

$F(T_1) \geq 1 - \beta$, 即 C 成立, 因此 $AB \subset C$, 故有 $P_\theta(C) \geq P_\theta(AB)$, 由此可得

$$\begin{aligned} P_\theta(C) &= P_\theta\left\{F_\theta(T_2) - F_\theta(T_1) \geq 1 - \beta\right\} \geq P(AB) \\ &= P(A) + P(B) - P(A \cup B) \\ &\geq (1 - \gamma/2) + (1 - \gamma/2) - 1 = 1 - \gamma. \end{aligned}$$

引理证毕.

4.5.2 正态总体的容忍区间和容忍限

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为自总体 $X \sim N(\mu, \sigma^2)$ 中抽取的简单随机样本, $\theta = (\mu, \sigma^2)$ 的充分统计分量为

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

将基于充分统计量 (\bar{X}, S^2) 来构造正态总体的容忍限和容忍区间. 在总体 $N(\mu, \sigma^2)$ 中, 若 μ 和 σ^2 已知, 则水平为 $(1-\beta, 1-\gamma)$ 的容忍上下限和容忍区间分别为 $\mu + \sigma u_\beta$, $\mu - \sigma u_\beta$ 和 $[\mu - \sigma u_{\beta/2}, \mu + \sigma u_{\beta/2}]$. 但现在 μ 和 σ^2 未知, 知道 \bar{X} 和 S^2 分别是 μ 和 σ^2 的良好估计, 因此将上述容忍上限中的 μ 和 σ 用 \bar{X} 和 S 代替得到 $\bar{X} + Su_\beta$. 由于估计而带来的随机性, 水平 $(1-\beta, 1-\gamma)$ 的容忍上限不见得正好是 $\bar{X} + Su_\beta$, 而可能要将系数 u_β 修改为某个 λ , λ 既与 β 有关, 也与 γ 有关 (注意 u_β 与 γ 无关). 容忍下限和容忍区间也如此处理.

因此首先来求容忍上限, 即找 λ 使 $\bar{X} + \lambda S$ 为水平 $(1-\beta, 1-\gamma)$ 的容忍上限. 按定义, 对给定的 β 和 γ , $0 < \beta, \gamma < 1$, 要确定 λ , 使得

$$P_\theta\left\{P_\theta^*(X \leq \bar{X} + \lambda S) \geq 1 - \beta\right\} \geq 1 - \gamma.$$

由于 $(X - \mu)/\sigma \sim N(0, 1)$, 其分布函数为 $\Phi(\cdot)$ 所以上式左边也为

$$\begin{aligned} &P_\theta\left\{P_\theta^*\left(\frac{X - \mu}{\sigma} \leq \frac{\bar{X} - \mu + \lambda S}{\sigma}\right) \geq 1 - \beta\right\} \\ &= P_\theta\left\{\Phi\left(\frac{\bar{X} - \mu + \lambda S}{\sigma}\right) \geq 1 - \beta\right\} \\ &= P_\theta\left\{\frac{\bar{X} - \mu + \lambda S}{\sigma} \geq \Phi^{-1}(1 - \beta) = u_\beta\right\}. \end{aligned} \tag{4.5.9}$$

令 $Z = \sqrt{n}(\bar{X} - \mu)/\sigma$, $S_* = S/\sigma$, 则 $Z \sim N(0, 1)$, $S_* \sim \sqrt{\chi^2_{n-1}/(n-1)}$. 因此

$$\frac{Z - \sqrt{n}u_\beta}{S/\sigma} = \frac{Z - \sqrt{n}u_\beta}{S_*} \sim t_{n-1, \delta},$$

即自由度 $n - 1$, 非中心参数 $\delta = -\sqrt{n}u_\beta$ 的非中心 t 分布. 由式 (4.5.9) 可知

$$\begin{aligned} & P_\theta \left\{ P_\theta^* \left(X \leq \bar{X} + \lambda S \right) \geq 1 - \beta \right\} \geq 1 - \gamma \\ \Leftrightarrow & P_\theta \left\{ \frac{\bar{X} - \mu + \lambda S}{\sigma} \geq u_\beta \right\} \geq 1 - \gamma \\ \Leftrightarrow & P_\theta \left\{ \frac{Z - \sqrt{n}u_\beta}{S_*} \geq -\sqrt{n}\lambda \right\} \geq 1 - \gamma. \end{aligned} \quad (4.5.10)$$

若记 $\lambda = \lambda(n, \beta, \gamma)$, 故由 $-\sqrt{n}\lambda = t_{n-1, \delta}(1 - \gamma)$, 解得 $\lambda(n, \beta, \gamma) = -t_{n-1, \delta}(1 - \gamma)/\sqrt{n}$. 因此可知, 水平为 $(1 - \beta, 1 - \gamma)$ 的容忍上限为 $\bar{X} + \lambda S$, 其中 $\lambda = t_{n-1, \delta}(\gamma)/\sqrt{n}$. 类似可求水平为 $(1 - \beta, 1 - \gamma)$ 的容忍下限为 $\bar{X} - \lambda S$, λ 同上. 对常见的 n, λ, γ 已编制了 $\lambda(n, \beta, \gamma)$ 值的表, 见附表 6.

求正态总体 $N(\mu, \sigma^2)$ 的容忍区间, 可利用引理 4.5.1 可得 $[\bar{X} - \lambda S, \bar{X} + \lambda S]$, 此处 $\lambda(n, \beta, \gamma) = t_{n-1, \delta^*}(\gamma/2)/\sqrt{n}$, 其中 $\delta^* = -\sqrt{n}u_{\beta/2}$. 当然, 现有的非中心 t 分布表还不够大, 不一定能从表上直接查到非中心 t 分布的分位数的值. 但附表 7 给出了 $\lambda(n, \beta, \gamma)$ 值的表.

例 4.5.3 某厂生产乐器用的镍合金线. 经验表明: 镍合金线的抗拉强度服从正态分布. 今从一批产品中随机抽取 10 个样品, 测得起抗拉强度 (单位: kg/mm^2) 为

10512, 10623, 10668, 10554, 10776, 1071, 10557, 10581, 10666, 10670.

求该镍合金线抗拉强度容忍下限 (设水平为 $(0.95, 0.95)$).

解 此问题中 $n = 10$, 水平为 $(0.95, 0.95)$, 即 $\beta = 0.05$, $\gamma = 0.05$. 因此 $1 - \beta = 0.95$, $1 - \gamma = 0.95$. 由数据算得

$$\bar{X} = 10632.4, \quad S^2 = 6738.77, \quad S = 82.09.$$

查附表 6 得 $\lambda = 2.91$, 因此得容许下限 $T_L = \bar{X} - \lambda S = 10632.4 - 2.91 \times 82.09 = 10393.52$. 因此这批镍合金线抗拉强度不低于 $10393.52 \text{ kg}/\text{mm}^2$.

例 4.5.4 经验表明棉纱的断裂负荷 (单位: 百分之一牛顿) 服从正态分布, 现从一批棉纱中随机抽取 12 根, 测得其断裂负荷为

$$\begin{aligned} & 228.6, 232.7, 238.8, 317.2, 315.8, 275.1, \\ & 222.2, 236.7, 224.7, 251.2, 210.4, 270.7. \end{aligned}$$

求棉纱断裂负荷的水平为 $(0.95, 0.99)$ 的容忍区间.

解 此问题中 $n = 12$, $1 - \beta = 0.95$, $1 - \gamma = 0.99$, 由数据算得

$$\bar{X} = 252.0, \quad S^2 = 1263.4, \quad S = 35.5.$$

查附表 7 得 $\lambda(12, 0.95, 0.99) = 3.87$, 由此算得

$$T_1 = \bar{X} - \lambda S = 252.0 - 3.87 \times 35.5 = 114.6,$$

$$T_2 = \bar{X} + \lambda S = 252.0 + 3.87 \times 35.5 = 389.4.$$

因此棉纱断裂负荷的水平 $(0.95, 0.99)$ 的容忍区间为 $[T_1, T_2] = [114.6, 389.4]$.

4.5.3 非参数容忍限和容忍区间

在实际问题中，还会经常遇到这样的问题，人们只知道总体分布 F 是连续的，要求此分布的容忍限和容忍区间。由于这时不知道分布的具体形式，谈不上运用分布的性质，只能利用样本给出的信息。下面讨论基于次序统计量如何给出 F 的容忍限和容忍区间。先证明一个预备知识。

引理 4.5.2 设一维随机变量 $X \sim F(x)$, $F(x)$ 是分布函数且处处连续，则 $Y = F(X)$ 服从均匀分布 $U(0, 1)$ 。

证 因为 $0 < Y < 1$, 故只需对 $0 < y < 1$ 证明 $P(Y < y) = y$ 即可。记 $t = \inf\{x : F(x) \geq y\}$, 则由 $F(x)$ 处处连续且非降, 易见 $F(t) = y$, 以及 $F(x) < y \Leftrightarrow x < t$, 故有

$$P(Y < y) = P(F(X) < y) = P(X < t) = F(t) = y.$$

引理证毕。

现设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $X \sim F(x)$ 中抽取的简单随机样本, $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ 为其次序统计量, 根据引理 4.5.2 可知, 若记 $U_i = F(X_i)$, $i = 1, 2, \dots, n$, 则 U_1, \dots, U_n i.i.d. $\sim U(0, 1)$. $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(n)}$ 为其次序统计量。

1. 求 F 的水平 $(1 - \beta, 1 - \gamma)$ 的容忍区间和容忍限

记 $U_i = F(X_i)$, $i = 1, 2, \dots, n$; $V_{ij} = U_{(j)} - U_{(i)}$, $1 \leq i < j \leq n$, 则 V_{ij} 的密度 g_{ij} 已由式 (2.3.13) 给出。若以 $[X_{(i)}, X_{(j)}]$ 作为容忍区间, 按定义有

$$P\left\{P^*(X_{(i)} \leq X \leq X_{(j)}) \geq 1 - \beta\right\} \geq 1 - \gamma. \quad (4.5.11)$$

此处 P^* 的意义与式 (4.5.1) 中 P_θ^* 的意义相同。上式左边为

$$P\left\{F(X_{(j)}) - F(X_{(i)}) \geq 1 - \beta\right\} = P(V_{ij} \geq 1 - \beta) = \int_{1-\beta}^1 g_{ij}(v) dv. \quad (4.5.12)$$

如果选择适当的 i, j 使式 (4.5.11) 中的积分不小于给定的 $1 - \gamma$, 则 $[X_{(i)}, X_{(j)}]$ 就是 F 的一个水平 $(1 - \beta, 1 - \gamma)$ 的容忍区间。形如式 (2.3.13) 的密度 g_{ij} , 称为 Beta 分布, 其参数为 $j - i$ 和 $n - j + i + 1$, 记为 $Be(j - i, n - j + i + 1)$ 。Beta 分布的参数不必为整数, 只要大于 0 就行。当 $p > 0, q > 0$ 时, $Be(p, q)$ 表示一分布, 其分布函数为

$$I_{p,q}(x) = \frac{1}{B(p, q)} \int_0^x t^{p-1} (1-t)^{q-1} dt, \quad (4.5.13)$$

其中 $B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt$, 称为 Beta 积分. 它与 Gamma 函数 $\Gamma(x)$ 有关系 $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$, 这个公式在微积中已经给出过.

式 (4.5.13) 当 $0 < x < 1$ 时, 称为不完全的 Beta 积分, Pearson 曾给它造了表. 这表可用于选择 i, j 的问题.

对 $F(x)$ 的水平 $(1-\beta, 1-\gamma)$ 的容忍上下限的问题也同样处理. 对容忍上限有

$$P\left\{ P^*(X \leq X_{(i)}) \geq 1 - \beta \right\} \geq 1 - \gamma,$$

此式左边是 $P\{F(X_{(i)}) \geq 1 - \beta\} = P(U_{(i)} \geq 1 - \beta)$. 利用式 (2.3.1), 在其中令 $F(x) = x, f(x) = 1$ 得 $U_{(i)}$ 的密度

$$f_i(x) = i \binom{n}{i} x^{i-1} (1-x)^{n-i} I_{(0,1)}(x).$$

因此有

$$P\left\{ F(X_{(i)}) \geq 1 - \beta \right\} = P(U_{(i)} \geq 1 - \beta) = \int_{1-\beta}^1 f_i(v) dv. \quad (4.5.14)$$

选择 i 使式 (4.5.14) 不小于 $1 - \gamma$, 则根据定义 $X_{(i)}$ 就是 F 的一个水平 $(1-\beta, 1-\gamma)$ 的容忍上限.

同理, 选择 j , 使得下式不小于 $1 - \gamma$, 即

$$P\left\{ F(X_{(j)}) \leq \beta \right\} = P(U_{(j)} \leq \beta) = \int_0^\beta f_j(v) dv \geq 1 - \gamma, \quad (4.5.15)$$

则 $X_{(j)}$ 就是 F 的一个水平 $(1-\beta, 1-\gamma)$ 的容忍下限.

在式 (4.5.14) 和式 (4.5.15) 中选择 i, j 使积分不小于 $1 - \gamma$, 可借助于不完全 Beta 函数求得.

2. 特例

(1) 假如取 $X_{(n)}$ 作为 F 的一个水平 $(1-\beta, 1-\gamma)$ 的容忍上限, 样本容量 n 必须有一定的要求, 否则它不能作为合适的容忍上限. 那么样本容量 n 至少应为多少? 这就把求容忍上限的问题转化为确定样本容量的问题.

由于 $X \sim F(x)$ 且 $F(x)$ 处处连续, 故 $F(X) \sim U(0, 1)$. 而 $F(X_{(n)}) = U_{(n)}$ 是来自总体 $U(0, 1)$ 的容量为 n 的样本的最大次序统计量, 其密度函数为

$$f_n(y) = ny^{n-1} I_{(0,1)}(y).$$

于是要求确定 n , 使得

$$P\{F(X_{(n)}) \geq 1 - \beta\} = P(U_{(n)} \geq 1 - \beta) \geq 1 - \gamma,$$

即

$$\int_{1-\beta}^1 ny^{n-1} dy \geq 1 - \gamma \iff 1 - (1 - \beta)^n \geq 1 - \gamma.$$

因此有

$$n \geq \frac{\ln \gamma}{\ln(1 - \beta)}.$$

对给定的 β, γ , 可以算得满足上述不等式的最小自然数 n .

(2) 类似计算可知, 若取 $X_{(1)}$ 作为 F 的一个水平 $(1 - \beta, 1 - \gamma)$ 的容忍下限, 可知 $F(X_{(1)}) = U_{(1)}$ 密度函数为

$$f_1(y) = n(1 - y)^{n-1} I_{(0,1)}(y).$$

故要使

$$P\{F(X_{(1)}) \leq \beta\} = P(U_{(1)} \leq \beta) = \int_0^\beta n(1 - y)^{n-1} dy \geq 1 - \gamma,$$

同样解出 $n \geq \ln \gamma / \ln(1 - \beta)$.

对通常用的 β, γ , 确定 F 的水平 $(1 - \beta, 1 - \gamma)$ 的容忍上下限所需要最小样本容量 n , 已编制了表, 详见附表 8, 如 $1 - \beta = 0.90, 1 - \gamma = 0.95$, 从附表 8 上查得 $n = 29$, 若 $1 - \beta = 0.95, 1 - \gamma = 0.99$ 从表上查得 $n = 90$.

(3) 若取 $[X_{(1)}, X_{(n)}]$ 作为 $F(x)$ 的一个水平为 $(1 - \beta, 1 - \gamma)$ 的容忍区间, 按定义

$$P\{F(X_{(n)}) - F(X_{(1)}) \geq 1 - \beta\} = P(U_{(n)} - U_{(1)} \geq 1 - \beta) \geq 1 - \gamma. \quad (4.5.16)$$

由于 $(U_{(1)}, U_{(n)})$ 的联合密度由式 (2.3.10) 给出, 只要取 $i = 1, j = n$, 即

$$p(y_1, y_2) = n(n-1)(y_2 - y_1)^{n-2} I_{[0 < y_1 < y_2 < 1]},$$

所以式 (4.5.16) 可改写为

$$\begin{aligned} \iint_{y_2 - y_1 \geq 1 - \beta} p(y_1, y_2) dy_1 dy_2 &= \int_0^{\beta} \int_{y_1 + (1 - \beta)}^1 n(n-1)(y_2 - y_1)^{n-2} dy_2 dy_1 \\ &\geq 1 - \gamma, \end{aligned}$$

解之可得

$$n(1 - \beta)^{n-1} - (n-1)(1 - \beta)^n \leq \gamma.$$

对给定的 β 和 γ (或等价的给定 $1 - \beta, 1 - \gamma$) 可以从上述不等式中解出 n 来.

对常用的 $1 - \beta$ 和 $1 - \gamma$, 也已编造了确定 F 的水平 $(1 - \beta, 1 - \gamma)$ 的容忍区间所需最小样本容量表, 见附表 9. 例如, $1 - \beta = 0.90$, $1 - \gamma = 0.95$, 从附表 9 上查得 $n = 46$; $1 - \beta = 0.95$, $1 - \gamma = 0.99$, 从表上查得 $n = 130$.

习 题 4

1. 对物体某指标进行 5 次测量, 得其数据为 4.781, 4.795, 4.769, 4.792, 4.779. 设指标值服从 $N(\mu, \sigma^2)$, 其中 $\sigma = 0.01$, 试求物体该指标平均值的置信系数为 0.95 的置信区间.
2. 设在上题中 σ 未知, 作置信系数分别为 0.95 和 0.99 的置信区间.
3. 设某种电子管的使用寿命服从正态分布, 从中随机抽取 15 个进行试验. 得样本均值为 1950 小时, 样本标准差 300 小时, 以 95% 的可靠度求整批电子管平均使用寿命的置信区间.
4. 设样本 X_1, \dots, X_n 取自正态总体 $N(\mu, 16)$, 为使 $[\bar{X} - 1, \bar{X} + 1]$ 是 μ 的置信系数为 0.90 的置信区间, 样本容量 n 至少应为多少?
5. 设样本 X_1, \dots, X_n 取自正态总体 $N(\mu, 16)$, 为使得 μ 的置信系数为 $1 - \alpha$ 的置信区间的长度不大于给定的 L , 样本容量 n 至少应为多少?
6. 设 X_1, \dots, X_m i.i.d. $\sim N(a, \sigma_1^2)$, Y_1, \dots, Y_n i.i.d. $\sim N(ca, \sigma_2^2)$, $c, \sigma_1^2, \sigma_2^2$ 已知, a 未知, $c \neq 0$. 又设合样本 $X_1, \dots, X_m, Y_1, \dots, Y_n$ 独立.
 - (1) 找 a 的 UMVUE;
 - (2) 基于此 UMVUE 构造 a 的一个置信系数为 $1 - \alpha$ 的置信区间.
7. 调查 10 个企业的研究经费, 得到的情况如下:

企业	1	2	3	4	5	6	7	8	9	10
研究经费所占比例	2	0	6	3	4	2	6	8	7	4

假设研究经费所占比例服从正态分布, 求方差的区间估计, 置信系数为 0.95.

8. 某电子产品的某一参数服从正态分布, 从某天生产的产品中抽取 15 只, 测得该参数为

$$3.0, 2.7, 2.9, 2.8, 3.1, 2.6, 2.5, 2.8, 2.4, 2.9, 2.7, 2.6, 3.2, 3.0, 2.8.$$

试分别对该参数的均值和方差作置信系数为 95% 的置信区间.

9. 设有 3 个同方差的正态总体, 现从中各取一个样本, 样本容量 n 和 $Q^2 = \sum_{i=1}^n (X_i - \bar{X})^2$ 的值分别为

n	6	3	8
Q^2	40	20	50

试求共同方差 σ^2 的置信系数为 0.95 的单侧置信上限.

10. 设 X_1, \dots, X_n 为取自正态总体 $N(\mu, \sigma^2)$ 的样本. 为使 $\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^{1/2} / 4$ 成为 σ 的置信系数为 0.95 的(单侧)置信下限, 样本容量 n 至少应取多少?
11. 随机地从 A 批导线中抽取 4 根, 从 B 批导线中抽取 5 根, 测量电阻(单位: Ω)为

A 批导线	0.143	0.142	0.143	0.137
B 批导线	0.140	0.142	0.136	0.138

设测试数据分别服从 $N(a_1, \sigma^2)$ 和 $N(a_2, \sigma^2)$, 并且它们相互独立, a_1, a_2, σ^2 均未知, 求 $a_1 - a_2$ 的 95% 的置信区间.

12. 枪弹的速度(单位: m/s)服从正态分布, 为了比较两种枪弹的速度, 在相同的条件下进行速度测定. 算得数据如下:

$$\begin{aligned} \text{枪弹甲: } & m = 110, \bar{X} = 2805, S_1 = 120.41; \\ \text{枪弹乙: } & n = 100, \bar{Y} = 2680, S_2 = 105.00. \end{aligned}$$

试求这两种枪弹的平均速度之差的置信水平近似为 95% 的置信区间.

13. 某电子产品有两种型号. 为了比较它们的某项参数值, 分别从这两种型号电子产品中随机抽取若干个, 测量该项参数值, 得如下数据:

$$\begin{aligned} \text{型号甲: } & 10.1, 10.3, 10.4, 9.7, 9.8; \\ \text{型号乙: } & 12.5, 12.2, 12.1, 12.0, 11.9, 11.8, 12.8. \end{aligned}$$

假设这两种型号的电子产品的该项参数值皆服从正态分布, 而且它们的方差相等. 试求它们的平均参数之差的置信系数为 95% 的置信区间.

- *14. 设 X_1, \dots, X_m 是自正态总体 $N(a, \sigma_1^2)$ 抽取的简单随机样本, Y_1, \dots, Y_n 是自正态总体 $N(b, \sigma_2^2)$ 抽取的简单随机样本, 且两组样本独立. 当 $\sigma_2^2 / \sigma_1^2 = \lambda$ 且 λ 已知时, 求 $b - a$ 的置信系数为 $1 - \alpha$ 的置信区间.

15. 有甲、乙两台机床加工同样产品, 分别从它们加工的产品中抽取若干个产品, 分别测得产品直径后, 算得如下数据:

$$\begin{aligned} \text{机床甲: } & m = 8, \bar{x} = 30.97, S_1 = 36.7; \\ \text{机床乙: } & n = 7, \bar{y} = 21.99, S_2 = 8.1. \end{aligned}$$

假设产品直径服从正态分布. 试求

- (1) 这两台机床加工产品的平均直径之差的置信水平近似为 95% 的置信区间;
 (2) 这两台机床加工精度(即方差)之比的置信水平为 90% 的置信区间.

16. 有两个化验员 A, B, 他们独立地对某种聚合物的含氯量用相同的方法作了 10 次测量, 其测定值的方差 S^2 分别为 0.5419 和 0.6065, 设 σ_A^2 和 σ_B^2 分别为 A, B 所测数据总体(设为正态总体)的方差, 求 σ_A^2 / σ_B^2 的 95% 的置信区间.

17. 设从两台机器所生产的滚珠轴承中分别抽取容量为 10 的样本, 算得它们直径的标准差分别为 $S_1 = 0.042\text{cm}$ 和 $S_2 = 0.035\text{cm}$, 假定轴承直径服从正态分布, 试求两个总体方差比的置信系数为 0.99 的置信区间.
- *18. 设 X_1, \dots, X_n 取自正态总体 $N(\mu, \sigma^2)$ 的随机样本. 求 $\theta = (\mu, \sigma^2)$ 的置信水平为 $1 - \alpha$ 的置信域 (即参数 μ 和 σ^2 同时的置信集).
- *19. 设 X_1, \dots, X_n 是来自均匀分布 $U(\theta_1, \theta_2)$ 的简单随机样本, 此处 $-\infty < \theta_1 < \theta_2 < \infty$, 记 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ 为其次序统计量. 求
- (1) $\theta_2 - \theta_1$ 的置信系数为 $1 - \alpha$ 的置信区间 (提示: 证明 $[(\theta_2 - \theta_1) - (X_{(n)} - X_{(1)})]/(\theta_2 - \theta_1)$ 分布与 θ_1, θ_2 无关);
 - (2) 求 $[X_{(1)} + X_{(n)} - (\theta_1 + \theta_2)]/(X_{(n)} - X_{(1)})$ 的分布;
 - (3) 求 $(\theta_1 + \theta_2)/2$ 的置信水平为 $1 - \alpha$ 的置信区间.
- *20. 设 X_1, \dots, X_m i.i.d. $\sim U(0, \theta_1)$, Y_1, \dots, Y_n i.i.d. $\sim U(0, \theta_2)$, $\theta_1 > 0, \theta_2 > 0$ 皆未知, 且合样本独立. 求 θ_1/θ_2 的一个置信系数为 $1 - \alpha$ 的置信区间 (提示: 令 $T_1 = X_{(m)}$, $T_2 = Y_{(n)}$, 证明 $(T_2\theta_1)/(T_1\theta_2)$ 的分布与 θ_1, θ_2 无关).
21. 设 X_1, \dots, X_n 是来自均匀分布 $U(\theta - 1/2, \theta + 1/2)$ 的简单随机样本, 求 θ 的置信系数 $1 - \alpha$ 的置信区间.
- *22. 设 r.v. X 具有密度
- $$f(x, \theta_1, \theta_2) = \begin{cases} \frac{\theta_2}{\theta_1} \left(\frac{x}{\theta_1}\right)^{\theta_2-1} e^{-(x/\theta_1)^{\theta_2}}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$
- 则称 X 服从 Weibull 分布, $\theta_1 > 0, \theta_2 > 0$ 皆未知. 设 X_1, \dots, X_n 是从此总体中抽取的简单的随机样本.
- (1) 写出估计 θ_1, θ_2 的似然方程;
 - (2) 令 $Y_i = (X_i/\theta_1)^{\theta_2}$, 证明 Y_1, \dots, Y_n i.i.d. 且 Y_1 分布与 θ_1, θ_2 无关;
 - (3) 将似然方程转化到 Y_1, \dots, Y_n , 以证明若 $\hat{\theta}_2$ 为似然方程对 θ_2 之解, 则 $\hat{\theta}_2/\theta_2$ 的分布与 θ_1, θ_2 无关;
 - (4) 指出利用 (3) 中结果构造 θ_2 的置信区间的方法.
23. 设 X_1, \dots, X_n 为从具有密度 $f(x, \theta) = e^{-(x-\theta)} I_{[x>\theta]}$ 的总体中抽取的简单随机样本, 此处 $-\infty < \theta < +\infty$, θ 未知.
- (1) 证明 $X_{(1)} - \theta$ 的分布与 θ 无关, 并求出此分布. 此处 $X_{(1)} = \min\{X_1, \dots, X_n\}$ (提示: 记 $X'_i = X_i - \theta, i = 1, 2, \dots, n$. 证明 X'_1, \dots, X'_n 的分布皆与 θ 无关, 又注意到 $X_{(1)} - \theta = X'_{(1)}$).
 - (2) 求 θ 的置信系数为 $1 - \alpha$ 的置信区间.
24. 设 X_1, \dots, X_n 是来自具有密度函数
- $$f(x, \theta) = \frac{\theta}{x^2} I_{[0 < \theta < x < \infty]}$$

的总体的简单随机样本, 求 θ 的置信系数为 $1 - \alpha$ 的置信区间.

25. 某学校计划组织一次大的郊游活动, 为此要了解学生对该活动的支持程度, 随机地对 100 名学生进行了解, 其中有 22 名支持者. 若记该校学生中支持这项活动的人数比例为 p , 并设该校学生数足够多.
- (1) 求 p 的置信系数为 0.99 的置信区间;
 - (2) 若只关心 p 的下限, 求出置信系数为 0.95 的置信下限.
26. 设 $X \sim \text{二项分布 } b(m, p_1)$, $Y \sim \text{二项分布 } b(n, p_2)$, m, n 很大, p_1, p_2 是未知参数, 且 X, Y 独立, 做出基于 X, Y 的 $p_2 - p_1$ 的置信系数近似为 $1 - \alpha$ 的置信区间.
27. 设 X_1, \dots, X_m 和 Y_1, \dots, Y_n 分别为自具有位置参数 θ_1 和 θ_2 的 Cauchy 分布总体中抽取的简单随机样本, 且样本 $X_1, \dots, X_m, Y_1, \dots, Y_n$ 相互独立, m, n 很大, 求 $\theta_2 - \theta_1$ 的置信系数近似为 $1 - \alpha$ 的置信区间. (Cauchy 分布密度 $f(x, \theta) = 1 / \{\pi[1 + (x - \theta)^2]\}$, $-\infty < x < +\infty$. 提示: 利用样本中位数及其极限定理.)
28. 电话总机在单位时间内接到呼唤次数服从 Poisson 分布 $P(\lambda)$. 观察单位时间内呼唤次数 40 次, 获得如下数据:

接到呼唤次数	0	1	2	3	4	5	6	7
观 察 次 数	5	10	12	8	3	2	0	0

试求 λ 的置信水平近似为 0.95 的置信区间.

29. 设 X_1, \dots, X_m 和 Y_1, \dots, Y_n 分别是从具有参数 λ_1 和 λ_2 的指数分布 (密度函数为 $f(x, \lambda_i) = \lambda_i e^{-\lambda_i x} I_{[x>0]}$, $i = 1, 2$) 中抽取的 i.i.d. 样本, 且合样本独立. 试确定 $\lambda_2 - \lambda_1$ 的一个信仰系数为 $1 - \alpha$ 的信仰区间 (提示: 利用 $2\lambda_1 m \bar{X} \sim \chi^2_{2m}$, $2\lambda_2 n \bar{Y} \sim \chi^2_{2n}$, 不必算到底, 指出方法即可).
30. 设 X_1, \dots, X_m i.i.d. $\sim F_1$, Y_1, \dots, Y_n i.i.d. $\sim F_2$, 且合样本独立, 仿上题的方法求
- (1) 若 F_1 和 F_2 分别为 $N(a, \sigma_1^2)$ 和 $N(b, \sigma_2^2)$, $a, b, \sigma_1^2, \sigma_2^2$ 皆未知, 求 $\sigma_2^2 - \sigma_1^2$ 的信仰系数为 $1 - \alpha$ 的信仰区间;
 - (2) 若 F_1 和 F_2 分别是表示均匀分布 $U(0, \theta_1)$ 和 $U(0, \theta_2)$, $\theta_1 > 0, \theta_2 > 0$ 皆未知, 求 $\theta_2 - \theta_1$ 的信仰系数为 $1 - \alpha$ 的信仰区间.
31. 设 X_1, \dots, X_n 为自指数分布 (密度函数为 $f(x, \lambda) = \lambda e^{-\lambda x} I_{(0, \infty)}(x)$) 中抽取的简单随机样本, 定出 λ 的信仰分布, 由之求出信仰系数为 $1 - \alpha$ 的最短信仰区间 (提示: 利用充分统计量 \bar{X}).
32. 设样本与总体同上题, 试求出形如 $c\bar{X}$ 的水平 $(1 - \beta, 1 - \gamma)$ 的容忍上、下限, 并利用引理 4.5.1 求出水平 $(1 - \beta, 1 - \gamma)$ 的容忍区间.
33. 经验表明, 某种型号玻璃纸的横向延伸率 (单位: %) 服从正态分布, 现从一批玻璃纸中随机抽取 8 个样品, 测得其横向延伸率为

$$35.5, 39.5, 43.5, 41.5, 47.5, 37.5, 49.5, 45.5.$$

试求玻璃纸的横向延伸率的水平为 $(0.95, 0.95)$ 的容忍区间.

34. 某种型号的圆钢的硬度 (单位: kg/mm^2) 服从正态分布. 现从一批圆钢中随机抽取 12 个样品, 测量其硬度, 由此算得 $\bar{X} = 207$, $S = 65$. 试求圆钢的水平为 $(0.90, 0.95)$ 的容忍下限.
35. (1) 为使 $X_{(1)}$ 成为 $F(x)$ 的水平 $(0.95, 0.99)$ 容忍下限, 样本容量 n 至少应为多少?
(2) 为使 $X_{(n)}$ 成为 $F(x)$ 的水平 $(0.90, 0.95)$ 容忍上限, 样本容量 n 至少应为多少?
36. 设 X_1, \dots, X_n i.i.d.~均匀分布 $U(0, \theta)$, $\theta > 0$ 未知, 试利用统计量 $\max_{1 \leq i \leq n} X_i$ 定出水平为 $(1 - \beta, 1 - \gamma)$ 的容忍上、下限.

第5章 参数假设检验

参数估计和假设检验是统计推断的两个主要形式. 关于参数的点估计和区间估计的问题已在前两章中讨论, 本章讨论假设检验问题. 假设检验问题就是通过从有关总体中抽取一定容量的样本, 利用样本去检验总体分布是否具有某种特征. 假设检验问题大致分为两大类.

(1) 参数型假设检验: 即总体的分布形式已知 (如正态、指数、二项分布等), 总体分布依赖于未知参数 (或参数向量) θ , 要检验的是有关未知参数的假设. 例如, 总体 $X \sim N(a, \sigma^2)$, a 未知, 检验

$$H_0: a = a_0 \leftrightarrow H_1: a \neq a_0 \quad \text{或} \quad H_0: a \leq a_0 \leftrightarrow H_1: a > a_0.$$

(2) 非参数型假设检验: 如果总体分布形式未知, 此时就需要有一种与总体分布族的具体数学形式无关的统计方法, 称为非参数方法. 例如, 检验一批数据是否来自某个已知的总体, 就属于这类问题.

本章讨论参数型假设检验问题. 非参数型假设检验问题将在第 6 章讨论.

同一个检验问题, 可以有几种不同的检验方法, 如何比较它们的优劣? 如何寻找在一定准则下的最优检验? 因此, 本章除了研究制定假设检验的种种方法, 还要讨论检验的优良性问题.

有关 Bayes 假设检验的内容放在本书的第 7 章介绍.

5.1 假设检验的若干基本概念

5.1.1 检验问题的提法

为了说明假设检验问题的提法, 考察下面的例子.

例 5.1.1 某工厂生产的一大批产品, 要卖给商店. 按规定次品率 p 不得超过 0.01, 今在其中抽取 100 件, 经检验有 3 件次品, 问这批产品可否出厂?

关于这个问题, 在面前存在两种可能性:

$$\text{甲: } 0 < p \leq 0.01; \quad \text{乙: } 0.01 < p < 1.$$

要通过从这批产品中抽样来决定甲, 乙两种可能性中哪个成立.

由于种种原因, 这个问题常以下述方式提出: 引进一个“假设”

$$H_0 : 0 < p \leq 0.01,$$

它称为零假设 (null hypothesis) 或原假设, 有时也简称为假设. 另一个可能是

$$H_1 : 0.01 < p < 1,$$

称为对立假设或备择假设 (alternative hypothesis).

目的是要通过样本决定接受 H_0 , 还是拒绝 H_0 . 可以形象地把问题写成

$$H_0 : 0 < p \leq 0.01 \leftrightarrow H_1 : 0.01 < p < 1.$$

注意这个提法中将 H_0 放在中心位置, 它是检验的对象. H_0 和 H_1 的位置不可颠倒. 从这个例子可将假设检验问题一般化, 提法如下.

设有参数分布族 $\{f(x, \theta), \theta \in \Theta\}$, 此处 Θ 为参数空间. X_1, \dots, X_n 是从上述分布族中抽取的简单随机样本. 在参数假设检验问题中, 感兴趣的是 θ 是否属于参数空间 Θ 的某个非空真子集 Θ_0 , 则命题 $H_0 : \theta \in \Theta_0$ 称为零假设或原假设, 其确切含义是: 存在一个 $\theta_0 \in \Theta_0$ 使得 X 的分布为 $f(x, \theta_0)$. 记 $\Theta_1 = \Theta - \Theta_0$, 则命题 $H_1 : \theta \in \Theta_1$ 称为 H_0 的对立假设或备择假设 (在例 5.1.1 中, $\Theta = (0, 1)$, $\Theta_0 = (0, 0.01]$, $\Theta_1 = (0.01, 1)$), 则假设检验问题表示为

$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1, \quad (5.1.1)$$

但要注意一点, 零假设总写在左边, 作为中心位置; 对立假设写在右边, 作为陪衬地位. 若没有对立假设, 检验问题就无完整的意义. 在以后的讨论中, 可以看出 H_0 放在中心地位的意义何在.

在式 (5.1.1) 中, 若 Θ_0 或 Θ_1 只包含参数空间 Θ 中的一个点, 则称为简单假设 (simple hypothesis); 否则, 称为复合假设 (composite hypothesis). 例如, 样本抽自 $N(a, \sigma^2)$, σ^2 已知, 则参数空间为 $\Theta = \{a : -\infty < a < +\infty\}$. 令 $H_0 : a = a_0 \leftrightarrow H_1 : a \neq a_0$, 则 H_0 为简单假设, H_1 为复合假设. 再如, 在上述问题中, 令 $H'_0 : a \leq a_0 \leftrightarrow H'_1 : a > a_0$, 则零假设 H'_0 和对立假设 H'_1 皆为复合假设.

5.1.2 否定域、检验函数和检验统计量

仍通过例子来说明这些概念.

例 5.1.2 设 $X = (X_1, \dots, X_n)$ 为从正态总体 $N(a, \sigma^2)$ 中抽取的随机样本, 其中 σ^2 已知. 考虑检验问题

$$H_0 : a = a_0 \leftrightarrow H_1 : a \neq a_0, \quad (5.1.2)$$

此处 a_0 为给定的常数.

这种检验的一种直观作法是：先求 a 的一个估计量，可以知道 $\bar{X} = \sum_{i=1}^n X_i / n$ 是 a 的一个优良估计。若 $|\bar{X} - a_0|$ 较大，就倾向于否定 H_0 ；反之，如果 $|\bar{X} - a_0|$ 较小，就认为抽样结果与 H_0 相接近，因而倾向于接受 H_0 。具体地说，要确定一个数 A ，由 X_1, \dots, X_n 算出样本均值 \bar{X} ，当 $|\bar{X} - a_0| > A$ 时就否定 H_0 ；当 $|\bar{X} - a_0| \leq A$ 时就接受 H_0 。称

$$D = \{\mathbf{X} = (X_1, \dots, X_n) : |\bar{X} - a_0| > A\} \quad (5.1.3)$$

为否定域，或拒绝域 (reject region)，即否定域是由样本空间 \mathcal{X} 中一切使 $|\bar{X} - a_0| > A$ 的那些样本 $\mathbf{X} = (X_1, \dots, X_n)$ 构成。有了否定域，等价于将样本空间 \mathcal{X} 分成不相交的两部分 D 和 $\bar{D} = \mathcal{X} - D$ 。一旦有了样本 \mathbf{X} ，当 $\mathbf{X} \in D$ 时，就否定 H_0 ；当 $\mathbf{X} \in \bar{D}$ 时，就接受 H_0 。称 \bar{D} 为接受域 (acceptance region)。只要 A 定下来了，则否定域（或接受域）也就确定了。因此，此问题中的检验可视为如下一种法则：

$$T : \begin{cases} \text{当 } |\bar{X} - a_0| > A \text{ 时, 拒绝 } H_0, \\ \text{当 } |\bar{X} - a_0| \leq A \text{ 时, 接受 } H_0. \end{cases}$$

上式中的 T 给出了一种法则，一旦有了样本，就可以在接受 H_0 或否定 H_0 这两个结论中选择一个。称这样一种法则 T 为检验问题 (5.1.2) 的一个检验。

为了便于数学上处理，引入检验函数 $\varphi(\mathbf{x})$ 的概念， $\varphi(\mathbf{x})$ 与检验 T 是一一对应的。在例 5.1.2 中，

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{当 } |\bar{x} - a_0| > A, \\ 0, & \text{当 } |\bar{x} - a_0| \leq A. \end{cases} \quad (5.1.4)$$

有如下定义。

定义 5.1.1 由式 (5.1.4) 给出的检验函数 $\varphi(\mathbf{x})$ 是定义在样本空间 \mathcal{X} 上，取值于 $[0,1]$ 上的函数。它表示当有了样本 \mathbf{X} 后，否定 H_0 的概率。

由定义可见，若 $\varphi(\mathbf{x}) = 1$ ，则以概率为 1 否定 H_0 ，若 $\varphi(\mathbf{x}) = 0$ ，则以概率为 0 否定 H_0 （即以概率为 1 接受 H_0 ）。若 $\varphi(\mathbf{x})$ 只取 0, 1 这两个值，则称这种检验为非随机化检验 (non-randomized test)。此时，否定域也可用检验函数表示如下： $D = \{\mathbf{X} = (X_1, \dots, X_n) : \varphi(\mathbf{X}) = 1\}$ 。

若对某些样本 \mathbf{X} ，有 $0 < \varphi(\mathbf{x}) < 1$ ，则称 $\varphi(\mathbf{x})$ 为随机化检验 (randomized test)。如在例 5.1.1 中，令 $X_i = 1$ ，若第 i 个产品为次品；否则为 0； $i = 1, \dots, n$ 。设 $\mathbf{X} = (X_1, \dots, X_n)$ 为样本，令 $T(\mathbf{X}) = \sum_{i=1}^{100} X_i$ ，当 $T(\mathbf{X}) < c$ 时认为这批产品合格，

接受 H_0 ; 当 $T(\mathbf{X}) > c$ 时, 认为不合格, 拒绝 H_0 . 当 $T(\mathbf{X}) = c$ 时, 若规定拒绝 H_0 , 厂方觉得被拒绝的可能性大了, 吃亏了. 反之, 若接受 H_0 , 买方(商店)接受不合格产品的可能性大了, 也觉得吃亏. 在双方僵持不下的情况下, 下列折中方案是双方都可以接受的: 定下一个数 $0 < r < 1$, 规定当 $T(\mathbf{X}) = c$ 时, 作一次成功概率为 r 的随机试验, 根据试验结果来决定拒绝还是接受这批产品, 如取 $r = 1/2$, 则可通过掷一枚均匀硬币来决定. 规定若出现正面则拒绝 H_0 , 否则接受 H_0 . 这样, 当出现 $T(\mathbf{X}) = c$, 双方都有 $1/2$ 的可能, 作出对自己不利的决定, 双方都觉得合理, 可以接受. 如果取 $r = 1/3$, 试验可以通过在装有 2 个白球和 1 个黑球的盒子中随机摸球来决定. 规定若摸到黑球(发生的概率为 $1/3$) 则拒绝 H_0 , 若摸到白球(发生的概率为 $2/3$) 则接受 H_0 . 这种随机化检验函数可表示为

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > c, \\ r, & T(\mathbf{x}) = c, \\ 0, & T(\mathbf{x}) < c. \end{cases} \quad (5.1.5)$$

在例 5.1.2 中要确定检验, 必须定出式 (5.1.3) 或式 (5.1.4) 中的 A , 此处 A 称为临界值(critical value). 要定下 A 的值需要确定检验统计量的分布. 此例中检验统计量是样本均值 \bar{X} . 同样, 在例 5.1.1 中, 检验函数 (5.1.5) 中的 c 称为临界值, 检验统计量是 $T(\mathbf{X}) = \sum_{i=1}^{100} X_i$. 确定检验统计量的分布是解决假设检验问题的关键. 当检验统计量的精确分布很难找到时, 若其极限分布比较简单, 可用极限分布代替精确分布, 获得假设检验问题的近似解.

5.1.3 两类错误与功效函数

统计推断是以样本为依据的, 由于样本的随机性, 不能保证统计推断方法的绝对正确性, 而只能以一定的概率去保证这种推断的可靠性. 在假设检验问题中可能出现下列两种情形会犯错误:

(1) 零假设 H_0 本来是对的, 由于样本的随机性, 观察值落入否定域 D , 错误地将 H_0 否定了, 称为弃真. 这时犯的错误称为第一类错误(type I error).

(2) 零假设 H_0 本来不对, 由于样本的随机性, 观察值落入接受域 \bar{D} , 错误地将 H_0 接受了, 称为取伪. 这时犯的错误称为第二类错误(type II error).

如在例 5.1.1 中确定了非随机化检验如下:

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > 3, \\ 0, & T(\mathbf{x}) \leq 3. \end{cases}$$

如果总体的真实次品率为 $p = 0.005 < 0.01$, 由于样本的随机性, 抽样结果显示

$T(\mathbf{x}) = 5$, 即样本落入否定域, 这时犯第一类错误. 但也有可能总体的真实次品率为 $p = 0.03 > 0.01$, 由于样本的随机性, 抽样结果显示 $T(\mathbf{x}) = 1$, 即样本落入了接受域. 这时犯第二类错误.

应当注意, 在每一具体场合, 只会犯两类错误中的一个. 当检验确定后, 犯两类错误的概率也就确定了. 希望犯两类错误的概率越小越好, 但这一点很难做到. 在样本大小 n 固定的前提下, 二者不可兼得. 这就如同区间估计问题中可靠度和精度二者不可兼得一样. 那么, 怎样去计算犯两类错误的概率呢? 为此, 需要引进功效函数的概念.

定义 5.1.2 设 $\varphi(\mathbf{x})$ 是 $H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1$ 的一个检验函数, 则

$$\beta_\varphi(\theta) = P_\theta \left\{ \text{用检验 } \varphi \text{ 否定了 } H_0 \right\} = E_\theta[\varphi(\mathbf{X})], \quad \theta \in \Theta$$

称为 φ 的功效函数 (power function), 也称为效函数或势函数.

若 $\varphi(\mathbf{x})$ 为非随机化检验, 否定域为 D , 则

$$\beta_\varphi(\theta) = P_\theta \left\{ \mathbf{X} = (X_1, \dots, X_n) \in D \right\}, \quad \theta \in \Theta.$$

因此功效函数表示当参数为 θ 时, 否定 H_0 的概率. 对例 5.1.1, 当检验函数为式 (5.1.5) 时, 利用 $T(\mathbf{X}) = \sum_{i=1}^n X_i \sim b(n, \theta)$, $0 < \theta < 1$ 可知检验的功效函数为

$$\begin{aligned} \beta_\varphi(\theta) &= E_\theta[\varphi(\mathbf{X})] = P \left(\sum_{i=1}^{100} X_i > c \right) + rP \left(\sum_{i=1}^{100} X_i = c \right) \\ &= \sum_{k=c+1}^{100} \binom{100}{k} \theta^k (1-\theta)^{100-k} + r \binom{100}{c} \theta^c (1-\theta)^{100-c}. \end{aligned}$$

以下讨论中假定 $\varphi(\mathbf{x})$ 皆为非随机化的检验函数, 除非特别申明, 不认为 $\varphi(\mathbf{x})$ 为随机化检验函数.

知道了检验 $\varphi(\mathbf{x})$ 的功效函数后, 就可以计算犯两类错误的概率. 若以 $\alpha_\varphi^*(\theta)$ 和 $\beta_\varphi^*(\theta)$ 分别记犯第一、二类错误的概率, 则犯第一类错误的概率可表示为

$$\alpha_\varphi^*(\theta) = \begin{cases} \beta_\varphi(\theta), & \theta \in \Theta_0, \\ 0, & \theta \in \Theta_1, \end{cases}$$

犯第二类错误的概率可表示为

$$\beta_\varphi^*(\theta) = \begin{cases} 0, & \theta \in \Theta_0, \\ 1 - \beta_\varphi(\theta), & \theta \in \Theta_1. \end{cases}$$

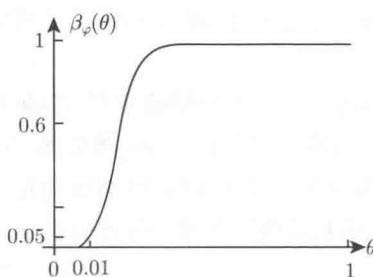


图 5.1.1 功效函数

如在例 5.1.1 中, $\Theta = (0, 1)$, 则检验问题 $H_0 : 0 < \theta \leq 0.01 \leftrightarrow H_1 : 0.01 < \theta < 1$ 的一个较好的功效函数之形状如图 5.1.1 所示.

一个好的检验 $\varphi(x)$, 犯两类错误的概率都应较小, 也就是功效函数 $\beta_\varphi(\theta)$ 在 Θ_0 中应尽可能地小, 在 Θ_1 中应尽可能地大.

还需要说明的一点是: 如前所述, 犯两类错误的概率完全由功效函数决定, 从这一点上看, 如果两个检验有同一功效函数, 则此两检验在性质上也完全相同.

5.1.4 检验水平和控制犯第一类错误概率的原则

前面说过, 希望一个检验犯两类错误的概率都很小, 但除了极例外情形, 一般说来在固定样本大小时, 对任何检验都办不到. 例如, 要使犯第一类错误的概率减小, 就要缩小拒绝域, 使接受域增大, 这必然导致犯第二类错误概率增大, 反之亦然. 因此, Neyman-Pearson 提出了一条原则, 就是限制犯第一类错误概率的原则, 即在保证犯第一类错误的概率不超过指定数值 α ($0 < \alpha < 1$, 通常取较小的数) 的检验中, 寻找犯第二类错误概率尽可能小的检验. 若记

$$S_\alpha = \{ \varphi : \beta_\varphi(\theta) \leq \alpha, \theta \in \Theta_0 \},$$

S_α 表示由所有犯第一类错误的概率都不超过 α 的检验函数构成的类. 只考虑 S_α 中的检验, 在 S_α 中挑选“犯第二类错误的概率尽可能小的检验”, 这种法则称为控制犯第一类错误概率的法则.

根据 Neyman-Pearson 原则, 在原假设 H_0 为真时, 作出错误决定 (即否定 H_0) 的概率受到了控制. 这表明, 原假设 H_0 受到保护, 不至于轻易被否定. 所以在具体问题中, 往往将有把握、不能轻易否定的命题作为原假设 H_0 , 而把没有把握的、不能轻易肯定的命题作为对立假设. 因此原假设 H_0 和对立假设 H_1 的地位是不平等的, 不能相互调换.

与犯第一类错误概率相联系的另一个概念是检验水平, 其定义如下.

定义 5.1.3 设 φ 是式 (5.1.1) 的一个检验, 而 $0 \leq \alpha \leq 1$. 如果 φ 犯第一类错误的概率总不超过 α (或等价地说, φ 满足 $\beta_\varphi(\theta) \leq \alpha$, 一切 $\theta \in \Theta_0$), 则称 α 是检验 φ 的一个水平, 而 φ 称为显著性水平为 α 的检验, 简称水平为 α 的检验.

按这一定义, 检验的水平不唯一. 若 α 为检验 φ 的水平, 而 $\alpha < \alpha' < 1$, 则 α' 也是检验 φ 的水平. 为避免这一问题, 有时称一个检验的最小水平为其真实水平. 也就是

$$\text{检验 } \varphi \text{ 的真实水平} = \sup\{ \beta_\varphi(\theta), \theta \in \Theta_0 \}. \quad (5.1.6)$$

至于水平的选择, 习惯上把 α 取得比较小且标准化, 如 $\alpha = 0.01, 0.05$ 或 0.10 等. 标准化是为了方便造表.

水平的选取, 对检验的性质有很大影响. 不难了解, 如果水平选得很低, 那么容许犯第一类错误的概率很小, 而为了达到这一点势必大大缩小否定域, 使接受域扩大, 从而增加了犯第二类错误的可能性. 反之, 若水平选得高, 则否定域扩大, 使接受域缩小, 从而犯第二类错误的概率将相应地降低. 这样看来, 水平的选择不是一个数学问题, 而是一个必须从实际角度来考虑的问题. 一般说来, 有以下几个因素影响水平的选定:

(1) 当一个检验涉及双方利益时, 水平的选定常是双方协议的结果. 以例 5.1.1 为例, 商店向工厂进货, 检验其次品率是否超过 0.01, 若水平选得低, 则可能有较多的次品被商店接受; 反之, 若水平定得高, 则将有较多的合格品被商店拒收. 因此水平定的大小涉及商店和工厂双方利益, 应由双方商定. 如前所述, 有时还要采取随机化的方法, 使双方利益达到平衡.

(2) 犯两种错误的后果一般在性质上有很大的不同. 如果犯第一类错误的后果在性质上很严重, 就力求在合理的范围内尽量减少犯这种错误的可能性, 这时相应的水平就取得更低一些. 例如, 制药厂要生产一种新药代替旧药治疗某种疾病, 安排了一些试验, 要对新旧药物疗效作出检验. 由于旧药已经长期临床使用, 有一定的疗效. 新药尚未经长期临床使用, 一旦效果不好时, 将危及病人的生命安全, 造成严重的后果. 所以在进行检验时, 将原假设 H_0 设为“旧药不比新药差”, 且使检验水平 α 定得更小一些, 这样使 H_0 被否定的可能性大大减小了. 这样就保证了“原假设被否定、新药被接受的检验”将是非常严格的.

(3) 一般说来, 试验者在试验前对问题的情况总不是一无所知的. 他对问题的了解使他对零假设是否能成立就有了一定的看法, 这种看法可能影响到他对水平的选择. 比方说, 一个物理学家根据某种理论推定随机变量 X 应有分布 F , 而他打算将这一理论付诸检验. 很明显, 如果他对这一理论很有信心, 他将非常倾向于认为假设能成立, 这时只有很有力的证据才可能使他认为这假设不对. 相应地, 他将把检验水平取得低一些.

在实际问题中, 零假设被否定, 常常意味着推翻一种理论或用新方法来代替一直使用的标准方法. 在大多数情况下, 人们希望这样做时有相当大的根据. 从这里可以看到, Neyman-Pearson 控制犯第一类错误的原则, 在零假设的选择中有很大的实际意义, 而决不单纯是一个数学问题. 同时, 也进一步解释了在假设检验问题中, 零假设处在突出地位的原因.

最后要说明的一点是: 若水平 α 很小, 原假设 H_0 不会轻易被否定. 如果样本落入了否定域, 作出“否定原假设 H_0 ”的结论就比较可靠(因为, 此时只会犯第一类错误, 且其概率很小). 反之, 当 α 很小时, 如果样本落入接受域, 作出“接受原假

设 H_0 " 的结论未必可靠. 这只能表明: 在所选定的水平下没有充分根据否定 H_0 , 但决不意味着有充分根据说明它正确 (因为此时会犯第二类错误, 其概率可能很大).

5.1.5 求解假设检验问题的一般步骤

- (1) 根据问题的要求提出零假设 H_0 和备择假设 H_1 ;
- (2) 导出否定域的形式, 确定检验统计 $T(\mathbf{X})$, 其中临界值 A 待定.
- (3) 选取适当水平, 利用检验统计量的分布求出临界值 A .
- (4) 由样本 \mathbf{X} 算出检验统计量 $T(\mathbf{X})$ 的具体值, 代入到否定域中, 与临界值相比较, 作出接受或者拒绝原假设 H_0 的结论.

5.1.6 假设检验发展简史

假设检验是一种有重要应用价值的统计推断形式. 从理论上看, 它是数理统计学的一个重要分支. 除去早期的片段情况, 假设检验方法和理论的系统发展, 始于 20 世纪初. 大致上按照时间顺序, 其过程经历了以下一些重大事件: K. Pearson 在 1900 年提出 χ^2 检验 (也称为拟合优度检验), 这个检验方法首先是在分类数据的检验问题中提出来的. R. A. Fisher 在 20 世纪 20 年代发表了几篇论文, 创立了显著性检验方法. 他建立了实用中常用的基于 t 分布和 F 分布的检验. Pearson 和 Fisher 工作的不足之处, 是没有提出如何制定一些原则和标准, 对不同的检验方法进行比较和选择. 而 J. Neyman 和 E. S. Pearson 正是从这一点入手, 从 1928 年开始, 花了大约 10 年时间, 合作发表了一系列论文, 建立了假设检验的一种数学理论, 通称为 Neyman-Pearson (NP) 理论. 这是假设检验乃至数理统计学中的一个重大事件. NP 理论重要性在于给出了一个样板, 将统计问题的解转化为数学最优化问题. 十余年后, A. Wald 把这一想法扩展到整个的数理统计学领域, 在 1950 建立了统计决策理论, 对战后时期数理统计的发展起了较大的影响. 假设检验发展史上另一个重要事件是 Bayes 学派的兴起. Bayes 方法的研究大约始于 20 世纪二三十年代, 到五六十年代引起人们广泛的关注, 迅速崛起, 达到可以与频率学派分庭抗礼的程度. 时至今日, 其影响日益扩大, 它的研究已渗透到数理统计的几乎所有领域.

作为刚刚接触假设检验这个统计推断形式的读者, 很有必要了解一下 Pearson 和 Fisher 这些大师的思想, 通过对他们的研究工作的了解, 可以加深对假设检验中一些基本概念的理解. 有兴趣的读者可参看文献 [2] 第 96 页和文献 [5] 第 9 章.

5.2 正态总体参数的假设检验

正态分布是最常见的分布, 关于它的参数的假设检验是实际中常遇到的问题,

因此也是最重要的一类检验问题. 本节将分下列几种情况来讨论正态总体参数的直观检验方法: 单个正态总体均值和方差的检验; 两个正态总体均值差和方差比的检验; 极限分布为正态分布的有关大样本检验.

在讨论正态分布总体参数的假设检验问题时, 定理 2.2.3 和推论 2.4.1 – 推论 2.4.5 在求检验统计量的分布中起到十分重要的作用.

5.2.1 单个正态总体均值的检验

设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从正态总体 $N(\mu, \sigma^2)$ 中抽取的简单随机样本, 求下列三类检验问题:

- (1) $H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0,$
- (2) $H'_0 : \mu \leq \mu_0 \leftrightarrow H'_1 : \mu > \mu_0,$
- (3) $H''_0 : \mu \geq \mu_0 \leftrightarrow H''_1 : \mu < \mu_0,$

其中 μ_0 和检验水平 α 给定.

检验问题 (1) 称为双边检验 (two-side test), 检验问题 (2) 和 (3) 称为单边检验 (one-side test).

1. 方差 σ^2 已知时的检验方法

- (i) 首先考虑检验问题 (1), 即

$$H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0.$$

用直观方法构造检验的否定域. 由于 $\bar{X} = \sum_{i=1}^n X_i/n$ 是 μ 的无偏估计且具有良好性质. 直观上看 $|\bar{X} - \mu_0|$ 越大, H_0 越不像成立. 因此检验的否定域可取如下形式: $\{\mathbf{X} = (X_1, \dots, X_n) : |\bar{X} - \mu_0| > A\}$, A 待定. 当 σ^2 已知时, 由定理 2.2.3 可知, 在 H_0 成立的条件下 $\bar{X} \sim N(\mu_0, \sigma^2/n)$, 将 \bar{X} 标准化得到

$$U = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0, 1). \quad (5.2.1)$$

因此取 $U = \sqrt{n}(\bar{X} - \mu_0)/\sigma$ 作为检验统计量, 则否定域的等价形式可取为

$$\left\{ (X_1, \dots, X_n) : |U| > c \right\}, \quad c \text{ 待定.}$$

由检验水平为 α , 可知

$$P(|U| > c | H_0) = P\left(\left|\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}\right| > c | H_0\right) = \alpha.$$

由于 $U | \mu = \mu_0 \sim N(0, 1)$, 故得 $c = u_{\alpha/2}$. 因此由否定域

$$D_1 = \left\{ (X_1, \dots, X_n) : |U| > u_{\alpha/2} \right\} \quad (5.2.2)$$

确定的检验为检验问题(1)的水平为 α 的检验.

(ii) 当方差 σ^2 已知时, 考虑检验问题(2)的水平为 α 的检验, 即

$$H'_0: \mu \leq \mu_0 \leftrightarrow H'_1: \mu > \mu_0.$$

从直观上看检验问题(2)的否定域为 $\{(X_1, \dots, X_n): \sqrt{n}(\bar{X} - \mu_0)/\sigma > c\}$, 其中 c 待定. 此时仍取 $U = \sqrt{n}(\bar{X} - \mu_0)/\sigma$ 作为检验统计量. 令

$$P(U > c | H'_0) = P\left(\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} > c \mid \mu = \mu_0\right) = \alpha, \quad (5.2.3)$$

由于 $U | \mu = \mu_0 \sim N(0, 1)$, 故取 $c = u_\alpha$. 得到否定域

$$D_2^* = \left\{ (X_1, \dots, X_n) : U = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} > u_\alpha \right\}. \quad (5.2.4)$$

由上式给出的否定域是否为检验问题(2)的检验水平为 α 的否定域呢? 显然由(5.2.3)可知, 当 $\mu = \mu_0$ 时由否定域 D_2^* 确定的检验的水平是 α . 问题是当 $\mu < \mu_0$ 时, 由否定域 D_2^* 确定的检验的水平是否仍然是 α ? 如果能证明: 以式(5.2.4)为否定域的检验的功效函数 $\beta_{D_2^*}(\mu)$ 为 μ 的非降函数, 则它必为检验问题(2)的检验水平为 α 的检验. 其理由如下: 因为由式(5.2.3)可知 $\beta_{D_2^*}(\mu_0) = \alpha$, 又由功效函数 $\beta_{D_2^*}(\mu)$ 关于 μ 的非降性可知 $\beta_{D_2^*}(\mu) \leq \beta_{D_2^*}(\mu_0) = \alpha$, 对一切 $\mu \leq \mu_0$. 这就保证了由式(5.2.4)给出的否定域也是检验问题(2)的水平为 α 的否定域.

下面证明功效函数 $\beta_{D_2^*}(\mu)$ 关于 μ 的非降性. 记 $\Phi(\cdot)$ 为 $N(0, 1)$ 的分布函数, 则

$$\begin{aligned} \beta_{D_2^*}(\mu) &= P_\mu\left(\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} > u_\alpha\right) \\ &= P_\mu\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} > u_\alpha + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma}\right) \\ &= 1 - \Phi\left(u_\alpha + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma}\right). \end{aligned} \quad (5.2.5)$$

显然由(5.2.5)确定的功效函数 $\beta_{D_2^*}(\mu)$ 是 μ 的非降函数, 这就证明了所要的结论. 因此由

$$D_2 = D_2^* = \left\{ (X_1, \dots, X_n) : U > u_\alpha \right\} \quad (5.2.6)$$

确定的检验是检验问题(2)的水平为 α 的检验. 此处 U 由式(5.2.1)给出.

(iii) 类似方法可得检验问题(3)的检验水平为 α 的检验的否定域

$$D_3 = \left\{ (X_1, \dots, X_n) : U < -u_\alpha \right\}.$$

这种基于检验统计量 U 服从 $N(0, 1)$ 分布的检验方法称为一样本 U 检验.

2. 方差未知时正态总体均值的检验方法

在样本分布为 $N(\mu, \sigma^2)$, 当 σ^2 未知且假定 $\mu = \mu_0$ 时, 由推论 2.4.2 可知

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim t_{n-1}. \quad (5.2.7)$$

因此取 $T = \sqrt{n}(\bar{X} - \mu_0)/S$ 作为检验统计量, 用完全与方差 σ^2 已知情形相同的方法 (所不同的就是用检验统计量 T 代替那儿的检验统计量 U) 可得检验问题 (1)-(3) 的检验水平为 α 的检验的否定域. 否定域中的临界值将 $N(0, 1)$ 的分位数改成相应的 t 分布的分位数. 详细结果见表 5.2.1.

表 5.2.1 单个正态总体均值的假设检验

	H_0	H_1	检验统计量及其分布	否定域
σ^2 已知	$\mu = \mu_0$	$\mu \neq \mu_0$	$U = \sqrt{n}(\bar{X} - \mu_0)/\sigma$ $U \mu = \mu_0 \sim N(0, 1)$	$ U > u_{\alpha/2}$
	$\mu \leq \mu_0$	$\mu > \mu_0$		$U > u_\alpha$
	$\mu \geq \mu_0$	$\mu < \mu_0$		$U < -u_\alpha$
σ^2 未知	$\mu = \mu_0$	$\mu \neq \mu_0$	$T = \sqrt{n}(\bar{X} - \mu_0)/S$ $T \mu = \mu_0 \sim t_{n-1}$	$ T > t_{n-1}(\alpha/2)$
	$\mu \leq \mu_0$	$\mu > \mu_0$		$T > t_{n-1}(\alpha)$
	$\mu \geq \mu_0$	$\mu < \mu_0$		$T < -t_{n-1}(\alpha)$

这种基于检验统计量 T 服从 t 分布的检验方法称为一样本 t 检验, T 称为一样本 t 检验统计量. 它是英国统计学家 W. S. Gosset 首先发现, 并于 1908 年用“Student”笔名在 Biometrika 杂志上发表的. 故 t 分布也常称为 Student 分布. 这个结果的发现在数理统计发展史上是一件大事, 因为它开了小样本理论发展的先声.

例 5.2.1 食品厂用自动装罐机装罐头食品, 每罐标准重量为 500g, 每天开工需检查机器的工作状况. 今抽得 10 罐, 测得其重量 (单位: g) 为

$$495, 510, 505, 498, 503, 492, 502, 512, 497, 506.$$

假定罐头重量 X 服从正态分布 $N(\mu, \sigma^2)$, 已知 $\sigma = 6.5$, 问机器是否工作正常 (取检验水平 $\alpha = 0.05$)?

解 检验问题为

$$H_0 : \mu = 500 \leftrightarrow H_1 : \mu \neq 500.$$

本题中 σ^2 已知, 故否定域由表 5.2.1 中第一行给出, 即

$$D = \left\{ (X_1, \dots, X_n) : \left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \right| > u_{\alpha/2} \right\},$$

其中 $n = 10$, $\alpha = 0.05$, $\sigma = 6.5$, $\mu_0 = 500$, 查表得 $u_{0.025} = 1.96$, 由样本算得

$\bar{X} = 502$, 因此

$$|U| = \left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \right| = \left| \frac{\sqrt{10}(502 - 500)}{6.5} \right| = 0.973 < 1.96,$$

所以在显著性水平 $\alpha = 0.05$ 下认为没有足够理由说明自动装罐机工作不正常, 故接受 H_0 .

例 5.2.2 某砖厂所生产的地砖的抗断强度 X 服从正态分布 $N(\mu, \sigma^2)$, 今从该厂生产的地砖中随机抽取 6 块测得抗断强度 (单位: kg/cm^2) 如下:

$$32.56, 29.66, 31.64, 30.00, 31.87, 31.03.$$

问这一批地砖的平均抗断强度可否认为不低于 $32.50 \text{ kg}/\text{cm}^2$ (取检验水平 $\alpha = 0.05$)?

解 检验问题为

$$H_0: \mu \geq 32.50 \leftrightarrow H_1: \mu < 32.50.$$

本题中 σ^2 未知, 故采用 t 检验法, 否定域由表 5.2.1 中最后一行给出, 即

$$D = \left\{ (X_1, \dots, X_n) : \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} < -t_{n-1}(\alpha) \right\},$$

其中 $n = 6$, $\alpha = 0.05$, $\mu_0 = 32.50$, 查表得 $t_5(0.05) = 2.015$, 由数据算得 $\bar{X} = 31.13$, $S = 1.123$, 因此有

$$T = \frac{\sqrt{6}(31.13 - 32.50)}{1.123} = -2.99 < -2.015,$$

故否定 H_0 , 即认为地砖强度达不到 $32.50 \text{ kg}/\text{cm}^2$.

5.2.2 单个正态总体方差的检验

设 X_1, \dots, X_n 为自正态总体 $N(\mu, \sigma^2)$ 中抽取的简单随机样本, 本段要讨论下列三类检验问题:

- (4) $H_0: \sigma^2 = \sigma_0^2 \leftrightarrow H_1: \sigma^2 \neq \sigma_0^2$,
- (5) $H'_0: \sigma^2 \leq \sigma_0^2 \leftrightarrow H'_1: \sigma^2 > \sigma_0^2$,
- (6) $H''_0: \sigma^2 \geq \sigma_0^2 \leftrightarrow H''_1: \sigma^2 < \sigma_0^2$,

其中 σ_0^2 和检验水平 α 给定.

1. 均值未知时单个正态总体方差的检验方法

首先考虑检验问题 (4), 即

$$H_0: \sigma^2 = \sigma_0^2 \leftrightarrow H_1: \sigma^2 \neq \sigma_0^2.$$

由于均值 μ 未知时 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$ 是 σ^2 的一个无偏估计, 且具有良好性质. 直观上看 S^2/σ_0^2 太小或者 S^2/σ_0^2 太大时, H_0 不像成立. 因此检验的否定域可取如下形式: $\{(X_1, \dots, X_n) : S^2/\sigma_0^2 < A_1 \text{ 或 } S^2/\sigma_0^2 > A_2\}$, A_1, A_2 待定. 在给定 $\sigma^2 = \sigma_0^2$ 的条件下, 由定理 2.2.3 可知

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2. \quad (5.2.8)$$

故取检验统计量为 $\chi^2 = (n-1)S^2/\sigma_0^2$. 因此, 否定域的等价形式可取为

$$D = \left\{ (X_1, \dots, X_n) : \frac{(n-1)S^2}{\sigma_0^2} < c_1 \text{ 或 } \frac{(n-1)S^2}{\sigma_0^2} > c_2 \right\}.$$

记 $\theta = (\mu, \sigma^2)$, 为了确定 c_1, c_2 , 令

$$\alpha = P_\theta \left(\frac{(n-1)S^2}{\sigma_0^2} < c_1 \text{ 或 } \frac{(n-1)S^2}{\sigma_0^2} > c_2 \mid H_0 \right),$$

满足上式的 c_1, c_2 的对子有很多, 存在一对 c_1, c_2 是最优的, 但计算较复杂且使用不方便. 确定 c_1, c_2 的一个简单实用的方法是令

$$P_\theta \left(\frac{(n-1)S^2}{\sigma_0^2} < c_1 \mid H_0 \right) = \alpha/2, \quad P_\theta \left(\frac{(n-1)S^2}{\sigma_0^2} > c_2 \mid H_0 \right) = \alpha/2.$$

由上述两式和式 (5.2.8) 易知临界值 $c_1 = \chi_{n-1}^2(1-\alpha/2)$, $c_2 = \chi_{n-1}^2(\alpha/2)$. 所以检验问题 (4) 的水平为 α 的接受域为

$$\bar{D}_4 = \left\{ (X_1, \dots, X_n) : \chi_{n-1}^2 \left(1 - \frac{\alpha}{2} \right) \leq \frac{(n-1)S^2}{\sigma_0^2} \leq \chi_{n-1}^2 \left(\frac{\alpha}{2} \right) \right\}.$$

此接受域的表达式比否定域简单, 使用上也方便, 故此处采用接受域代替否定域.

用完全类似于 5.2.1 节中求检验问题 (2) 和 (3) 的方法可分别求得检验问题 (5) 和 (6) 的水平为 α 的否定域如下:

$$D_5 = \left\{ (X_1, \dots, X_n) : \frac{(n-1)S^2}{\sigma_0^2} > \chi_{n-1}^2(\alpha) \right\},$$

$$D_6 = \left\{ (X_1, \dots, X_n) : \frac{(n-1)S^2}{\sigma_0^2} < \chi_{n-1}^2(1-\alpha) \right\}.$$

2. 当均值 μ 已知时方差 σ^2 的检验方法

简述如下: 当 μ 已知时, σ^2 的一个具有良好性质的无偏估计是 $S_\mu^2 = \sum_{i=1}^n (X_i - \mu)^2/n$. 当 $\sigma^2 = \sigma_0^2$ 时, 由推论 2.4.1 可知

$$\frac{nS_\mu^2}{\sigma_0^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma_0} \right)^2 \sim \chi_n^2. \quad (5.2.9)$$

因此取检验统计量为 $\chi^2_\mu = nS_\mu^2/\sigma_0^2$, 用它代替 $\chi^2 = (n-1)S^2/\sigma_0^2$. 采用完全类似于前面 μ 未知情形的讨论方法, 可得检验问题 (4)-(6) 的水平为 α 的否定域, 仅注意在否定域中将确定临界值的 χ^2 分布的自由度由 $n-1$ 改成 n 即可. 详细结果见表 5.2.2.

表 5.2.2 单个正态总体方差的假设检验

	H_0	H_1	检验统计量 及其分布	否定域
μ 已 知	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2_\mu = nS_\mu^2/\sigma_0^2$ $\chi^2_\mu \sigma_0^2 \sim \chi_n^2$	$nS_\mu^2/\sigma_0^2 < \chi_n^2(1-\alpha/2)$ 或 $nS_\mu^2/\sigma_0^2 > \chi_n^2(\alpha/2)$
	$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$		$nS_\mu^2/\sigma_0^2 > \chi_n^2(\alpha)$
	$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$		$nS_\mu^2/\sigma_0^2 < \chi_n^2(1-\alpha)$
μ 未 知	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ $\chi^2 \sigma_0^2 \sim \chi_{n-1}^2$	$(n-1)S^2/\sigma_0^2 < \chi_{n-1}^2(1-\alpha/2)$ 或 $(n-1)S^2/\sigma_0^2 > \chi_{n-1}^2(\alpha/2)$
	$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$		$(n-1)S^2/\sigma_0^2 > \chi_{n-1}^2(\alpha)$
	$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$		$(n-1)S^2/\sigma_0^2 < \chi_{n-1}^2(1-\alpha)$

这种基于检验统计量服从一定自由度的 χ^2 分布的检验方法称为 χ^2 检验.

例 5.2.3 某工厂生产的一种细纱支数服从正态分布, 其总体标准差为 1.2. 现从某日生产的一批产品中抽取 16 缕进行支数测量, 测得样本标准差为 2.1, 问纱的均匀度是否改变 ($\alpha = 0.05$)?

解 检验问题为

$$H_0 : \sigma^2 = 1.44 \leftrightarrow H_1 : \sigma^2 \neq 1.44.$$

检验的接受域为

$$\bar{D} = \left\{ (X_1, \dots, X_n) : \chi_{n-1}^2(1-\alpha/2) < \frac{(n-1)S^2}{\sigma_0^2} < \chi_{n-1}^2(\alpha/2) \right\},$$

其中 $n = 16$, $\alpha = 0.05$. 查表得到 $\chi_{15}^2(0.975) = 6.262$, $\chi_{15}^2(0.025) = 27.488$. 由已知数据算得 $S^2 = 2.1^2 = 4.41$, $\sigma_0^2 = 1.2^2 = 1.44$. 因此有

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{15 \times 4.41}{1.44} = 45.94 > 27.488.$$

故否定 H_0 , 即认为棉纱的均匀度发生改变.

5.2.3 两个正态总体均值差的检验

设 $\mathbf{X} = (X_1, \dots, X_m)$ 为自正态总体 $N(\mu_1, \sigma_1^2)$ 中抽取的简单随机样本, $\mathbf{Y} = (Y_1, \dots, Y_n)$ 为自正态总体 $N(\mu_2, \sigma_2^2)$ 中抽取的简单随机样本, 且样本 X_1, \dots, X_m 和 Y_1, \dots, Y_n 独立. 本段要讨论下列三类检验问题:

$$(7) H_0 : \mu_2 - \mu_1 = \mu_0 \leftrightarrow H_1 : \mu_2 - \mu_1 \neq \mu_0,$$

$$(8) H'_0 : \mu_2 - \mu_1 \leq \mu_0 \leftrightarrow H'_1 : \mu_2 - \mu_1 > \mu_0,$$

$$(9) H''_0 : \mu_2 - \mu_1 \geq \mu_0 \leftrightarrow H''_1 : \mu_2 - \mu_1 < \mu_0,$$

其中 μ_0 和检验水平 α 给定.

以上假设检验问题与求均值差的置信区间的 Behrens-Fisher 问题 (见 4.2.3 节) 一样, 除了它的几种特殊情形已获得圆满解决外, 一般情况至今尚没有得到简单精确的解法. 下面分几种情况分别讨论.

1. 当 σ_1^2 和 σ_2^2 已知时均值差的检验

首先考虑双边检验问题 (7), 即

$$H_0 : \mu_2 - \mu_1 = \mu_0 \leftrightarrow H_1 : \mu_2 - \mu_1 \neq \mu_0.$$

由于 $\bar{Y} - \bar{X}$ 是 $\mu_2 - \mu_1$ 的一个良好估计, 直观上看 $|\bar{Y} - \bar{X} - \mu_0|$ 越大, H_0 越不像成立, 故否定域可取如下形式 $\{(\mathbf{X}, \mathbf{Y}) : |\bar{Y} - \bar{X} - \mu_0| > A\}$, A 待定. 由于 $\bar{Y} - \bar{X} \sim N(\mu_2 - \mu_1, \sigma_1^2/m + \sigma_2^2/n)$, 故当 $\mu_2 - \mu_1 = \mu_0$ 时有

$$U = \frac{\bar{Y} - \bar{X} - \mu_0}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0, 1). \quad (5.2.10)$$

因此, 取检验统计量为 $U = (\bar{Y} - \bar{X} - \mu_0)/\sqrt{\sigma_1^2/m + \sigma_2^2/n}$, 则否定域等价形式为

$$\{(X_1, \dots, X_m; Y_1, \dots, Y_n) : |U| > c\}, \quad c \text{ 待定.}$$

记 $\theta = \mu_2 - \mu_1$, 为确定 c , 令

$$\alpha = P_\theta(|U| > c | H_0) = P\left(\left|\frac{\bar{Y} - \bar{X} - \mu_0}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}\right| > c | H_0\right) = 2 - 2\Phi(c),$$

由此可确定临界值 $c = u_{\alpha/2}$. 因此检验问题 (7) 的水平为 α 的检验否定域为

$$D_7 = \{(X_1, \dots, X_m; Y_1, \dots, Y_n) : |U| > u_{\alpha/2}\},$$

完全类似 5.2.1 节中的求检验问题 (2), (3) 方法, 可得检验问题 (8), (9) 的水平为 α 的检验的否定域为

$$D_8 = \{(X_1, \dots, X_m; Y_1, \dots, Y_n) : U > u_\alpha\},$$

$$D_9 = \{(X_1, \dots, X_m; Y_1, \dots, Y_n) : U < -u_\alpha\}.$$

这种基于由式 (5.2.10) 给出的检验统计量 U 的检验方法, 称为两样本 U 检验.

2. 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知时均值差的检验

若 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 已知, 则由前面刚刚讨论的两样本 U 检验可知, 此时检验统计量 U 变为

$$U = \frac{\bar{Y} - \bar{X} - \mu_0}{\sqrt{\sigma^2/m + \sigma^2/n}} = \frac{\bar{Y} - \bar{X} - \mu_0}{\sigma} \sqrt{\frac{mn}{m+n}}. \quad (5.2.11)$$

而在 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知时, 上述表达式中的 σ^2 常用

$$\begin{aligned} S_w^2 &= \frac{1}{n+m-2} [(m-1)S_1^2 + (n-1)S_2^2] \\ &= \frac{1}{n+m-2} \left[\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2 \right] \end{aligned} \quad (5.2.12)$$

来估计, 其中

$$S_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2 \quad (5.2.13)$$

分别为样本 X_1, \dots, X_m 和 Y_1, \dots, Y_n 的样本方差. 当 $\mu_2 - \mu_1 = \mu_0$ 时, 将式 (5.2.11) 中的 σ 用式 (5.2.12) 中的 S_w 代替, 得到下列的统计量 T_w , 由推论 2.4.3 可知

$$T_w = \frac{\bar{Y} - \bar{X} - \mu_0}{S_w} \sqrt{\frac{mn}{m+n}} \sim t_{n+m-2}. \quad (5.2.14)$$

因此取 T_w 作为检验统计量.

与 5.2.1 节中一样本 t 检验方法类似, 在两样本均值差的检验问题中, 用检验统计量 T_w 代替那里的一样本 t 检验统计量 T , 用完全相同的讨论方式, 可得检验问题 (7)-(9) 的水平为 α 的否定域, 只要注意在否定域中将确定临界值的 t 分布的自由度由 $n-1$ 改为 $n+m-2$ 即可. 详细结果见表 5.2.3.

表 5.2.3 两个正态总体均值差的假设检验

	H_0	H_1	检验统计量及其分布	否定域
已知	$\mu_2 - \mu_1 = \mu_0$	$\mu_2 - \mu_1 \neq \mu_0$	$U = \frac{\bar{Y} - \bar{X} - \mu_0}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}$ $U \mu_0 \sim N(0, 1)$	$ U > u_{\alpha/2}$
	$\mu_2 - \mu_1 \leq \mu_0$	$\mu_2 - \mu_1 > \mu_0$		$U > u_\alpha$
	$\mu_2 - \mu_1 \geq \mu_0$	$\mu_2 - \mu_1 < \mu_0$		$U < -u_\alpha$
未知	$\mu_2 - \mu_1 = \mu_0$	$\mu_2 - \mu_1 \neq \mu_0$	$T_w = \frac{\bar{Y} - \bar{X} - \mu_0}{S_w} \sqrt{\frac{mn}{m+n}}$ $T_w \mu_0 \sim t_{n+m-2}$ $S_w^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{n+m-2}$	$ T_w > t_{n+m-2} \left(\frac{\alpha}{2} \right)$
	$\mu_2 - \mu_1 \leq \mu_0$	$\mu_2 - \mu_1 > \mu_0$		$T_w > t_{n+m-2}(\alpha)$
	$\mu_2 - \mu_1 \geq \mu_0$	$\mu_2 - \mu_1 < \mu_0$		$T_w < -t_{n+m-2}(\alpha)$

这种基于检验统计量服从 t_{n+m-2} 分布的检验方法, 称为两样本 t 检验.

例 5.2.4 为研究正常成年男女血液红细胞平均数的差别, 检验某地正常成年男子 156 人, 女子 74 人, 计算男女红细胞的平均数和样本标准差分别为

$$\text{男: } \bar{X} = 465.13 \text{ 万/mm}^3, \quad S_1 = 54.80 \text{ 万/mm}^3,$$

$$\text{女: } \bar{Y} = 422.16 \text{ 万/mm}^3, \quad S_2 = 49.20 \text{ 万/mm}^3.$$

假定正常成年男女红细胞数分别服从正态分布, 且方差相同. 检验正常成年人红细胞数是否与性别有关 ($\alpha = 0.01$).

解 设 X_1, \dots, X_m i.i.d. $\sim N(\mu_1, \sigma^2)$; Y_1, \dots, Y_n i.i.d. $\sim N(\mu_2, \sigma^2)$, 且假定这两组样本独立. 检验问题为

$$H_0: \mu_2 - \mu_1 = 0 \leftrightarrow H_1: \mu_2 - \mu_1 \neq 0.$$

检验的否定域为

$$D = \left\{ (X_1, \dots, X_m; Y_1, \dots, Y_n) : \left| \frac{\bar{Y} - \bar{X}}{S_w} \sqrt{\frac{mn}{m+n}} \right| > t_{n+m-2} \left(\frac{\alpha}{2} \right) \right\},$$

其中

$$m = 156, \quad n = 74, \quad \bar{X} = 465.13, \quad \bar{Y} = 422.16, \quad S_1 = 54.80, \quad S_2 = 49.20,$$

$$S_w^2 = \frac{1}{n+m-2} [(m-1)S_1^2 + (n-1)S_2^2] = 2816.6, \quad S_w = 53.07,$$

查表得 $t_{228}(0.005) = 2.576$. 由

$$|T| = \left| \frac{\bar{Y} - \bar{X}}{S_w} \sqrt{\frac{mn}{m+n}} \right| = \left| \frac{422.16 - 465.13}{53.07} \sqrt{\frac{156 \times 74}{156+74}} \right| = 5.74 > 2.576,$$

故否定 H_0 , 即认为正常成年人的红细胞数与性别有关.

3. 当样本容量 $m = n$ 时均值差的检验 —— 成对比较问题

前面讨论的用于两个正态总体均值差的检验中, 假定了来自两个正态总体的样本是相互独立的. 但在实际问题中, 有时候情况不总是这样. 可能这两个正态总体的样本是来自同一个总体上的重复观察, 它们是成对出现的, 而且是相关的. 例如, 为了考察一种安眠药的效果, 记录了 n 个失眠患者服药前的每晚睡眠时间 X_1, X_2, \dots, X_n 和服用此安眠药后每晚睡眠时间 Y_1, Y_2, \dots, Y_n , 其中 (X_i, Y_i) 是第 i 个患者不服用安眠药和服用安眠药每晚的睡眠时间. 它们是有关系的, 不会相互独立. 另一方面, X_1, X_2, \dots, X_n 是 n 个不同失眠患者的睡眠时间, 由于个人体质诸方面的条件不同, 这 n 个观察值不能认为是来自同一个正态总体的样本. Y_1, Y_2, \dots, Y_n 也是一样. 这样的数据称为成对数据, 这样的数据模型用两样本 t 检

验就不合适. 因为 X_i 和 Y_i 是同在第 i 个患者身上观察到的夜晚睡眠时间, 所以 $Z_i = Y_i - X_i$ 就消除了人的体质诸方面的差异, 仅剩下安眠药的效果. 若安眠药无效, Z_i 的差异仅由随机误差引起, 随机误差可认为服从正态分布 $N(0, \sigma^2)$. 故可假定 Z_1, \dots, Z_n 为自 $N(\mu, \sigma^2)$ 中抽取的简单随机样本, μ 就是安眠药的平均效果. 安眠药是否有效, 就归结为检验如下假设

$$H_0: \mu = 0 \leftrightarrow H_1: \mu \neq 0.$$

因为 Z_1, \dots, Z_n 被认为是来自正态总体 $N(\mu, \sigma^2)$ 的简单样本, 故可用关于单个正态总体均值的 t 检验方法. 检验的否定域为

$$D = \{(Z_1, Z_2, \dots, Z_n) : |T_Z| > t_{n-1}(\alpha/2)\},$$

此处 α 为检验水平, $T_Z = \sqrt{n} \bar{Z}/S_Z$ 为检验统计量, 其中 \bar{Z} 和 S_Z^2 分别为 Z_1, Z_2, \dots, Z_n 的样本均值和样本方差.

例 5.2.5 今有两台测量材料中某种金属含量的光谱仪 A 和 B, 为鉴定它们的质量有无显著差异, 对金属含量不同的 9 件材料样品进行测量, 得到 9 对观察值为

u (单位: %): 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00.

v (单位: %): 0.10, 0.21, 0.52, 0.32, 0.78, 0.59, 0.68, 0.77, 0.89.

问根据实验结果, 在 $\alpha = 0.01$ 下, 能否判断这两台光谱仪的质量有无显著差异?

解 将光谱仪 A 和 B 对 9 件样品的测定值记为 X_1, X_2, \dots, X_9 , 和 Y_1, Y_2, \dots, Y_9 . 由于这 9 件样品金属含量不同, 所以 X_1, X_2, \dots, X_9 不能看成来自同一总体. Y_1, Y_2, \dots, Y_9 也一样; 每个对子 (X_i, Y_i) 中 X_i 与 Y_i 不独立. 故需用成对比较. 记

$$Z_i = Y_i - X_i, \quad i = 1, 2, \dots, 9.$$

若这两光谱仪质量一样, 测量得到的每对数据的差异仅由随机误差引起. 随机误差可认为服从正态分布 $N(0, \sigma^2)$. 故可假定 Z_1, \dots, Z_n 为自 $N(\mu, \sigma^2)$ 中抽取的随机样本, 要检验

$$H_0: \mu = 0 \leftrightarrow H_1: \mu \neq 0, \quad \alpha = 0.01.$$

由表 5.2.1 可知此检验的否定域为

$$\{(Z_1, \dots, Z_n) : |T_Z| > t_{n-1}(\alpha/2)\},$$

其中 $n = 9$. 由题中数据算得

$$\bar{Z} = \frac{1}{9} \sum_{i=1}^9 Z_i = 0.06, \quad S_Z^2 = \frac{1}{8} \sum_{i=1}^9 (Z_i - \bar{Z})^2 = 0.01505, \quad S_Z = 0.12268,$$

查表得 $t_{n-1}(\alpha/2) = t_8(0.005) = 3.3554$. 由于

$$|T_z| = \left| \frac{\sqrt{n}\bar{Z}}{S_z} \right| = \frac{3 \times 0.06}{0.12268} = 1.47 < 3.3554,$$

故无足够证据显示两台仪器有显著差异, 因此接受 H_0 .

5.2.4 两个正态总体方差比的检验

设 $\mathbf{X} = (X_1, \dots, X_m)$ 是从正态总体 $N(\mu_1, \sigma_1^2)$ 中抽取的简单随机样本, $\mathbf{Y} = (Y_1, \dots, Y_n)$ 是从正态总体 $N(\mu_2, \sigma_2^2)$ 中抽取的简单随机样本, 且样本 X_1, \dots, X_m 和 Y_1, \dots, Y_n 独立. 讨论下列三类假设检验问题:

$$(10) H_0 : \frac{\sigma_2^2}{\sigma_1^2} = 1 \leftrightarrow H_1 : \frac{\sigma_2^2}{\sigma_1^2} \neq 1,$$

$$(11) H'_0 : \frac{\sigma_2^2}{\sigma_1^2} \leq 1 \leftrightarrow H'_1 : \frac{\sigma_2^2}{\sigma_1^2} > 1,$$

$$(12) H''_0 : \frac{\sigma_2^2}{\sigma_1^2} \geq 1 \leftrightarrow H''_1 : \frac{\sigma_2^2}{\sigma_1^2} < 1,$$

检验水平 α 给定.

记 \bar{X} 和 S_1^2 为 X_1, \dots, X_m 的样本均值和样本方差; \bar{Y} 和 S_2^2 为 Y_1, \dots, Y_n 的样本均值和样本方差, 其中 S_1^2 和 S_2^2 如式 (5.2.13) 所示.

1. 当 μ_1 和 μ_2 未知时方差比的检验方法

首先讨论检验问题 (10), 即

$$H_0 : \frac{\sigma_2^2}{\sigma_1^2} = 1 \leftrightarrow H_1 : \frac{\sigma_2^2}{\sigma_1^2} \neq 1.$$

由于 S_1^2 和 S_2^2 分别是 σ_1^2 和 σ_2^2 的无偏估计, 并具有良好性质. 直观上看, S_2^2/S_1^2 太小或者太大时, H_0 不像成立. 可设想否定域的形式为

$$\left\{ (X_1, \dots, X_m; Y_1, \dots, Y_n) : \frac{S_2^2}{S_1^2} < c_1 \text{ 或 } \frac{S_2^2}{S_1^2} > c_2 \right\}, \quad c_1, c_2 \text{ 待定.}$$

在 $\sigma_2^2/\sigma_1^2 = 1$ 的条件下, 由推论 2.4.4 可知

$$F = \frac{S_2^2}{S_1^2} \sim F_{n-1, m-1}. \tag{5.2.15}$$

因此取检验统计量为 $F = S_2^2/S_1^2$. 记 $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, 为了确定否定域中的临界值 c_1, c_2 , 令

$$P_\theta \left(\frac{S_2^2}{S_1^2} < c_1 \text{ 或 } \frac{S_2^2}{S_1^2} > c_2 \mid H_0 \right) = \alpha.$$

满足上式要求的 c_1 和 c_2 有很多, 其中存在一对 c_1, c_2 最优, 但计算复杂, 使用不方便. 确定 c_1, c_2 的一个简单实用的方法是令

$$P_\theta \left(\frac{S_2^2}{S_1^2} < c_1 \mid H_0 \right) = \frac{\alpha}{2}, \quad P_\theta \left(\frac{S_2^2}{S_1^2} > c_2 \mid H_0 \right) = \frac{\alpha}{2}.$$

由上述两式和式 (5.2.15) 易知临界值 $c_1 = F_{n-1, m-1}(1 - \alpha/2)$, $c_2 = F_{n-1, m-1}(\alpha/2)$, 所以检验问题 (10) 的水平为 α 的接受域为

$$\bar{D}_{10} = \left\{ (\mathbf{X}, \mathbf{Y}) : F_{n-1, m-1}(1 - \alpha/2) \leq \frac{S_2^2}{S_1^2} \leq F_{n-1, m-1}(\alpha/2) \right\}.$$

此接受域的表达式比否定域简单, 使用方便, 故此处采用接受域代替否定域.

用完全类似于 5.2.1 节中求检验问题 (2), (3) 的方法可分别求得检验问题 (11) 和 (12) 的水平为 α 的否定域如下:

$$D_{11} = \left\{ (X_1, \dots, X_m; Y_1, \dots, Y_n) : \frac{S_2^2}{S_1^2} > F_{n-1, m-1}(\alpha) \right\},$$

$$D_{12} = \left\{ (X_1, \dots, X_m; Y_1, \dots, Y_n) : \frac{S_2^2}{S_1^2} < F_{n-1, m-1}(1 - \alpha) \right\}.$$

这里要注意一点的是: 当 α 是较小的数, 如 $\alpha = 0.01, 0.05$ 时, 从 F 分布的分位数表上查不到 $F_{n-1, m-1}(1 - \alpha)$ 的数值, 但利用习题 2, 第 32 题已证明的事实

$$F_{n, m}(1 - \alpha) = \frac{1}{F_{m, n}(\alpha)}, \quad (5.2.16)$$

使问题获得解决. 例如, 从表上查不到 $F_{5, 10}(1 - 0.01)$ 的值, 但可查到 $F_{10, 5}(0.01)$ 的值, 利用式 (5.2.16) 可知 $F_{5, 10}(1 - 0.01) = 1/F_{10, 5}(0.01)$, 从而可求得所要的数值.

2. 当 μ_1 和 μ_2 已知时, 方差比的检验方法

简述如下: 当 μ_1 和 μ_2 已知时, σ_1^2 和 σ_2^2 具有良好性质的无偏估计分别是

$$S_{1*}^2 = \frac{1}{m} \sum_{i=1}^m (X_i - \mu_1)^2, \quad S_{2*}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \mu_2)^2.$$

当 $\sigma_1^2/\sigma_2^2 = 1$ 时, 利用推论 2.4.1 和 F 分布的定义, 容易证明在 H_0 成立的条件下, 有

$$F_* = \frac{S_{2*}^2}{S_{1*}^2} \sim F_{n, m}. \quad (5.2.17)$$

因此, 取检验统计量 F_* 代替 $F = S_2^2/S_1^2$, 完全类似于 μ_1 和 μ_2 未知情形的讨论, 可得到检验问题 (10)-(12) 的水平为 α 的否定域, 只要注意在否定域中, 将确定临界值的 F 分布的自由度由 $n-1, m-1$ 分别改为 n, m 即可. 详细结果见表 5.2.4.

这种基于检验统计量服从 F 分布的检验方法, 称为 F 检验.

表 5.2.4 两个正态总体方差比的假设检验

	H_0	H_1	检验统计量及其分布	否定域
已知	$\sigma_1^2 = \sigma_2^2$	$\sigma_2^2 \neq \sigma_1^2$	$F_* = S_{2*}^2 / S_{1*}^2$ $F_* _{\sigma_2^2 = \sigma_1^2} \sim F_{n,m}$	$F_* < F_{n,m}(1 - \alpha/2)$ 或 $F_* > F_{n,m}(\alpha/2)$
	$\sigma_2^2 \leq \sigma_1^2$	$\sigma_2^2 > \sigma_1^2$	$S_{1*}^2 = \frac{1}{m} \sum_{i=1}^m (X_i - \mu_1)^2$	$F_* > F_{n,m}(\alpha)$
	$\sigma_2^2 \geq \sigma_1^2$	$\sigma_2^2 < \sigma_1^2$	$S_{2*}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \mu_2)^2$	$F_* < F_{n,m}(1 - \alpha)$
	$\sigma_1^2 = \sigma_2^2$	$\sigma_2^2 \neq \sigma_1^2$	$F = S_2^2 / S_1^2$ $F _{\sigma_2^2 = \sigma_1^2} \sim F_{n-1,m-1}$	$F < F_{n-1,m-1}(1 - \alpha/2)$ 或 $F > F_{n-1,m-1}(\alpha/2)$
未知	$\sigma_2^2 \leq \sigma_1^2$	$\sigma_2^2 > \sigma_1^2$	$S_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$	$F > F_{n-1,m-1}(\alpha)$
	$\sigma_2^2 \geq \sigma_1^2$	$\sigma_2^2 < \sigma_1^2$	$S_2^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2$	$F < F_{n-1,m-1}(1 - \alpha)$

例 5.2.6 测得两批样本大小皆为 6 的电子器材电阻的均值 $\bar{X} = 0.14$, $\bar{Y} = 0.139$, 样本标准差分别为 $S_1 = 0.0026$, $S_2 = 0.0024$, 假设这两批器材的电阻分别服从 $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$, 均值方差皆未知且两组样本独立, 问这两批电子器件的电阻是否相同? ($\alpha = 0.05$)

解 这个问题表面看是对两个正态总体均值差的检验, 但不知道是否有 $\sigma_1^2 = \sigma_2^2$, 因此首先求两个正态总体方差是否相同的检验. 如果检验认为 $\sigma_1^2 = \sigma_2^2$, 然后再作两样本 t 检验. 如果经检验否定了 $\sigma_1^2 = \sigma_2^2$, 则不能用两样本的 t 检验方法去检验均值差, 这就变成 Behrens-Fisher 问题, 将留在本节最后解决.

首先考虑下列检验问题.

$$(1) H_0 : \sigma_1^2 = \sigma_2^2 \leftrightarrow H_1 : \sigma_1^2 \neq \sigma_2^2, \quad \alpha = 0.05.$$

由表 5.2.4 可知, 此检验的接受域是

$$\left\{ (\mathbf{X}, \mathbf{Y}) : F_{m-1,n-1}(1 - \alpha/2) \leq \frac{S_1^2}{S_2^2} \leq F_{m-1,n-1}(\alpha/2) \right\},$$

此处 $m = n = 6$, $S_1^2/S_2^2 = 0.0026^2/0.0024^2 = 1.17$. 由 $\alpha = 0.05$, 查 F 分布表得 $F_{5,5}(0.025) = 7.15$. 由于

$$\frac{1}{7.15} = F_{5,5}(1 - 0.025) < F = \frac{S_1^2}{S_2^2} = 1.17 < F_{5,5}(0.025) = 7.15,$$

故认为没有足够的证据否定 H_0 , 因此接受 H_0 .

在接受上述检验后, 可以假定 $\sigma_1^2 = \sigma_2^2$, 进一步考虑下列检验问题.

$$(2) H'_0 : \mu_1 = \mu_2 \leftrightarrow H'_1 : \mu_1 \neq \mu_2, \quad \alpha = 0.05.$$

由表 5.2.3 可知此检验的否定域为

$$\{(\bar{X}, \bar{Y}) : |\bar{T}_w| > t_{n+m-2}(\alpha/2)\},$$

此处 $m = n = 6$, $\bar{X} = 0.14$, $\bar{Y} = 0.139$, $S_1 = 0.0026$, $S_2 = 0.0024$, 因此有

$$S_w^2 = \frac{1}{10}(5 \times 0.0026^2 + 5 \times 0.0024^2) = 6.26 \times 10^{-6}, \quad S_w = 0.0025.$$

由 $\alpha = 0.05$, 查 t 分布表得 $t_{10}(0.025) = 2.228$. 由于

$$|\bar{T}| = \sqrt{\frac{nm}{n+m}} \left| \frac{\bar{Y} - \bar{X}}{S_w} \right| = \sqrt{3} \times \left| \frac{0.14 - 0.139}{0.0025} \right| = 0.6928 < 2.228,$$

故没有充足的理由否定两批电子器件的电阻值相同, 因此接受 H'_0 .

5.2.5 极限分布为正态分布的检验

本段讨论 Behrens-Fisher 问题的大样本检验, 附带也给出这一问题的一个小样本检验的近似方法. 同时也讨论二项分布和 Poisson 分布参数的大样本检验问题.

1. Behrens-Fisher 检验问题的近似方法

设 X_1, \dots, X_m 是从正态总体 $N(\mu_1, \sigma_1^2)$ 中抽取的简单随机样本, Y_1, \dots, Y_n 是正态总体 $N(\mu_2, \sigma_2^2)$ 中抽取的简单随机样本, 且样本 X_1, \dots, X_m 和 Y_1, \dots, Y_n 独立. 考虑 $\mu_2 - \mu_1$ 的下列三类检验问题:

- (a) $H_0 : \mu_2 - \mu_1 = \mu_0 \leftrightarrow H_1 : \mu_2 - \mu_1 \neq \mu_0$,
- (b) $H'_0 : \mu_2 - \mu_1 \leq \mu_0 \leftrightarrow H'_1 : \mu_2 - \mu_1 > \mu_0$,
- (c) $H''_0 : \mu_2 - \mu_1 \geq \mu_0 \leftrightarrow H''_1 : \mu_2 - \mu_1 < \mu_0$,

其中 μ_0 和检验水平 α 给定.

上述检验问题的几个特例情形: ① σ_1^2 和 σ_2^2 已知; ② $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知; ③ 成对比较问题 ($m = n$); 这三种情形的检验问题已经在 5.2.3 节中解决. 当 σ_1^2 和 σ_2^2 未知且不相等时的检验问题称为 Behrens-Fisher 问题, 是尚未解决的问题, 下面将介绍处理这类检验问题的两个近似方法.

记 S_1^2 和 S_2^2 分别为两组样本的样本方差, 如式 (5.2.13) 所示. 将分下列两种情形来讨论:

(1) 当 σ_1^2 和 σ_2^2 未知, 但 m, n 充分大时, 可利用基于中心极限定理的近似解法. 由于 $\bar{Y} - \bar{X} \sim N(\mu_2 - \mu_1, \sigma_1^2/m + \sigma_2^2/n)$, 将其标准化有

$$\frac{\bar{Y} - \bar{X} - (\mu_2 - \mu_1)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0, 1).$$

由于 S_1^2 和 S_2^2 分别为 σ_1^2 和 σ_2^2 的相合估计, 用 S_1^2 和 S_2^2 代替 σ_1^2 和 σ_2^2 , 由引理 2.5.1 可知, 当 $m, n \rightarrow \infty$ 时有

$$\frac{\bar{Y} - \bar{X} - (\mu_2 - \mu_1)}{\sqrt{S_1^2/m + S_2^2/n}} = \frac{\bar{Y} - \bar{X} - (\mu_2 - \mu_1)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \cdot \frac{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}{\sqrt{S_1^2/m + S_2^2/n}} \xrightarrow{\mathcal{L}} N(0, 1).$$

特别当 $\mu_2 - \mu_1 = \mu_0$ 且 $m, n \rightarrow \infty$ 时, 有

$$U^* = \frac{\bar{Y} - \bar{X} - \mu_0}{\sqrt{S_1^2/m + S_2^2/n}} \xrightarrow{\mathcal{L}} N(0, 1). \quad (5.2.18)$$

取 U^* 作为检验统计量, 用与 5.2.1 小节中完全类似的方法得到检验问题 (a)–(c) 的下列的否定域, 但要注意到此处的检验水平不是精确为 α , 而是近似为 α .

$$\begin{aligned} D_a &= \{(\mathbf{X}, \mathbf{Y}) : |U^*| > u_{\alpha/2}\}; \\ D_b &= \{(\mathbf{X}, \mathbf{Y}) : U^* > u_\alpha\}; \\ D_c &= \{(\mathbf{X}, \mathbf{Y}) : U^* < -u_\alpha\}. \end{aligned} \quad (5.2.19)$$

*(2) 当 σ_1^2 和 σ_2^2 未知, 但 m, n 不都是充分大时, 可用基于 t 分布的近似解法. 在 4.2 节 Behrens-Fisher 区间估计问题中, 已说明了

$$\frac{\bar{Y} - \bar{X} - (\mu_2 - \mu_1)}{\sqrt{S_1^2/m + S_2^2/n}}$$

近似服从自由度为 r 的 t 分布, 其中

$$r = S_*^4 / \left[\frac{S_1^4}{m^2(m-1)} + \frac{S_2^4}{n^2(n-1)} \right], \quad S_*^2 = \frac{S_1^2}{m} + \frac{S_2^2}{n}.$$

当 $\mu_2 - \mu_1 = \mu_0$ 时, 近似地有

$$T_* = \frac{\bar{Y} - \bar{X} - \mu_0}{\sqrt{S_1^2/m + S_2^2/n}} \sim t_r, \quad (5.2.20)$$

即 T_* 近似服从自由度为 r 的 t 分布. 取 T_* 作为检验统计量, 类似于 5.2.1 节中一样本 t 检验法得到检验问题 (a)–(c) 检验水平近似为 α 的检验的否定域

$$\begin{aligned} D_a &= \{(\mathbf{X}, \mathbf{Y}) : |T_*| > t_r(\alpha/2)\}; \\ D_b &= \{(\mathbf{X}, \mathbf{Y}) : T_* > t_r(\alpha)\}; \\ D_c &= \{(\mathbf{X}, \mathbf{Y}) : T_* < -t_r(\alpha)\}, \end{aligned} \quad (5.2.21)$$

例 5.2.7 在例 5.2.4 中假定正常成年男女红细胞数皆服从正态分布, 数据作如下修改: 不假定两组样本方差相等, 正常男女红细胞的样本平均数不变, 样本标

准差改为 $S_1 = 54.80 \text{ 万/mm}^3$, $S_2 = 39.20 \text{ 万/mm}^3$, 要求检验正常成年人红细胞数是否与性别有关. 取 $\alpha = 0.02$.

解 由于两组样本方差的相等性取消了. 用 F 检验来检验两组样本方差是否相等的结果是: 否定 H_0 , 即有足够的把握认为两总体方差不相等. 这是一个 Behrens-Fisher 问题. 由于样本容量较大, 故可用大样本方法来求下列检验问题:

$$H_0: \mu_2 - \mu_1 = 0 \leftrightarrow H_1: \mu_2 - \mu_1 \neq 0.$$

由式 (5.2.19) 可知检验的否定域为

$$\{(\mathbf{X}, \mathbf{Y}): |U^*| > u_{\alpha/2}\},$$

其中检验统计量 $U^* = (\bar{Y} - \bar{X}) / \sqrt{S_1^2/m + S_2^2/n}$, $m = 176$, $n = 74$, $\bar{X} = 465.13$, $\bar{Y} = 422.16$, $S_1 = 54.80$, $S_2 = 39.20$, 查表得 $u_{0.01} = 2.30$. 由于

$$|U^*| = \frac{|465.13 - 422.16|}{\sqrt{54.8^2/176 + 39.2^2/74}} = \frac{42.97}{6.15} = 6.99 > 2.30,$$

故否定 H_0 , 即认为正常成年人红细胞数与性别有关.

* **例 5.2.8** 分别抽取甲种矿石的 10 个样品和乙种矿石的 5 个样品, 测其含铁量 (单位: %), 由数据算得

$$\text{甲矿石: } \bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i = 16.01, \quad S_1^2 = \frac{1}{9} \sum_{i=1}^{10} (X_i - \bar{X})^2 = 10.80,$$

$$\text{乙矿石: } \bar{Y} = \frac{1}{5} \sum_{j=1}^5 Y_j = 18.98, \quad S_2^2 = \frac{1}{4} \sum_{i=1}^5 (Y_i - \bar{Y})^2 = 0.27.$$

在显著性水平 $\alpha = 0.01$ 下, 求检验问题: 甲矿石含铁量不低于乙矿石的含铁量.

解 设 X_1, \dots, X_m i.i.d. $\sim N(\mu_1, \sigma_1^2)$, Y_1, \dots, Y_n i.i.d. $\sim N(\mu_2, \sigma_2^2)$, 此处 $m = 10$, $n = 5$. 因 S_1^2 和 S_2^2 相差甚大, 假定 $\sigma_1^2 = \sigma_2^2$ 是不合理的. 由于 m 和 n 都较小, 不宜用大样本方法, 故此问题属于 Behrens-Fisher 问题, 用基于 t 分布的检验方法. 只能在方差未知的一般情形下, 检验假设

$$H_0: \mu_1 \geq \mu_2 \leftrightarrow H_1: \mu_1 < \mu_2.$$

检验的否定域

$$D = \left\{ (\mathbf{X}, \mathbf{Y}): T_* = \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/m + S_2^2/n}} < -t_r(\alpha) \right\}.$$

首先计算 t 分布的自由度

$$r = \frac{(S_1^2/m + S_2^2/n)^2}{S_1^4/m^2(m-1) + S_2^4/n^2(n-1)} = 9.88 \approx 10,$$

查表得 $t_{10}(0.01) = 2.764$, 由数据得

$$T_* = \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/m + S_2^2/n}} = \frac{-2.97}{1.065} = -2.79 < -2.764,$$

否定 H_0 , 即认为甲矿石含铁量低于乙矿石.

2. 二项分布参数的大样本检验

设 X_1, \dots, X_n i.i.d. $\sim b(1, p)$, 显见 $T = \sum_{i=1}^n X_i \sim b(n, p)$, 考虑下列检验问题:

$$H_0 : p = p_0 \leftrightarrow H_1 : p \neq p_0, \quad (5.2.22)$$

其中 p_0 和检验水平 α 给定.

由独立同分布场合的中心极限定理可知 $(T - np)/\sqrt{np(1-p)} \xrightarrow{\mathcal{L}} N(0, 1)$, 当 $n \rightarrow \infty$ 时. 故当 H_0 成立, 即 $p = p_0$ 时有

$$U = \frac{T - np_0}{\sqrt{np_0(1-p_0)}} \xrightarrow{\mathcal{L}} N(0, 1), \quad \text{当 } n \rightarrow \infty.$$

因此取 U 作为检验统计量. 当 n 较大时, U 近似服从 $N(0, 1)$ 分布. 由 U 检验法可知检验问题 (5.2.22) 水平近似为 α 的否定域为

$$D_1 = \{(X_1, \dots, X_n) : |U| > u_{\alpha/2}\}.$$

类似可知 p 的两个单边检验问题及水平近似为 α 的否定域如下:

$$H'_0 : p \leq p_0 \leftrightarrow H'_1 : p > p_0, \quad D_2 = \{(X_1, \dots, X_n) : U > u_\alpha\}.$$

$$H''_0 : p \geq p_0 \leftrightarrow H''_1 : p < p_0, \quad D_3 = \{(X_1, \dots, X_n) : U < -u_\alpha\}.$$

3. Poisson 分布参数的大样本检验

设 X_1, \dots, X_n 为自 Poisson 总体 $P(\lambda)$ 中抽取的随机样本, 考虑检验问题

$$H_0 : \lambda = \lambda_0 \leftrightarrow H_1 : \lambda \neq \lambda_0, \quad (5.2.23)$$

其中 λ_0 和检验水平 α 给定.

由于 $T = \sum_{i=1}^n X_i$ 服从参数为 $n\lambda$ 的 Poisson 分布 $P(n\lambda)$. 由中心极限定理可知 $(T - n\lambda)/\sqrt{n\lambda} \xrightarrow{\mathcal{L}} N(0, 1)$, 当 $n \rightarrow \infty$. 因此当 H_0 成立, 即 $\lambda = \lambda_0$ 时有

$$U_0 = \frac{T - n\lambda_0}{\sqrt{n\lambda_0}} \xrightarrow{\mathcal{L}} N(0, 1), \quad \text{当 } n \rightarrow \infty.$$

因此取 U_0 作为检验统计量. 当 n 较大时, U_0 近似服从 $N(0, 1)$ 分布. 由 U 检验法可知双边检验问题 (5.2.23) 的水平近似为 α 的否定域为

$$D_1 = \{(X_1, \dots, X_n) : |U_0| > u_{\alpha/2}\}.$$

类似可知 λ 的两个单边检验问题及水平近似为 α 的否定域如下:

$$H'_0 : \lambda \leq \lambda_0 \leftrightarrow H'_1 : \lambda > \lambda_0, \quad D_2 = \{(X_1, \dots, X_n) : U_0 > u_\alpha\}.$$

$$H''_0 : \lambda \geq \lambda_0 \leftrightarrow H''_1 : \lambda < \lambda_0, \quad D_2 = \{(X_1, \dots, X_n) : U_0 < -u_\alpha\}.$$

4. 两样本检验问题

设 X_1, \dots, X_m i.i.d. $\sim b(1, p_1)$, Y_1, \dots, Y_n i.i.d. $\sim b(1, p_2)$, 且样本 X_1, \dots, X_m 和 Y_1, \dots, Y_n 独立. 求检验问题

$$H_0 : p_2 - p_1 = 0 \leftrightarrow H_1 : p_2 - p_1 \neq 0, \quad (5.2.24)$$

检验水平 α 给定.

记 \bar{X} 和 \bar{Y} 分别为两组样本的均值. 由中心极限定理可知

$$\frac{\bar{Y} - \bar{X} - (p_2 - p_1)}{\sqrt{p_1(1-p_1)/m + p_2(1-p_2)/n}} \xrightarrow{\mathcal{L}} N(0, 1), \quad \text{当 } n, m \rightarrow \infty.$$

当 H_0 成立, 即 $p_1 = p_2 = p$ 时, 将 p 用合样本估计, 即取

$$\hat{p} = \frac{1}{m+n} \left(\sum_{i=1}^m X_i + \sum_{j=1}^n Y_j \right),$$

显然 \hat{p} 为 p 的相合估计, 故由引理 2.5.1 可知, 当 $m, n \rightarrow \infty$ 时, 有

$$\tilde{U} = \frac{\bar{Y} - \bar{X}}{\sqrt{\hat{p}(1-\hat{p})}} \sqrt{\frac{mn}{m+n}} = \frac{\bar{Y} - \bar{X}}{\sqrt{p(1-p)}} \sqrt{\frac{mn}{m+n}} \cdot \frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{\mathcal{L}} N(0, 1).$$

因此取 \tilde{U} 为检验统计量, 当 m, n 都较大时, \tilde{U} 近似服从 $N(0, 1)$ 分布. 由 U 检验法得到双边检验问题 (5.2.24) 的检验水平近似为 α 的否定域为

$$D_1 = \{(X_1, \dots, X_m; Y_1, \dots, Y_n) : |\tilde{U}| > u_{\alpha/2}\}.$$

类似可知两个单边检验问题及水平近似为 α 的否定域如下:

$$H'_0 : p_2 \leq p_1 \leftrightarrow H'_1 : p_2 > p_1, \quad D_2 = \{(X_1, \dots, X_m; Y_1, \dots, Y_n) : \tilde{U} > u_\alpha\}$$

$$H''_0 : p_2 \geq p_1 \leftrightarrow H''_1 : p_2 < p_1, \quad D_3 = \{(X_1, \dots, X_m; Y_1, \dots, Y_n) : \tilde{U} < -u_\alpha\}.$$

Poisson 分布的两样本检验问题可用类似方法讨论, 检验统计量的选取和检验否定域的形式留给读者作为练习.

5.3 似然比检验

似然比检验是 Neyman 和 Pearson 在 1928 年提出的构造假设检验的一般方法。它在假设检验中的地位，相当于极大似然估计在点估计中的地位。它可视为 Fisher 的极大似然原理在假设检验问题中的体现。由这种方法构造出来的检验，一般说有比较良好的性质。这个方法的一个重要优点就是适用面广。就是说，它对分布族的形式没有什么特殊的要求。

5.3.1 似然比检验的定义

设有分布族 $\{f(x, \theta), \theta \in \Theta\}$, Θ 为参数空间。令 $\mathbf{X} = (X_1, \dots, X_n)$ 为自上述分布族中抽取的简单随机样本, $f(\mathbf{x}, \theta)$ 为样本的概率函数。要考虑检验问题 (5.1.1)。在有了样本 \mathbf{x} 后将 $f(\mathbf{x}, \theta)$ 视为 θ 的函数, 称为似然函数。如第 3 章介绍极大似然估计时所述。若 $f(\mathbf{x}, \theta_1) < f(\mathbf{x}, \theta_2)$, 则认为真参数为 θ_2 的“似然性”较其为 θ_1 的“似然性”大。由于假设检验在 “ $\theta \in \Theta_0$ 与 $\theta \in \Theta_1$ ” 这二者中选其一, 自然考虑以下两个量:

$$\begin{aligned} L_{\Theta_0}(\mathbf{x}) &= \sup_{\theta \in \Theta_0} f(\mathbf{x}, \theta), \\ L_{\Theta_1}(\mathbf{x}) &= \sup_{\theta \in \Theta_1} f(\mathbf{x}, \theta). \end{aligned}$$

考虑其比值 $L_{\Theta_1}(\mathbf{x})/L_{\Theta_0}(\mathbf{x})$, 若此比值较大, 则说明真参数在 Θ_1 内的“似然性”较大, 因而倾向于否定假设 “ $\theta \in \Theta_0$ ”。反之, 若此比值较小, 倾向于接受假设 “ $\theta \in \Theta_0$ ”。

若记 $\lambda(\mathbf{x}) = L_{\Theta_1}(\mathbf{x})/L_{\Theta_0}(\mathbf{x})$, 其中 $L_{\Theta}(\mathbf{x}) = \sup_{\theta \in \Theta} f(\mathbf{x}, \theta)$ 。由于 $\lambda(\mathbf{x})$ 与 $L_{\Theta_1}(\mathbf{x})/L_{\Theta_0}(\mathbf{x})$ 同增或同减, 可用 $\lambda(\mathbf{x})$ 代替比值 $L_{\Theta_1}(\mathbf{x})/L_{\Theta_0}(\mathbf{x})$, 这样做的好处是 $L_{\Theta}(\mathbf{x}) = \sup_{\theta \in \Theta} f(\mathbf{x}, \theta)$ 的计算比 $L_{\Theta_1}(\mathbf{x})$ 要容易。因此得到如下定义。

定义 5.3.1 设样本 \mathbf{X} 有概率函数 $f(\mathbf{x}, \theta)$, $\theta \in \Theta$, 而 Θ_0 为参数空间 Θ 的真子集, 考虑检验问题 (5.1.1), 则统计量

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta} f(\mathbf{x}, \theta)}{\sup_{\theta \in \Theta_0} f(\mathbf{x}, \theta)} \quad (5.3.1)$$

称为关于该检验问题的似然比。而由下述定义的检验函数

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \lambda(\mathbf{x}) > c, \\ r, & \lambda(\mathbf{x}) = c, \\ 0, & \lambda(\mathbf{x}) < c, \end{cases} \quad (5.3.2)$$

其中 c, r ($0 \leq r \leq 1$) 为待定常数, 称为检验问题 (5.1.1) 的一个似然比检验 (likelihood ratio test), 有些文献中也称其为广义似然比检验.

若样本分布为连续分布时, 在式 (5.3.2) 中令 $r = 0$, 即

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \lambda(\mathbf{x}) > c, \\ 0, & \lambda(\mathbf{x}) \leq c. \end{cases}$$

在式 (5.3.2) 中常数 c 和 r 的选择是要使检验具有给定的水平 α .

根据上面所说, 找似然比检验有以下步骤:

- (1) 求似然函数 $f(\mathbf{x}, \theta)$, 并明确参数空间 Θ 和 Θ_0 是什么.
- (2) 算出 $L_\Theta(\mathbf{x}) = \sup_{\theta \in \Theta} f(\mathbf{x}, \theta)$ 和 $L_{\Theta_0}(\mathbf{x}) = \sup_{\theta \in \Theta_0} f(\mathbf{x}, \theta)$.
- (3) 求出 $\lambda(\mathbf{x})$ 或与其等价的统计量的分布.
- (4) 确定 c 和 r 使式 (5.3.2) 具有给定的检验水平 α .

其中最关键的是第 3 步. 一般 $\lambda(\mathbf{x})$ 的表达式复杂, 求其分布不易. 但若 $\lambda(\mathbf{x}) = g(T(\mathbf{x}))$ 为 $T(\mathbf{x})$ 的单调上升 (或下降) 函数, 则检验式 (5.3.2) 显然等价于

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > c', \\ r, & T(\mathbf{x}) = c', \\ 0, & T(\mathbf{x}) < c'. \end{cases}$$

因此代替求 $\lambda(\mathbf{X})$ 的分布, 只要求出 $T(\mathbf{X})$ 的分布即可 (若 $\lambda(\mathbf{x})$ 为 $T(\mathbf{x})$ 的单调下降函数, 则将 $\varphi(\mathbf{x})$ 中的不等式反向).

如果 $\lambda(\mathbf{X})$ 分布无法求得, 可用其极限分布近似代替, 这一情形将在本节最后一段介绍.

5.3.2 若干例子

例 5.3.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态分布族 $\{N(\mu, \sigma^2), -\infty < \mu < +\infty, \sigma^2 > 0\}$ 中抽取的随机样本, 求下列检验问题的水平为 α 的似然比检验:

$$H_0: \mu = \mu_0 \leftrightarrow H_1: \mu \neq \mu_0. \quad (5.3.3)$$

解 记 $\boldsymbol{\theta} = (\mu, \sigma^2)$, 则 $\boldsymbol{\theta}$ 的似然函数为

$$f(\mathbf{x}, \boldsymbol{\theta}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\}, \quad (5.3.4)$$

在这里, 参数空间为

$$\Theta = \{\boldsymbol{\theta} = (\mu, \sigma^2) : -\infty < \mu < +\infty, \sigma^2 > 0\}.$$

零假设 H_0 对应的 θ 的子集为

$$\Theta_0 = \{\boldsymbol{\theta} = (\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}.$$

在 θ 上, μ 和 σ^2 的极大似然估计 (MLE) 为

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2;$$

在 Θ_0 上, σ^2 的 MLE 为

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2.$$

故有

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \Theta} f(\mathbf{x}, \boldsymbol{\theta}) &= f(\mathbf{x}, \hat{\mu}, \hat{\sigma}^2) = \left(\frac{2\pi e}{n} \right)^{-n/2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^{-n/2}, \\ \sup_{\boldsymbol{\theta} \in \Theta_0} f(\mathbf{x}, \boldsymbol{\theta}) &= f(\mathbf{x}, \mu_0, \tilde{\sigma}^2) = \left(\frac{2\pi e}{n} \right)^{-n/2} \left(\sum_{i=1}^n (x_i - \mu_0)^2 \right)^{-n/2}. \end{aligned} \quad (5.3.5)$$

从而有

$$\begin{aligned} \lambda(\mathbf{x}) &= \left[\sum_{i=1}^n (x_i - \bar{x})^2 \Big/ \sum_{i=1}^n (x_i - \mu_0)^2 \right]^{-n/2} \\ &= \left[1 + n(\bar{x} - \mu_0)^2 \Big/ \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{n/2} \\ &= \left(1 + \frac{1}{n-1} T^2 \right)^{\frac{n}{2}}, \end{aligned}$$

其中 $T = \sqrt{n}(\bar{X} - \mu_0)/S$, 而 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$. 由于 $\lambda(\mathbf{x})$ 为 $|T|$ 的严格增函数, 故检验的否定域 $D = \{\mathbf{X} = (X_1, \dots, X_n) : \lambda(\mathbf{X}) > c'\} = \{\mathbf{X} : |T| > c\}$. 令

$$P(|T| > c | H_0) = \alpha.$$

利用下列事实: 当 H_0 成立时 $T \sim t_{n-1}$, 则可知 $c = t_{n-1}(\alpha/2)$. 因此

$$\varphi(\mathbf{x}) = \begin{cases} 1, & |T| > t_{n-1}(\alpha/2), \\ 0, & |T| \leq t_{n-1}(\alpha/2) \end{cases}$$

是检验问题 (5.3.3) 的一个水平为 α 的似然比检验. 这与在 5.2 节中用直观方法求得的检验结果是一致的.

例 5.3.2 问题与例 5.3.1 相同, 求下列检验:

$$H_0: \mu \leq \mu_0 \leftrightarrow H_1: \mu > \mu_0 \quad (5.3.6)$$

的水平为 α 的似然比检验.

解 此时似然函数 $f(\mathbf{x}, \theta)$ 和 θ 与例 5.3.1 中相同, 但

$$\Theta_0 = \{\theta = (\mu, \sigma^2) : \mu \leq \mu_0, \sigma^2 > 0\}.$$

因此, $L_\Theta(\mathbf{x}) = \sup_{\theta \in \Theta} f(\mathbf{x}, \mu, \sigma^2)$ 与例 5.3.1 中完全相同. 但要注意到

$$\begin{aligned} L_{\Theta_0}(\mathbf{x}) &= \sup_{\theta \in \Theta_0} (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \\ &= \sup_{\theta \in \Theta_0} (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right\}. \end{aligned}$$

记 $g(\mu) = \exp \{ -n(\bar{x} - \mu)^2 / (2\sigma^2) \}$. 当 σ^2 固定, $\mu \leq \bar{x}$ 时, $g'(\mu) \geq 0$, 故 $g(\mu)$ 关于 μ 单调增; 当 $\mu \geq \bar{x}$ 时, $g'(\mu) \leq 0$, 故 $g(\mu)$ 关于 μ 单调降. 因此,

(1) 当 $\bar{x} > \mu_0$ 时, 若 H_0 成立, $g(\mu)$ 在 $\mu = \mu_0$ 处达到最大, 故有

$$\min_{\mu \leq \mu_0} \sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \mu_0)^2.$$

(2) 当 $\bar{x} \leq \mu_0$ 时, 若 H_0 成立, $g(\mu)$ 在 $\mu = \bar{x}$ 处达到最大, 故有

$$\min_{\mu \leq \mu_0} \sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2,$$

因此有

$$L_{\Theta_0}(\mathbf{x}) = \begin{cases} L_\Theta(\mathbf{x}), & \bar{x} \leq \mu_0, \\ \left(\frac{2\pi e}{n} \right)^{-n/2} \left(\sum_{i=1}^n (x_i - \mu_0)^2 \right)^{-n/2}, & \bar{x} > \mu_0. \end{cases}$$

故

$$\lambda(\mathbf{x}) = \begin{cases} 1, & \bar{x} \leq \mu_0, \\ \left[\sum_{i=1}^n (x_i - \mu_0)^2 \Big/ \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{\frac{n}{2}} = \left(1 + \frac{1}{n-1} T^2 \right)^{\frac{n}{2}}, & \bar{x} > \mu_0. \end{cases}$$

此处 $T = \sqrt{n}(\bar{X} - \mu_0)/S$, 而 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$. 由于 $\lambda(\mathbf{x})$ 为 T 的严格增函数, 因此似然比检验的否定域为

$$D = \{\mathbf{X} = (X_1, \dots, X_n) : \lambda(\mathbf{X}) > c'\} = \{\mathbf{X} : T > c\}.$$

由于检验水平 α 给定, c 由下式确定:

$$P(T > c \mid \mu = \mu_0) = \alpha.$$

当 $\mu = \mu_0$ 时, $T \sim t_{n-1}$, 故知 $c = t_{n-1}(\alpha)$.

类似在 5.2 节中所述, 上述检验的功效函数 $\beta_\varphi(\mu)$ 是 μ 的单调增函数, 故有

$$\beta_\varphi(\mu) \leq \beta_\varphi(\mu_0) = \alpha, \quad \mu \leq \mu_0.$$

因此

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > t_{n-1}(\alpha), \\ 0, & T(\mathbf{x}) \leq t_{n-1}(\alpha) \end{cases}$$

为检验问题 (5.3.6) 的水平为 α 的似然比检验. 这与在 5.2 节中用直观方法求得的检验结果是一致的.

类似方法可求得检验问题

$$H_0 : \mu \geq \mu_0 \leftrightarrow H_1 : \mu < \mu_0 \quad (5.3.7)$$

的水平为 α 的似然比检验, 这留给读者作为练习.

例 5.3.3 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为自正态分布总体 $N(\mu, \sigma^2)$ 中抽取的随机样本, 考虑如下检验问题的水平为 α 的似然比检验:

$$H_0 : \sigma^2 = \sigma_0^2 \leftrightarrow H_1 : \sigma^2 \neq \sigma_0^2, \quad (5.3.8)$$

其中 σ_0^2 和 α 给定.

解 此时似然函数仍为式 (5.3.4), 参数空间 Θ 如例 5.3.1, 而

$$\Theta_0 = \{\boldsymbol{\theta} = (\mu, \sigma^2) : -\infty < \mu < +\infty, \sigma^2 = \sigma_0^2\}, \quad (5.3.9)$$

$L_{\Theta_0}(\mathbf{x})$ 也与例 5.3.1 相同, 但

$$\begin{aligned} L_{\Theta_0}(\mathbf{x}) &= \sup_{\mu} \left\{ (2\pi\sigma_0^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \right\} \\ &= (2\pi\sigma_0^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}, \end{aligned}$$

因此有

$$\lambda(\mathbf{x}) = \left(\frac{e}{n} \right)^{-n/2} \left[\frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{-n/2} \exp \left\{ \frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}.$$

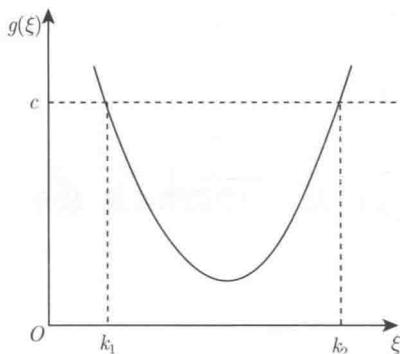


图 5.3.1

令 $\xi = \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x})^2$, $g(\xi) = \xi^{-\frac{n}{2}} e^{\xi/2}$, 则

在 $\xi > 0$ 时 $g(\xi)$ 关于 ξ 先降后升, 且当 $\xi \rightarrow 0$ 和 $\xi \rightarrow \infty$ 时, $g(\xi)$ 的极限皆为 $+\infty$, 其形状如图 5.3.1 所示. 故似然比检验的接受域为 $\bar{D} = \{X : g(\xi) \leq c\} = \left\{ X : k_1 \leq \sum_{i=1}^n (X_i - \bar{X})^2 / \sigma_0^2 \leq k_2 \right\}$.

由于在 H_0 成立时, $\sum_{i=1}^n (X_i - \bar{X})^2 / \sigma_0^2 \sim \chi_{n-1}^2$, 因而 k_1 和 k_2 为下列方程组的解:

$$\begin{cases} g(k_1) = g(k_2), \\ P(k_1 \leq \xi \leq k_2 | H_0) = 1 - \alpha \end{cases} \iff \begin{cases} k_1^{n/2} e^{-k_1/2} = k_2^{n/2} e^{-k_2/2}, \\ P(\xi < k_1 | H_0) + P(\xi > k_2 | H_0) = \alpha. \end{cases} \quad (5.3.10)$$

方程 (5.3.10) 的解不易得到, 一般取

$$P(\xi < k_1 | H_0) = \frac{\alpha}{2}, \quad P(\xi > k_2 | H_0) = \frac{\alpha}{2},$$

得到

$$k_1 = \chi_{n-1}^2(1 - \alpha/2), \quad k_2 = \chi_{n-1}^2(\alpha/2).$$

因此检验问题 (5.3.8) 的水平为 α 的似然比检验是

$$\varphi(x) = \begin{cases} 0, & \chi_{n-1}^2(1 - \alpha/2) < \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x})^2 < \chi_{n-1}^2(\alpha/2), \\ 1, & \text{其他.} \end{cases}$$

这与在 5.2 节中用直观方法求得的单个正态总体方差的检验结果是一致的.

单边检验问题 $H_0 : \sigma^2 \leq \sigma_0^2 \leftrightarrow H_1 : \sigma^2 > \sigma_0^2$ 的水平为 α 的似然比检验留给读者作为练习.

关于两样本正态总体均值差和方差比的似然比检验的方法与前面的例子相同, 只是表达要复杂一些, 已将其放到习题中, 供读者练习.

例 5.3.4 设 $X = (X_1, \dots, X_n)$ 为自均匀分布总体 $U(0, \theta)$ 中抽取的随机样本, 求

$$H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0 \quad (5.3.11)$$

水平为 α 的似然比检验, 其中 α 和 θ_0 给定.

解 此时似然函数为

$$f(x, \theta) = \begin{cases} \theta^{-n}, & 0 < x_{(n)} < \theta, \\ 0, & \text{其他.} \end{cases}$$

参数空间 $\Theta = (0, \infty)$, $\Theta_0 = (0, \theta_0]$. 由于 $X_{(n)} = \max\{X_1, \dots, X_n\}$ 为 θ 的 MLE, 故有

$$L_\Theta(\mathbf{x}) = \sup_{\theta \in \Theta} f(\mathbf{x}, \theta) = (x_{(n)})^{-n}$$

和

$$L_{\Theta_0}(\mathbf{x}) = \sup_{\theta \in \Theta_0} f(\mathbf{x}, \theta) = \begin{cases} L_\Theta(\mathbf{x}), & 0 < x_{(n)} \leq \theta_0, \\ 0, & x_{(n)} > \theta_0. \end{cases}$$

因此有

$$\lambda(\mathbf{x}) = \begin{cases} 1, & 0 < x_{(n)} \leq \theta_0, \\ \infty, & x_{(n)} > \theta_0. \end{cases}$$

由于 $\lambda(\mathbf{x})$ 为 $T(\mathbf{x}) = x_{(n)}$ 的非降函数, 故检验的否定域为

$$D = \{\mathbf{X} = (X_1, \dots, X_n) : X_{(n)} > c\}.$$

注 5.3.1 为使检验水平等于 α , 将集合 $G = \{\mathbf{x} = (x_1, \dots, x_n) : \lambda(\mathbf{x}) = 1\}$ 分为两部分 $G_1 = \{\mathbf{x} : c < x_{(n)} \leq \theta_0\}$, $G_2 = G - G_1$, 则 $G_1 \cup \{\mathbf{x} : \lambda(\mathbf{x}) = \infty\} = \{\mathbf{x} : x_{(n)} > c\}$ 为检验问题 (5.3.11) 的否定域是合理的.

由于 $T = X_{(n)}$ 的密度函数为 $g(t) = nt^{n-1}/\theta^n \cdot I_{(0,\theta)}(t)$, 故由

$$\alpha = P(X_{(n)} > c | \theta = \theta_0) = \int_c^{\theta_0} \frac{nt^{n-1}}{\theta_0^n} dt = 1 - \left(\frac{c}{\theta_0}\right)^n,$$

解出 $c = \theta_0 \sqrt[n]{1-\alpha}$, 故否定域为

$$D = \{\mathbf{X} : X_{(n)} > \theta_0 \sqrt[n]{1-\alpha}\}.$$

检验的功效函数为

$$\begin{aligned} \beta_\varphi(\theta) &= P_\theta(X_{(n)} > \theta_0 \sqrt[n]{1-\alpha}) = \int_{\theta_0 \sqrt[n]{1-\alpha}}^\theta \frac{nt^{n-1}}{\theta^n} dt \\ &= \frac{1}{\theta^n} [\theta^n - \theta_0^n (1-\alpha)] = 1 - (1-\alpha) \left(\frac{\theta_0}{\theta}\right)^n. \end{aligned}$$

它是 θ 的单调增函数, 故有

$$\beta_\varphi(\theta) \leq \beta_\varphi(\theta_0), \quad \theta \leq \theta_0,$$

因此以 D 为否定域的检验水平为 α . 因此

$$\varphi(\mathbf{x}) = \begin{cases} 1, & X_{(n)} > \theta_0 \sqrt[n]{1-\alpha}, \\ 0, & X_{(n)} \leq \theta_0 \sqrt[n]{1-\alpha} \end{cases}$$

为检验问题 (5.3.11) 的水平为 α 的似然比检验.

例 5.3.5 设样本 $\mathbf{X} = (X_1, \dots, X_n)$ 取自下列指数分布总体, 其密度函数为

$$f(x, \lambda) = \frac{1}{\lambda} \exp\left\{-\frac{x}{\lambda}\right\} I_{(0, \infty)}(x).$$

求检验问题

$$H_0 : \lambda = \lambda_0 \leftrightarrow H_1 : \lambda \neq \lambda_0 \quad (5.3.12)$$

的检验水平为 α 的似然比检验, 此处 λ_0 和 α 给定.

解 参数 λ 的似然函数为

$$L(\lambda, \mathbf{x}) = \begin{cases} \lambda^{-n} \exp\{-n\bar{x}/\lambda\}, & \text{当 } x_1, \dots, x_n > 0, \\ 0, & \text{其他,} \end{cases}$$

参数空间和 H_0 对应的参数空间的子集分别为 $\Theta = (0, \infty)$ 和 $\Theta_0 = \{\lambda : \lambda = \lambda_0\}$. 由于 λ 的极大似然估计为 $\hat{\lambda} = \bar{X}$, 故在 Θ 和 Θ_0 上似然函数的最大值分别为

$$L_\Theta(\mathbf{x}) = \sup_{\lambda \in \Theta} L(\lambda, \mathbf{x}) = e^{-n}/\bar{x}^n,$$

$$L_{\Theta_0}(\mathbf{x}) = L(\lambda_0, \mathbf{x}) = \lambda_0^{-n} \exp\{-n\bar{x}/\lambda_0\}.$$

记 $T = \sum_{i=1}^n x_i = n\bar{x}$, 则似然比为

$$\lambda(T) = \frac{L_\Theta(\mathbf{x})}{L_{\Theta_0}(\mathbf{x})} = \frac{n^n \lambda_0^n}{e^n T^n} \exp\{T/\lambda_0\} = c \cdot g(T).$$

此处 $c = (n\lambda_0/e)^n$, $g(T) = T^{-n} e^{T/\lambda_0}$. 显见当 $T \rightarrow \infty$ 和 $T \rightarrow 0$ 时 $g(T) \rightarrow \infty$, $g(t)$ 的形状与图 5.3.1 类似. 因此有

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{当 } \lambda(T) > c \\ 0, & \text{当 } \lambda(T) \leq c \end{cases} = \begin{cases} 1, & \text{当 } T < k_1 \text{ 或 } T > k_2, \\ 0, & \text{其他.} \end{cases}$$

由推论 2.4.5 可知, 当 H_0 成立时, $2T/\lambda_0 \sim \chi_{2n}^2$. 为确定临界值 k_1 和 k_2 , 令

$$P(T < k_1 | H_0) = P\left(\frac{2T}{\lambda_0} < \frac{2k_1}{\lambda_0}\right) = \frac{\alpha}{2},$$

$$P(T > k_2 | H_0) = P\left(\frac{2T}{\lambda_0} > \frac{2k_2}{\lambda_0}\right) = \frac{\alpha}{2},$$

得到

$$k_1 = \frac{\lambda_0}{2} \chi_{2n}^2 \left(1 - \frac{\alpha}{2}\right), \quad k_2 = \frac{\lambda_0}{2} \chi_{2n}^2 \left(\frac{\alpha}{2}\right).$$

因此检验问题 (5.3.12) 的检验水平为 α 的似然比检验为

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{当 } \lambda(T) < \frac{\lambda_0}{2} \chi_{2n}^2 \left(1 - \frac{\alpha}{2}\right) \text{ 或 } \lambda > \frac{\lambda_0}{2} \chi_{2n}^2 \left(\frac{\alpha}{2}\right), \\ 0, & \text{其他.} \end{cases}$$

*5.3.3 似然比的渐近分布

在似然比检验的定义 5.3.1 中, 为了确定式 (5.3.2) 中的 c 和 r , 就需要知道似然比 $\lambda(\mathbf{X})$ 在零假设成立时的分布。在简单的例子中, 如本节第 2 部分的几个例子中, 似然比的精确分布可以求得。但在许多情况下, 似然比有很多复杂的形状, 其精确分布无法求得。1938 年, S. S. Wilks 证明了: 若 X_1, \dots, X_n 是简单随机样本, 则当 $n \rightarrow \infty$ 时, 在零假设成立之下, 似然比有一个简单的极限分布。利用它的极限分布可近似决定式 (5.3.2) 中的 c 和 r 。

Wilks 定理的确切陈述需要陈述一大堆关于总体概率分布的假定, 其证明也很复杂。略去这些陈述, 只强调其中一个至关重要之点, 即要求参数空间 Θ 的维数要高于零假设成立时的 Θ_0 的维数, 如样本 X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma^2)$, $H_0: \mu = \mu_0 \leftrightarrow H_1: \mu \neq \mu_0$, 则 $\Theta = \{\theta = (\mu, \sigma^2) : -\infty < \mu < +\infty, \sigma^2 > 0\}$ 是 R_2 中的上半平面, Θ 的维数是 2; 而 $\Theta_0 = \{\theta = (\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}$, 它是 Θ 中的一条直线, 其维数为 1。因此此例中 Θ 维数高于 Θ_0 的维数。又如球体是三维集, 空间的一个点是零维集。明确了这一点, Wilks 的定理大致可表达为

定理 5.3.1 设 Θ 的维数为 k , Θ_0 的维数为 s , 若 $k - s = t > 0$, 且样本的概率分布满足一定的正则条件, 则对检验问题 (5.1.1), 在零假设 H_0 成立之下, 当样本大小 $n \rightarrow \infty$ 时有

$$2 \log \lambda(\mathbf{X}) \xrightarrow{\mathcal{L}} \chi_t^2.$$

定理的详细陈述及证明见文献 [1] 第 326 页。还有一点需要明确: 零假设 Θ_0 中可以包含不止一个点, 这时定理 5.3.1 的含义是: 不论真参数落在 Θ_0 中何处, $2 \log \lambda(\mathbf{X})$ 的极限分布总是自由度为 t 的 χ^2 分布。

例 5.3.6 设样本 $\mathbf{X}_i = (X_{i1}, \dots, X_{in_i})$ 为从正态总体 $N(\mu_i, \sigma_i^2)$, $i = 1, \dots, m$ 中抽取的简单样本, 且全部样本独立。要检验假设

$$H_0: \sigma_1^2 = \dots = \sigma_m^2 \leftrightarrow H_1: \sigma_1^2, \dots, \sigma_m^2 \text{ 不完全相同}.$$

解 记 $\boldsymbol{\theta} = (\mu_1, \dots, \mu_m; \sigma_1^2, \dots, \sigma_m^2)$, 易见 $\boldsymbol{\theta}$ 的似然函数为

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_m) &= \prod_{i=1}^m \left[(2\pi\sigma_i^2)^{-\frac{n_i}{2}} \exp \left\{ -\frac{1}{2\sigma_i^2} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \right\} \right] \\ &= \prod_{i=1}^m \left[(2\pi\sigma_i^2)^{-\frac{n_i}{2}} \exp \left\{ -\frac{1}{2\sigma_i^2} \sum_{j=1}^{n_i} [(x_{ij} - \bar{x}_i)^2 + (\bar{x}_i - \mu)^2] \right\} \right] \end{aligned}$$

参数空间为 $\Theta = \{\boldsymbol{\theta} = (\mu_1, \dots, \mu_m; \sigma_1^2, \dots, \sigma_m^2) : \mu_i \in R_1, \sigma_i^2 > 0\}$, 其维数为 $k = 2m$ 。零假设对应的 $\boldsymbol{\theta}$ 的子空间为: $\Theta_0 = \{\boldsymbol{\theta} = (\mu_1, \dots, \mu_m; \sigma_1^2, \dots, \sigma_m^2) : \mu_i \in$

$R_1, \sigma_1^2 = \dots = \sigma_m^2 > 0\}$, 其维数为 $s = m + 1$. 记

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \quad S_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2, \quad S^2 = \frac{1}{n} \sum_{i=1}^m n_i S_i^2, \quad (5.3.13)$$

此处 $n = \sum_{i=1}^m n_i$. 易见, μ_i 和 σ_i^2 的 MLE 分别为 $\hat{\mu}_i = \bar{X}_i$, $\hat{\sigma}_i^2 = S_i^2$, $i = 1, \dots, m$. 故

$$L_{\Theta}(\mathbf{x}_1, \dots, \mathbf{x}_m) = \sup_{\theta \in \Theta} L(\theta; \mathbf{x}_1, \dots, \mathbf{x}_m) = (2\pi e)^{-\frac{n}{2}} \prod_{i=1}^m S_i^{-n_i}.$$

当 $\sigma_1^2 = \dots = \sigma_m^2 = \sigma^2$ 时, σ^2 的 MLE 为 $\hat{\sigma}^2 = S^2$, 故

$$L_{\Theta_0}(\mathbf{x}_1, \dots, \mathbf{x}_m) = \sup_{\theta \in \Theta_0} L(\theta; \mathbf{x}_1, \dots, \mathbf{x}_m) = (2\pi e)^{-\frac{n}{2}} S^{-n}.$$

因此不难算出

$$\lambda(\mathbf{x}_1, \dots, \mathbf{x}_m) = \frac{L_{\Theta}(\mathbf{x}_1, \dots, \mathbf{x}_m)}{L_{\Theta_0}(\mathbf{x}_1, \dots, \mathbf{x}_m)} = S^n / \left(\prod_{i=1}^m S_i^{n_i} \right).$$

令

$$Y_n \equiv 2 \log \lambda(\mathbf{x}_1, \dots, \mathbf{x}_m) = n \log S^2 - \sum_{i=1}^n n_i \log S_i^2.$$

由定理 5.3.1 可知, 当 H_0 成立且当 $\min\{n_1, n_2, \dots, n_m\} \rightarrow \infty$ 时, 有

$$Y_n \xrightarrow{\mathcal{L}} \chi_t^2 = \chi_{m-1}^2,$$

其中 $t = k - s = 2m - (m + 1) = m - 1$. 由此得到大样本检验有水平近似为 α 的否定域

$$D = \{(\mathbf{X}_1, \dots, \mathbf{X}_m) : Y_n > \chi_{m-1}^2(\alpha)\}.$$

*5.4 一致最优检验与无偏检验

5.4.1 引言及定义

设有分布族 $\{f(x, \theta), \theta \in \Theta\}$, 其中 Θ 为参数空间. 样本 $\mathbf{X} = (X_1, \dots, X_n)$ 为从上述分布族抽取的简单样本, 如 5.1 节所述, 参数 θ 的假设检验问题可以表示成如下的一般形式:

$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1, \quad (5.4.1)$$

其中 Θ_0 为参数空间 Θ 的非空真子集, $\Theta_1 = \Theta - \Theta_0$.

对检验问题 (5.4.1) 可用几种不同方法去检验, 这就产生不同检验的比较问题, 以及在一定准则下寻求“最优”检验的问题. 这与在第 3 章参数估计问题中, 在无偏估计中找一致最小方差估计的问题完全相似. 下面先给出一致最优检验的定义.

定义 5.4.1 设有检验问题 (5.4.1), 令 $0 < \alpha < 1$, 记 Φ_α 为式 (5.4.1) 的一切水平为 α 的检验的集合. 若 $\varphi \in \Phi_\alpha$, 且对任何检验 $\varphi_1 \in \Phi_\alpha$, 有

$$\beta_\varphi(\theta) \geq \beta_{\varphi_1}(\theta), \quad \theta \in \Theta_1, \quad (5.4.2)$$

则称 φ 为式 (5.4.1) 的一个水平为 α 的一致最优检验 (uniformly most powerful test, UMPT). 当 φ 为水平 α 的 UMPT 时, 它在限制第一类错误概率不超过 α 的条件下, 总使犯第二类错误概率达到最小 (即使不犯第二类错误的概率最大). 因此若以错误概率作为衡量检验优劣的唯一度量, 且接受限制第一类错误概率的原则, 则 UMPT 是最好的检验. 不过, UMPT 的存在一般是例外而不常见的. 理由如下: 若 Θ_1 包含不止一个点, 当在其中取两个不同点 θ_1 和 θ_2 时, 为使 $\beta_\varphi(\theta_1)$ 达到最大的那种检验 φ , 不见得同时也能使 $\beta_\varphi(\theta_2)$ 达到最大. 在 Θ_0 和 Θ_1 都只包含一个点时, 一般说来 UMPT 存在. 这就是下面 Neyman-Pearson 引理 (NP 引理) 的内容.

5.4.2 Neyman-Pearson 引理

定理 5.4.1 (NP 基本引理) 设样本 \mathbf{X} 的分布有概率函数 $f(\mathbf{x}, \theta)$, 参数 θ 只有两个可能的值 θ_0 和 θ_1 , 考虑下列检验问题:

$$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta = \theta_1, \quad (5.4.3)$$

则对任给的 $0 < \alpha < 1$ 有

(1) 存在性. 对检验问题 (5.4.3) 必存在一个检验函数 $\varphi(\mathbf{x})$ 及非负常数 c 和 $0 \leq r \leq 1$, 满足条件

(i)

$$\varphi(\mathbf{x}) = \begin{cases} 1, & f(\mathbf{x}, \theta_1)/f(\mathbf{x}, \theta_0) > c, \\ r, & f(\mathbf{x}, \theta_1)/f(\mathbf{x}, \theta_0) = c, \\ 0, & f(\mathbf{x}, \theta_1)/f(\mathbf{x}, \theta_0) < c. \end{cases} \quad (5.4.4)$$

(ii)

$$E_{\theta_0}[\varphi(\mathbf{X})] = \alpha. \quad (5.4.5)$$

(2) 一致最优性. 任何满足式 (5.4.4) 和式 (5.4.5) 的检验 $\varphi(\mathbf{x})$ 是检验问题 (5.4.3) 的 UMPT.

注 5.4.1 (1) 在定理 5.4.1 中, 当样本分布为连续分布时, 式 (5.4.4) 中的随机化是不必要的. 这时取 $r = 0$, 即式 (5.4.4) 变为

$$\varphi(\mathbf{x}) = \begin{cases} 1, & f(\mathbf{x}, \theta_1)/f(\mathbf{x}, \theta_0) > c, \\ 0, & f(\mathbf{x}, \theta_1)/f(\mathbf{x}, \theta_0) \leq c, \end{cases}$$

其中 c 由 $E_{\theta_0}[\varphi(\mathbf{X})] = P(f(\mathbf{X}, \theta_1)/f(\mathbf{X}, \theta_0) > c | H_0) = \alpha$ 来确定.

(2) 从“似然性”的观点去看 NP 基本引理是很清楚的: 对每个样本 \mathbf{X} , θ_1 和 θ_0 的“似然度”分别为 $f(\mathbf{x}, \theta_1)$ 和 $f(\mathbf{x}, \theta_0)$. 比值 $f(\mathbf{x}, \theta_1)/f(\mathbf{x}, \theta_0)$ 越大, 就反映在得到样本 \mathbf{X} 时, θ 越像 θ_1 而非 θ_0 , 这样的样本 \mathbf{X} 就越倾向于否定 “ $H_0 : \theta = \theta_0$ ” 的假设.

证 (1) 先证明存在性. 记随机变量 $f(\mathbf{X}, \theta_1)/f(\mathbf{X}, \theta_0)$ 的分布函数为

$$G(y) = P\left(\frac{f(\mathbf{X}, \theta_1)}{f(\mathbf{X}, \theta_0)} < y\right), \quad -\infty < y < \infty,$$

则 $G(y)$ 具有分布函数的性质: 单调、非降、左连续且 $\lim_{y \rightarrow -\infty} G(y) = 0$, $\lim_{y \rightarrow +\infty} G(y) = 1$. 从而由 $0 < \alpha < 1$ 和 $G(y)$ 的单调性可知: 必存在 c , 使得

$$G(c) \leq 1 - \alpha \leq G(c + 0).$$

如何确定 r , 分下列三种情形讨论:

(i) 若 $G(c) = 1 - \alpha$, 则取 $r = 1$, 这时由式 (5.4.4) 确定的 $\varphi(\mathbf{x})$ 满足

$$\begin{aligned} E_{\theta_0}[\varphi(\mathbf{X})] &= P_{\theta_0}\left(\frac{f(\mathbf{X}, \theta_1)}{f(\mathbf{X}, \theta_0)} \geq c\right) \\ &= 1 - P_{\theta_0}\left(\frac{f(\mathbf{X}, \theta_1)}{f(\mathbf{X}, \theta_0)} < c\right) = 1 - G(c) = \alpha. \end{aligned}$$

(ii) 若 $G(c + 0) = 1 - \alpha$, 则取 $r = 0$, 此时由式 (5.4.4) 定义的 $\varphi(\mathbf{x})$ 满足

$$\begin{aligned} E_{\theta_0}[\varphi(\mathbf{X})] &= 1 - P(f(\mathbf{X}, \theta_1)/f(\mathbf{X}, \theta_0) \leq c) \\ &= 1 - [P(f(\mathbf{X}, \theta_1)/f(\mathbf{X}, \theta_0) < c) + P(f(\mathbf{X}, \theta_1)/f(\mathbf{X}, \theta_0) = c)] \\ &= 1 - [G(c) + (G(c + 0) - G(c))] = 1 - G(c + 0) = \alpha. \end{aligned}$$

(iii) 若 $G(c) < 1 - \alpha < G(c + 0)$, 则取 $r = [G(c + 0) - (1 - \alpha)]/[G(c + 0) - G(c)]$, 显然, 此时对由式 (5.4.4) 定义的 $\varphi(\mathbf{x})$, 有

$$\begin{aligned} E_{\theta_0}[\varphi(\mathbf{X})] &= P(f(\mathbf{X}, \theta_1)/f(\mathbf{X}, \theta_0) > c) + r \cdot P(f(\mathbf{X}, \theta_1)/f(\mathbf{X}, \theta_0) = c) \\ &= 1 - G(c) - (G(c + 0) - G(c)) + \frac{G(c + 0) - (1 - \alpha)}{G(c + 0) - G(c)} \cdot (G(c + 0) - G(c)) \\ &= 1 - (1 - \alpha) = \alpha. \end{aligned}$$

故存在性证毕.

(2) 再证由式 (5.4.4) 和式 (5.4.5) 定义的 $\varphi(\mathbf{x})$ 具有 UMP 性质. 设 $\varphi_1(\mathbf{x})$ 为检验问题 (5.4.3) 的任一水平为 α 的检验, 要证明 $E_{\theta_1}[\varphi(\mathbf{X})] \geq E_{\theta_1}[\varphi_1(\mathbf{X})]$. 为此定义样本空间 \mathcal{X} 上的子集

$$S^+ = \{\mathbf{x} : \varphi(\mathbf{x}) > \varphi_1(\mathbf{x})\}, \quad S^- = \{\mathbf{x} : \varphi(\mathbf{x}) < \varphi_1(\mathbf{x})\},$$

则在 S^+ 上有 $\varphi(\mathbf{x}) > \varphi_1(\mathbf{x}) \geq 0$, 因此有 $\varphi(\mathbf{x}) > 0$, 故由式 (5.4.4) 可知此时

$$\frac{f(\mathbf{x}, \theta_1)}{f(\mathbf{x}, \theta_0)} \geq c \iff f(\mathbf{x}, \theta_1) - cf(\mathbf{x}, \theta_0) \geq 0.$$

当 $\mathbf{x} \in S^-$ 时有 $\varphi(\mathbf{x}) < \varphi_1(\mathbf{x}) \leq 1$, 因此有 $\varphi(\mathbf{x}) < 1$, 故由式 (5.4.4) 可知此时

$$\frac{f(\mathbf{x}, \theta_1)}{f(\mathbf{x}, \theta_0)} \leq c \iff f(\mathbf{x}, \theta_1) - cf(\mathbf{x}, \theta_0) \leq 0.$$

故在 $S = S^+ \cup S^-$ 上必有

$$(\varphi(\mathbf{x}) - \varphi_1(\mathbf{x}))(f(\mathbf{x}, \theta_1) - cf(\mathbf{x}, \theta_0)) \geq 0$$

(因为在 S^+ 上两因子皆非负, 在 S^- 上两因子皆非正). 因此

$$\begin{aligned} & \int_{\mathcal{X}} (\varphi(\mathbf{x}) - \varphi_1(\mathbf{x}))(f(\mathbf{x}, \theta_1) - cf(\mathbf{x}, \theta_0)) d\mathbf{x} \\ &= \int_{S^+ \cup S^-} (\varphi(\mathbf{x}) - \varphi_1(\mathbf{x}))(f(\mathbf{x}, \theta_1) - cf(\mathbf{x}, \theta_0)) d\mathbf{x} \geq 0, \end{aligned}$$

即

$$\begin{aligned} & \int_{\mathcal{X}} \varphi(\mathbf{x}) f(\mathbf{x}, \theta_1) d\mathbf{x} - \int_{\mathcal{X}} \varphi_1(\mathbf{x}) f(\mathbf{x}, \theta_1) d\mathbf{x} \\ & \geq c \left[\int_{\mathcal{X}} \varphi(\mathbf{x}) f(\mathbf{x}, \theta_0) d\mathbf{x} - \int_{\mathcal{X}} \varphi_1(\mathbf{x}) f(\mathbf{x}, \theta_0) d\mathbf{x} \right]. \end{aligned} \quad (5.4.6)$$

由式 (5.4.5) 知 $E_{\theta_0}[\varphi(\mathbf{X})] = \int_{\mathcal{X}} \varphi(\mathbf{x}) f(\mathbf{x}, \theta_0) d\mathbf{x} = \alpha$, 而 $\varphi_1(\mathbf{x})$ 是检验问题 (5.4.3) 的水平为 α 的任一检验, 即 $E_{\theta_0}[\varphi_1(\mathbf{X})] \leq \alpha$, 故知式 (5.4.6) 右边非负, 从而左边也非负. 因此有

$$\beta_{\varphi}(\theta_1) = \int_{\mathcal{X}} \varphi(\mathbf{x}) f(\mathbf{x}, \theta_1) d\mathbf{x} \geq \int_{\mathcal{X}} \varphi_1(\mathbf{x}) f(\mathbf{x}, \theta_1) d\mathbf{x} = \beta_{\varphi_1}(\theta_1).$$

这就证明了 $\varphi(\mathbf{x})$ 为式 (5.4.3) 的水平为 α 的 UMPT. 定理证毕.

例 5.4.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为自正态总体 $N(\mu, 1)$ 中抽取的随机样本, 其中 μ 为未知参数, 求假设检验问题

$$H_0 : \mu = 0 \leftrightarrow H_1 : \mu = \mu_1 (\mu_1 > 0)$$

的水平为 α 的 UMPT, 其中 μ_1 和 α 给定.

解 由 NP 引理, 先求 $f_0(\mathbf{x})$ 和 $f_1(\mathbf{x})$ 的表达式

$$\begin{aligned} f_0(\mathbf{x}) &= (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n x_i^2 \right\}, \\ f_1(\mathbf{x}) &= (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \mu_1)^2 \right\}. \end{aligned}$$

似然比可表示为

$$\lambda(\mathbf{x}) = \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} = \exp \left\{ -\frac{1}{2} n \mu_1^2 + n \mu_1 \bar{x} \right\}.$$

显然当 $\mu_1 > 0$ 时, $\lambda(\mathbf{x})$ 为 \bar{x} 的严格增函数, 故 UMPT 的否定域为

$$D = \{\mathbf{X} : \lambda(\mathbf{X}) > c'\} = \{\mathbf{X} : \sqrt{n}\bar{X} > c\},$$

当 H_0 成立时, $U = \sqrt{n}\bar{X} \sim N(0, 1)$, 故由 NP 引理可知

$$E_0[\varphi(\mathbf{X})] = P(\sqrt{n}\bar{X} > c | H_0) = \alpha,$$

显然 $c = u_\alpha$. 因此检验水平为 α 的 UMPT 的检验函数为

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \bar{x} > u_\alpha / \sqrt{n}, \\ 0, & \bar{x} \leq u_\alpha / \sqrt{n}. \end{cases}$$

可见 $\varphi(\mathbf{x})$ 与 μ_1 无关, 因此上述检验函数 $\varphi(\mathbf{x})$ 也是检验问题

$$H_0 : \mu = 0 \leftrightarrow H'_1 : \mu > 0$$

的水平为 α 的 UMPT.

注 5.4.2 此例告诉我们: 在某些情况下, 如果由 NP 引理得到的 UMPT 不依赖于对立假设的具体值, 则可由此得到一个对立假设是复合假设, 即 $H_0 : \mu = 0 \leftrightarrow H'_1 : \mu > 0$ 的水平为 α 的 UMPT.

类似本例可以求得检验问题 $H_0 : \mu = 0 \leftrightarrow H''_1 : \mu < 0$ 的检验水平为 α 的 UMPT, 具体的推导留给读者作为练习.

例 5.4.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从两点分布 $b(1, p)$ 中抽取的随机样本, 其中 p 为未知参数. 求检验问题

$$H_0 : p = p_0 \leftrightarrow H_1 : p = p_1 \quad (p_1 > p_0)$$

的水平为 α 的 UMPT, 其中 p_0, p_1 和 α 给定.

解 由 NP 引理, 先求 f_0 和 f_1 的表达式

$$\begin{aligned} f_0(\mathbf{x}) &= p(\mathbf{x}, p_0) = p_0^{\sum_{i=1}^n x_i} (1-p_0)^{n - \sum_{i=1}^n x_i}, \\ f_1(\mathbf{x}) &= p(\mathbf{x}, p_1) = p_1^{\sum_{i=1}^n x_i} (1-p_1)^{n - \sum_{i=1}^n x_i}. \end{aligned}$$

记 $T(\mathbf{x}) = \sum_{i=1}^n x_i$, 似然比

$$\lambda(\mathbf{x}) = \frac{p(\mathbf{x}, p_1)}{p(\mathbf{x}, p_0)} = \left(\frac{1-p_1}{1-p_0} \right)^n \left[\frac{p_1(1-p_0)}{p_0(1-p_1)} \right]^{T(\mathbf{x})}.$$

由于 $p_1 > p_0, 1-p_0 > 1-p_1$, 所以 $p_1(1-p_0)/p_0(1-p_1) > 1$, 故 $\lambda(\mathbf{x})$ 关于 $T(\mathbf{x})$ 严格单调增. 由于 r.v. $T(\mathbf{X})$ 服从离散型分布, 故需要随机化. 由 NP 引理可知检验函数为

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > c, \\ r, & T(\mathbf{x}) = c, \\ 0, & T(\mathbf{x}) < c. \end{cases}$$

当 H_0 成立时 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 服从二项分布 $b(n, p_0)$, 当 α 给定时, c 由下列不等式确定:

$$\alpha_1 = \sum_{k=c+1}^n \binom{n}{k} p_0^k (1-p_0)^{n-k} \leq \alpha \leq \sum_{k=c}^n \binom{n}{k} p_0^k (1-p_0)^{n-k}.$$

取

$$r = \frac{\alpha - \alpha_1}{\binom{n}{c} p_0^c (1-p_0)^{n-c}},$$

则必有

$$E_{p_0}[\varphi(\mathbf{X})] = P_{p_0}(T(\mathbf{X}) > c) + r \cdot P_{p_0}(T(\mathbf{X}) = c) = \alpha.$$

因此 $\varphi(\mathbf{x})$ 为水平为 α 的 UMPT.

由于上述检验函数 $\varphi(\mathbf{x})$ 与 p_1 无关, 故它也是检验问题

$$H_0 : p = p_0 \leftrightarrow H'_1 : p > p_0$$

的水平为 α 的 UMPT.

注 5.4.3 关于随机化检验问题. 本例中当出现 $T(\mathbf{x}) = \sum_{i=1}^n x_i = c$ 时, 先做一个具有成功率为 r 的 Bernoulli 试验. 若该试验成功, 则否定 H_0 ; 若不然, 则接受 H_0 . 例如, $r = 1/2$, 则可通过掷一均匀硬币, 规定出现正面为成功. 若掷出正面, 则否定 H_0 ; 不然, 则接受 H_0 .

如在 5.1 节中所述, 对随机化检验分两步走: ① 首先通过试验获得样本观察; ② 有了样本后, 当样本出现特殊值 (如本例中 $\sum_{i=1}^n x_i = c$) 需随机化时再做一次试验. 试验结果为 A 或 \bar{A} , 成功概率为 $P(A) = r$. 若 A 发生, 则拒绝 H_0 ; 否则接受 H_0 .

例 5.4.3 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是来自均匀分布 $U(0, \theta)$ 的随机样本, 其中 $\theta > 0$ 为未知参数. 求下列检验问题

$$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta = \theta_1 (\theta_1 > \theta_0 > 0)$$

的水平为 α 的 UMPT.

解 服从均匀分布的样本 \mathbf{X} 的密度函数和似然比分别为

$$f(\mathbf{x}, \theta) = \frac{1}{\theta^n} I_{(0, \theta)}(x_{(n)}),$$

$$\lambda(\mathbf{x}) = \frac{f(\mathbf{x}, \theta_1)}{f(\mathbf{x}, \theta_0)} = \begin{cases} \left(\frac{\theta_0}{\theta_1}\right)^n, & 0 < x_{(n)} < \theta_0, \\ \infty, & \theta_0 < x_{(n)} < \infty. \end{cases}$$

此处定义了 $0/0 = \infty$, 因 $\lambda(\mathbf{x})$ 关于 $T(\mathbf{x}) = x_{(n)}$ 非降, 故由 NP 引理, 可知水平为 α 的 UMPT 函数有形式

$$\varphi(\mathbf{x}) = \begin{cases} 1, & x_{(n)} > c, \\ 0, & x_{(n)} \leq c. \end{cases}$$

$T = X_{(n)}$ 的密度函数为

$$g_\theta(t) = \frac{nt^{n-1}}{\theta^n} I_{(0, \theta)}(t).$$

故当 H_0 成立时, $T(\mathbf{X})$ 的密度函数为 $g_{\theta_0}(t) = [nt^{n-1}/\theta_0^n] \cdot I_{(0, \theta_0)}(t)$, 因此有

$$E_{\theta_0}[\varphi(\mathbf{X})] = \int_0^\infty \varphi(t) g_{\theta_0}(t) dt = \int_c^{\theta_0} \frac{nt^{n-1}}{\theta_0^n} dt = 1 - \frac{c^n}{\theta_0^n} = \alpha,$$

故得 $c = \theta_0 \sqrt[n]{1 - \alpha}$, 因此

$$\varphi(\mathbf{x}) = \begin{cases} 1, & x_{(n)} > \theta_0 \sqrt[n]{1 - \alpha}, \\ 0, & x_{(n)} \leq \theta_0 \sqrt[n]{1 - \alpha} \end{cases}$$

为一个水平为 α 的 UMPT.

由于此检验 $\varphi(\mathbf{x})$ 与 θ_1 无关, 故它也是

$$H_0 : \theta = \theta_0 \leftrightarrow H'_1 : \theta > \theta_0$$

的水平为 α 的 UMPT.

注 5.4.4 由上面三个例子可见 UMPT 检验函数 $\varphi(\mathbf{x})$ 皆为充分统计量的函数, 这是否具有普遍意义呢? 有下列结论:

设 r.v. X 的密度函数为 $f(x, \theta)$, $\theta \in \Theta$ 为未知参数, $\mathbf{X} = (X_1, \dots, X_n)$ 为自总体 X 中抽取的随机样本, $T = T(\mathbf{X})$ 为 θ 的充分统计量, 则由式 (5.4.4) 和式 (5.4.5) 定义的检验函数 $\varphi(\mathbf{x})$ 是充分统计量 T 的函数.

这一结果的证明并不难, 只要利用充分统计量的因子分解定理即可证得. 详细证明留给读者作为练习.

5.4.3 利用 NP 引理求一致最优检验 (UMPT)

NP 引理的作用主要不在于求像检验问题 (5.4.3) 那样的 UMPT, 因为实际应用中像式 (5.4.3) 那样的检验问题是不常见的. 一般情形是零假设和对立假设都是复合的情形. NP 引理的主要作用是在于它是求更复杂情形下 UMPT 的工具. 在前面的例 5.4.1– 例 5.4.3 这 3 个例子中已经将检验问题推广到对立假设是复合的情形. 更一般的假设检验问题如式 (5.4.1) 所示, 即 $H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1$, 其中 Θ_0 和 Θ_1 皆为复合情形 (即其中包含参数空间 Θ 中的点不止一个). 寻找这类检验问题的 UMPT 的一般想法是: 在 Θ_0 中挑一个 θ_0 尽可能与 Θ_1 接近, 再在 Θ_1 中挑一个 θ_1 , 用 NP 引理作出如式 (5.4.4) 和式 (5.4.5) 的 UMPT φ_{θ_1} . 一般当 θ_1 在 Θ_1 中变动时, φ_{θ_1} 不随 θ_1 的变化而变化, 即不论 θ_1 在 Θ_1 中如何变化, $\varphi_{\theta_1} = \varphi$ 与 θ_1 无关, 则 φ 也是 $H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \in \Theta_1$ 的 UMPT. 因此, 更进一步若能证明: 此检验对任何 $\theta \in \Theta_0$ 皆有检验水平 α , 则 φ 也是 $H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1$ 的水平为 α 的 UMPT.

此法要行得通也不容易. 只有在参数空间为一维欧氏空间 R_1 或其一子区间, 而检验的假设是单边的, 即为 $H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0$ 或者 $H_0 : \theta \geq \theta_0 \leftrightarrow H_1 : \theta < \theta_0$ 且对样本分布有一定要求时, 上述方法才可行. 特别当样本分布为指数族分布时, 上述两类单边检验的 UMPT 是存在的. 下面就来讨论之.

设样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的分布族为下列指数族:

$$f(\mathbf{x}, \theta) = c(\theta) \exp \{Q(\theta)T(\mathbf{x})\} h(\mathbf{x}), \quad (5.4.7)$$

其中 $c(\theta) > 0$ 和 $Q(\theta)$ 为 θ 的函数, $T(\mathbf{x})$ 和 $h(\mathbf{x})$ 是样本 \mathbf{x} 的函数.

对如下单边检验问题:

$$H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0, \quad (5.4.8)$$

有下列重要结论.

定理 5.4.2 设样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的分布为指数族 (5.4.7), 参数空间 Θ 为 $R_1 = (-\infty, +\infty)$ 的一有限或无限区间, θ_0 为 Θ 的一个内点且 $Q(\theta)$ 为 θ 的严格增函数, 则检验问题 (5.4.8) 的水平为 α 的 UMPT 存在 ($0 < \alpha < 1$), 且有形式

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > c, \\ r, & T(\mathbf{x}) = c, \\ 0, & T(\mathbf{x}) < c, \end{cases} \quad (5.4.9)$$

其中 c 和 r ($0 \leq r \leq 1$) 满足条件

$$E_{\theta_0}[\varphi(\mathbf{X})] = P_{\theta_0}(T(\mathbf{X}) > c) + r \cdot P_{\theta_0}(T(\mathbf{X}) = c) = \alpha. \quad (5.4.10)$$

证 任取 $\theta_1 > \theta_0$, 首先考虑检验问题

$$H'_0 : \theta = \theta_0 \leftrightarrow H'_1 : \theta = \theta_1 \quad (5.4.11)$$

有似然比

$$\lambda(\mathbf{x}) = \frac{f(\mathbf{x}, \theta_1)}{f(\mathbf{x}, \theta_0)} = \frac{c(\theta_1)}{c(\theta_0)} \exp \{(Q(\theta_1) - Q(\theta_0))T(\mathbf{x})\}.$$

由于 $Q(\theta_1) - Q(\theta_0) > 0$, $c(\theta_1)/c(\theta_0) > 0$, 上式右边为 $T(x)$ 的严格增函数. 因此由 NP 引理可知检验问题 (5.4.11) 的水平为 α 的 UMPT 检验函数为

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \lambda(\mathbf{x}) > c', \\ r, & \lambda(\mathbf{x}) = c', \\ 0, & \lambda(\mathbf{x}) < c' \end{cases} \iff \varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > c, \\ r, & T(\mathbf{x}) = c, \\ 0, & T(\mathbf{x}) < c, \end{cases}$$

其中常数 c 和 r 满足下式:

$$E_{\theta_0}[\varphi(\mathbf{X})] = P_{\theta_0}(T(\mathbf{X}) > c) + rP_{\theta_0}(T(\mathbf{X}) = c) = \alpha.$$

由于 c 和 r 与 θ_1 无关, 故由式 (5.4.9) 和式 (5.4.10) 确定的检验函数 $\varphi(\mathbf{x})$ 也是下述检验问题:

$$H'_0 : \theta = \theta_0 \leftrightarrow H'_1 : \theta > \theta_0$$

的水平为 α 的 UMPT.

只要证明 $\varphi(\mathbf{x})$ 作为检验问题 (5.4.8) 的检验, 具有水平 α , 即可完成证明. 为此只需证明 $\varphi(\mathbf{x})$ 的功效函数 $\beta_\varphi(\theta)$ 是 θ 的单调增函数即可. 下面来证明这一事实.

任取 $\theta' < \theta''$, 由式 (5.4.7) 可知

$$\frac{f(\mathbf{x}, \theta'')}{f(\mathbf{x}, \theta')} = \frac{c(\theta'')}{c(\theta')} \exp \{(Q(\theta'') - Q(\theta'))T(\mathbf{x})\}.$$

因为 $Q(\theta)$ 为 θ 的严格增函数且 $\theta' < \theta''$, 故有 $Q(\theta'') - Q(\theta') > 0$, 又 $c(\theta'')/c(\theta') > 0$, 因此 $f(\mathbf{x}, \theta'')/f(\mathbf{x}, \theta')$ 只与 $T(\mathbf{x})$ 有关, 且为 $T(\mathbf{x})$ 的严格增函数. 找 t_0 , 使

$$\frac{c(\theta'')}{c(\theta')} \exp \{(Q(\theta'') - Q(\theta')) t_0\} = 1,$$

这样的 t_0 必存在, 否则恒有 $f(\mathbf{x}, \theta'')/f(\mathbf{x}, \theta') < 1$ 或 $f(\mathbf{x}, \theta'')/f(\mathbf{x}, \theta') > 1$, 这与 $\int_{\mathcal{X}} f(\mathbf{x}, \theta'') d\mathbf{x} = \int_{\mathcal{X}} f(\mathbf{x}, \theta') d\mathbf{x} = 1$ 矛盾. 令

$$S_1 = \{\mathbf{x} : T(\mathbf{x}) > t_0\},$$

$$S_2 = \{\mathbf{x} : T(\mathbf{x}) < t_0\},$$

$$S_3 = \{\mathbf{x} : T(\mathbf{x}) = t_0\},$$

则由 $\int_{\mathcal{X}} f(\mathbf{x}, \theta') d\mathbf{x} = \int_{\mathcal{X}} f(\mathbf{x}, \theta'') d\mathbf{x} = 1$ 以及 t_0 的定义, 易知

$$\begin{aligned} & \int_{S_1} (f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')) d\mathbf{x} + \int_{S_2} (f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')) d\mathbf{x} \\ &= \int_{\mathcal{X}} (f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')) d\mathbf{x} = 0, \end{aligned}$$

故有

$$0 \leq \int_{S_1} [f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')] d\mathbf{x} = - \int_{S_2} [f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')] d\mathbf{x}.$$

由式 (5.4.9) 可知 $\varphi(\mathbf{x})$ 只与 $T(\mathbf{x})$ 有关, 且是 $T(\mathbf{x})$ 的非降函数, 故有

$$\inf_{\mathbf{x} \in S_1} \varphi(\mathbf{x}) \geq \sup_{\mathbf{x} \in S_2} \varphi(\mathbf{x}).$$

因此

$$\begin{aligned} \beta_\varphi(\theta'') - \beta_\varphi(\theta') &= \int_{\mathcal{X}} \varphi(\mathbf{x}) [f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')] d\mathbf{x} \\ &= \int_{S_1} \varphi(\mathbf{x}) [f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')] d\mathbf{x} + \int_{S_2} \varphi(\mathbf{x}) [f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')] d\mathbf{x} \\ &= \int_{S_1} \varphi(\mathbf{x}) [f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')] d\mathbf{x} - \int_{S_2} \varphi(\mathbf{x}) [-(f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta'))] d\mathbf{x} \\ &\geq \inf_{\mathbf{x} \in S_1} \varphi(\mathbf{x}) \cdot \int_{S_1} [f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')] d\mathbf{x} \\ &\quad - \sup_{\mathbf{x} \in S_2} \varphi(\mathbf{x}) \cdot \int_{S_2} [-(f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta'))] d\mathbf{x} \\ &= (\inf_{\mathbf{x} \in S_1} \varphi(\mathbf{x}) - \sup_{\mathbf{x} \in S_2} \varphi(\mathbf{x})) \int_{S_1} [f(\mathbf{x}, \theta'') - f(\mathbf{x}, \theta')] d\mathbf{x} \geq 0, \end{aligned}$$

即 $\beta_\varphi(\theta'') \geq \beta_\varphi(\theta')$, 对任给的 $\theta'' > \theta'$ 成立, 这就证明了 $\beta_\varphi(\theta)$ 为 θ 的非降函数, 故

$$\sup_{\theta \leq \theta_0} \beta_\varphi(\theta) \leq \beta_\varphi(\theta_0) = E_{\theta_0}[\varphi(\mathbf{X})] = \alpha.$$

因此由式 (5.4.9) 和式 (5.4.10) 确定的 $\varphi(\mathbf{x})$ 为检验问题 (5.4.8) 的水平为 α 的 UMPT. 定理证毕.

注 5.4.5 (1) 在定理 5.4.2 中若样本分布是连续分布, 则 UMPT 不需要随机化. 故检验问题 (5.4.8) 的水平为 α 的 UMPT, 通过式 (5.4.9) 和式 (5.4.10) 中令 $r = 0$ 获得.

(2) 若在定理 5.4.2 条件中改 “ $Q(\theta)$ 为 θ 的严格增函数” 为 “ $Q(\theta)$ 为 θ 的严格降函数”, 其余不变, 则检验问题 (5.4.8) 的水平为 α 的 UMPT, 需要通过将式 (5.4.9) 和式 (5.4.10) 中的不等号反向 (等号不变) 即可得到.

考虑与式 (5.4.8) 相反的单边检验问题

$$H_0 : \theta \geq \theta_0 \leftrightarrow H_1 : \theta < \theta_0, \quad (5.4.12)$$

关于这一检验问题的水平为 α 的 UMPT 有下列定理.

定理 5.4.3 若定理 5.4.2 的条件成立, 则检验问题 (5.4.12) 的水平为 α 的 UMPT 存在, 且有形式

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) < c, \\ r, & T(\mathbf{x}) = c, \\ 0, & T(\mathbf{x}) > c, \end{cases} \quad (5.4.13)$$

其中 c 和 r ($0 \leq r \leq 1$) 满足条件

$$E_{\theta_0}[\varphi(\mathbf{X})] = P_{\theta_0}(T(\mathbf{X}) < c) + r \cdot P_{\theta_0}(T(\mathbf{X}) = c) = \alpha, \quad (5.4.14)$$

此定理的证明方法与定理 5.4.2 类似, 从略.

注 5.4.6 (1) 在定理 5.4.3 中若样本分布为连续分布, 则 UMPT 不需要随机化, 故检验问题 (5.4.12) 的水平为 α 的 UMPT, 可通过在式 (5.4.13) 和式 (5.4.14) 中令 $r = 0$ 获得.

(2) 在定理 5.4.3 中, 若改 “ $Q(\theta)$ 为 θ 的严格增函数” 为 “ $Q(\theta)$ 为 θ 的严格降函数”, 其余条件不变, 则检验问题 (5.4.12) 的水平为 α 的 UMPT, 需要通过将式 (5.4.13) 和式 (5.4.14) 中的不等号反向 (等号不变), 即可以得到.

例 5.4.4 问题与例 5.4.1 相同, 即设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从正态总体 $N(\theta, 1)$ 中抽取的简单样本, 求检验问题 $H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0$ 的 UMPT, 此处 θ_0 和检验水平 α 给定.

解 正态分布为指数族分布, 样本密度为

$$\begin{aligned} f(x_1, \dots, x_n, \theta) &= (2\pi)^{-\frac{n}{2}} \exp\{-n\theta^2/2\} \exp\{n\theta\bar{x}\} \exp\left\{-\sum_{i=1}^n x_i^2/2\right\} \\ &= c(\theta) \exp\{Q(\theta)T(\mathbf{x})\} h(\mathbf{x}), \end{aligned}$$

其中 $c(\theta) = (2\pi)^{n/2} \exp\{-n\theta^2/2\}$, $h(\mathbf{x}) = \exp\left\{-\sum_{i=1}^n x_i^2/2\right\}$, $T(\mathbf{x}) = \bar{x}$, $Q(\theta) = n\theta$ 为 θ 的严格增函数, 由定理 5.4.2 (由于正态分布为连续分布, 检验函数不需要随机化) 可知水平为 α 的 UMPT 由下式给出:

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > c, \\ 0, & T(\mathbf{x}) \leq c. \end{cases}$$

由于 $T(\mathbf{X}) = \bar{X} \sim N(\theta, 1/n)$, 故 $\sqrt{n}(\bar{X} - \theta) \sim N(0, 1)$, 令

$$\alpha = E_{\theta_0}[\varphi(\mathbf{X})] = P_{\theta_0}(T(\mathbf{X}) > c) = P_{\theta_0}(\sqrt{n}(\bar{X} - \theta_0) > \sqrt{n}(c - \theta_0)),$$

可知 $\sqrt{n}(c - \theta_0) = u_\alpha$, 即 $c = \theta_0 + u_\alpha/\sqrt{n}$. 因 $T(\mathbf{x}) = \bar{x}$, 故水平为 α 的 UMPT 为

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \bar{x} > \theta_0 + u_\alpha/\sqrt{n}, \\ 0, & \bar{x} \leq \theta_0 + u_\alpha/\sqrt{n}. \end{cases}$$

特别取 $\theta_0 = 0$ 就与例 5.4.1 中的检验问题相同, 这是对例 5.4.1 的补充.

例 5.4.5 从一大批产品中抽取 n 个检查其结果, 得样本 $\mathbf{X} = (X_1, \dots, X_n)$, 其中 $X_i = 1$, 若第 i 个产品为废品; 否则为 0, $i = 1, \dots, n$. 求

$$H_0 : p \leq p_0 \leftrightarrow H_1 : p > p_0$$

的水平为 α 的 UMPT, 其中 p_0 和 α 给定.

解 令 $T = T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为 n 个产品中的废品数, 则充分统计量 $T \sim$ 二项分布 $b(n, p)$. 二项分布族为指数族, 其概率分布为

$$f(t, p) = \binom{n}{t} p^t (1-p)^{n-t} = c(p) \exp\{Q(p) \cdot t\} h(t),$$

其中 $c(p) = (1-p)^n$, $t = \sum_{i=1}^n x_i$ 为 T 的观察值, $h(t) = \binom{n}{t}$, $Q(p) = \log(p/(1-p))$ 为 p 的严格单调增函数, 故由定理 5.4.2 可知

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > c, \\ r, & T(\mathbf{x}) = c, \\ 0, & T(\mathbf{x}) < c, \end{cases}$$

其中 c 由下列不等式决定:

$$\alpha_1 = \sum_{i=c+1}^n \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha \leq \sum_{i=c}^n \binom{n}{i} p_0^i (1-p_0)^{n-i}.$$

取 r 为

$$r = \frac{\alpha - \alpha_1}{\binom{n}{t} p_0^c (1-p_0)^{n-c}},$$

则必有

$$E_{p_0}[\varphi(\mathbf{X})] = P_{p_0}(T(\mathbf{X}) > c) + r \cdot P_{p_0}(T(\mathbf{X}) = c) = \alpha,$$

因此上述检验 $\varphi(\mathbf{x})$ 为水平为 α 的 UMPT. 这是对例 5.4.2 的补充.

例 5.4.6 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为自 Poisson 总体 $P(\lambda)$ 中抽取的随机样本, $\lambda > 0$ 为未知参数. 求

$$H_0 : \lambda \leq \lambda_0 \leftrightarrow H_1 : \lambda > \lambda_0$$

的水平为 α 的 UMPT, 其中 λ_0 和 α 给定.

解 Poisson 分布为指数族分布. 样本 \mathbf{X} 的密度函数为

$$f(\mathbf{x}, \lambda) = \frac{\lambda^{T(\mathbf{x})} e^{-n\lambda}}{x_1! \cdots x_n!} = c(\lambda) \exp\{Q(\lambda)T(\mathbf{x})\} h(\mathbf{x}),$$

其中 $c(\lambda) = e^{-n\lambda}$, $T(\mathbf{x}) = \sum_{i=1}^n x_i$, $h(\mathbf{x}) = 1/(x_1! \cdots x_n!)$, $Q(\lambda) = \log \lambda$ 为 λ 的严格增函数, 由定理 5.4.2 可知

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > c, \\ r, & T(\mathbf{x}) = c, \\ 0, & T(\mathbf{x}) < c, \end{cases}$$

其中 c 由下列不等式确定 (注意检验统计量 $T(\mathbf{X}) = \sum_{i=1}^n X_i \sim P(n\lambda)$):

$$\alpha_1 = \sum_{k=c+1}^{\infty} \frac{(n\lambda_0)^k e^{-n\lambda_0}}{k!} \leq \alpha \leq \sum_{k=c}^{\infty} \frac{(n\lambda_0)^k e^{-n\lambda_0}}{k!}.$$

取 r 为

$$r = \frac{(\alpha - \alpha_1)c!}{(n\lambda_0)^c e^{-n\lambda_0}},$$

则必有

$$E_{\lambda_0}[\varphi(\mathbf{X})] = P_{\lambda_0}(T(\mathbf{X}) > c) + r \cdot P_{\lambda_0}(T(\mathbf{X}) = c) = \alpha,$$

故上述检验 $\varphi(\mathbf{x})$ 为水平为 α 的 UMPT.

例 5.4.7 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为自指数分布总体 $\text{Exp}(\lambda)$ 中抽取的随机样本, $\lambda > 0$ 为未知参数. 求

$$H_0 : \lambda \leq \lambda_0 \leftrightarrow H_1 : \lambda > \lambda_0$$

的水平为 α 的 UMPT, 其中 λ_0 和 α 给定.

解 指数分布属于指数族. 样本 \mathbf{X} 的密度函数为

$$f(\mathbf{x}, \lambda) = \lambda^n \exp\left\{-\lambda \sum_{i=1}^n x_i\right\} I_{[x_i > 0, i=1,2,\dots,n]},$$

其中 $c(\lambda) = \lambda^n$, $h(\mathbf{x}) = I_{[x_i > 0, i=1,2,\dots,n]}$, $T(\mathbf{x}) = \sum_{i=1}^n x_i$, $Q(\lambda) = -\lambda$ 为 λ 的单调降函数, 故由注 5.4.5 可知

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) < c, \\ 0, & T(\mathbf{x}) \geq c. \end{cases}$$

由推理 2.4.5 可知 $2\lambda T(\mathbf{X}) \sim \chi_{2n}^2$, 故有

$$\alpha = P_{\lambda_0}(T(\mathbf{X}) < c) = P_{\lambda_0}(2\lambda_0 T(\mathbf{X}) < 2\lambda_0 c),$$

因此 $2\lambda_0 c = \chi_{2n}^2(1 - \alpha)$, 即 $c = \chi_{2n}^2(1 - \alpha)/(2\lambda_0)$. 因此

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) < \frac{1}{2\lambda_0} \chi_{2n}^2(1 - \alpha), \\ 0, & T(\mathbf{x}) \geq \frac{1}{2\lambda_0} \chi_{2n}^2(1 - \alpha) \end{cases}$$

为水平为 α 的 UMPT.

例 5.4.8 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为自正态总体 $N(0, \sigma^2)$ 中抽取的随机样本, σ^2 为未知参数. 求

$$H_0 : \sigma^2 \geq \sigma_0^2 \leftrightarrow H_1 : \sigma^2 < \sigma_0^2$$

的水平为 α 的 UMPT, 其中 σ_0^2 和 α 给定.

解 正态分布 $N(0, \sigma^2)$ 为指数族, 样本 \mathbf{X} 密度函数为

$$f(\mathbf{x}, \sigma^2) = \prod_{i=1}^n f(x_i, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2\right\},$$

其中 $c(\sigma) = (2\pi\sigma^2)^{-n/2}$, $h(\mathbf{x}) \equiv 1$, $T(\mathbf{x}) = \sum_{i=1}^n x_i^2$, $Q(\sigma^2) = -1/(2\sigma^2)$ 为 σ^2 的严格单调增函数, 由定理 5.4.3 可知

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) < c, \\ 0, & T(\mathbf{x}) \geq c. \end{cases}$$

由于 $\sum_{i=1}^n X_i^2 / \sigma^2 \sim \chi_n^2$, 令

$$\alpha = E_{\sigma_0^2} [\varphi(\mathbf{X})] = P_{\sigma_0^2} \left(\sum_{i=1}^n X_i^2 < c \right) = P_{\sigma_0^2} \left(\sum_{i=1}^n X_i^2 / \sigma_0^2 < \frac{c}{\sigma_0^2} \right),$$

故有 $c/\sigma_0^2 = \chi_n^2(1 - \alpha)$, 即 $c = \sigma_0^2 \chi_n^2(1 - \alpha)$. 因此

$$\varphi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) < \sigma_0^2 \chi_n^2(1 - \alpha), \\ 0, & T(\mathbf{x}) \geq \sigma_0^2 \chi_n^2(1 - \alpha) \end{cases}$$

为水平为 α 的 UMPT.

*5.4.4 无偏检验

前面已经说过, UMPT 存在是很少的例外. 因此作为一致最优检验的准则, 它的作用是有限的. 为了得到适用范围更广的检验准则, 可采取下列办法: 先对所考虑的检验施加某种合理的一般性限制, 这样就缩小了所考虑的检验的范围, 然后在这缩小了的范围内找一致最优检验. 正如在点估计问题中, 先限制估计量是无偏的, 然后在无偏估计类中, 去寻找方差一致最小的无偏估计. 基于这种想法引进无偏检验的概念.

定义 5.4.2 设 φ 为检验问题 (5.4.1) 的一个检验, 若其功效函数 $\beta_\varphi(\theta)$ 满足条件: 对 $\forall \theta_0 \in \Theta_0$ 有 $\beta_\varphi(\theta_0) \leq \alpha$, 对 $\forall \theta_1 \in \Theta_1$ 有 $\beta_\varphi(\theta_1) \geq \alpha$, 则称 φ 为水平为 α 的无偏检验 (unbiased test), 或简称为无偏检验.

无偏检验的直观意义很清楚: 若 φ 为 $H_0 \leftrightarrow H_1$ 的无偏检验, 则其犯第一类错误的概率不应超过不犯第二类错误的概率.

下面给出一致最优无偏检验的定义. 记

$$\mathcal{U}_\alpha = \{\varphi : \varphi \text{ 为检验问题 (5.4.1) 的水平 } \alpha \text{ 的无偏检验}\},$$

即 \mathcal{U}_α 为一切水平为 α 的无偏检验的类.

定义 5.4.3 若 $\varphi \in \mathcal{U}_\alpha$ 且对任何 $\varphi_1 \in \mathcal{U}_\alpha$, 有

$$\beta_\varphi(\theta) \geq \beta_{\varphi_1}(\theta), \quad \theta \in \Theta_1,$$

则称 φ 是式 (5.4.1) 的一个水平为 α 的一致最优无偏检验 (uniformly most powerful unbiased test, UMPUT).

注 5.4.7 由上述定义可知任一 UMPT 必为 UMPUT. 说明如下: 记 UMPT φ 的功效函数为 $\beta_\varphi(\theta)$, 由 UMPT 的定义可知 $\beta_\varphi(\theta) \leq \alpha$, 对一切 $\theta \in \Theta_0$, 又显

见 $\varphi^* \equiv \alpha$ 是水平为 α 的检验, 由 UMPT 定义可知 $\beta_\varphi(\theta) \geq \beta_{\varphi^*}(\theta) \equiv \alpha$, 对一切 $\theta \in \Theta_1$. 可见有

$$\beta_\varphi(\theta_1) \geq \alpha \geq \beta_\varphi(\theta_0), \quad \forall \theta_0 \in \Theta_0, \theta_1 \in \Theta_1,$$

故检验 φ 是无偏的, 又是 UMPT, 因此必为 UMPUT.

UMPUT 存在的情况比 UMPT 要广一些. 对下列单参数指数族:

$$f(x, \theta) = c(\theta) \exp \{Q(\theta)T(x)\} h(x),$$

在前面的定理 5.4.2 和定理 5.4.3 中已证明了下列检验问题的水平为 α 的 UMPT 存在, 因而也是 UMPUT 的.

$$(1) H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0,$$

$$(2) H_0 : \theta \geq \theta_0 \leftrightarrow H_1 : \theta < \theta_0.$$

还可进一步证明下列两类单参数指数族的水平为 α 的 UMPUT 是存在的:

$$(3) H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0,$$

$$(4) H_0 : \theta_1 \leq \theta \leq \theta_2 \leftrightarrow H_1 : \theta < \theta_1 \text{ 或 } \theta > \theta_2,$$

其中 (3) 和 (4) 两类检验问题的 UMPUT 的存在性已超出本书的范围, 有兴趣的同学可参看文献 [1] 第 359 页.

5.5 假设检验与区间估计

假设检验与区间估计这两个统计推断的形式表面上看好像完全不同, 而实际上两者之间有着非常密切的关系. 由单参数假设检验问题的水平为 α 的双边检验和单边检验, 可以分别得到该参数的置信系数为 $1 - \alpha$ 的置信区间和置信限, 反之亦然. 具体说明如下.

5.5.1 如何由假设检验得到置信区间

设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $\{f(x, \theta), \theta \in \Theta\}$ 中抽取的简单样本, 目的是求参数 θ 的置信系数为 $1 - \alpha$ 的置信区间. 考虑双边检验问题

$$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0.$$

求出此检验的水平为 α 的接受域 \bar{D} , 则有

$$P(\bar{D} | H_0) = 1 - \alpha, \tag{5.5.1}$$

解由 \bar{D} 确定的不等式, 得到如下不等式: $\hat{\theta}_1(\mathbf{X}) \leq \theta_0 \leq \hat{\theta}_2(\mathbf{X})$, 由于式 (5.5.1) 是在条件 “ $H_0 : \theta = \theta_0$ ” 下成立, 改 θ_0 为 θ 得 $\hat{\theta}_1(\mathbf{X}) \leq \theta \leq \hat{\theta}_2(\mathbf{X})$, 则 $[\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X})]$ 即为所求的置信系数为 $1 - \alpha$ 的置信区间.

若要求 θ 的置信上、下限, 就需要考虑单边检验 $H_0'': \theta \geq \theta_0 \leftrightarrow H_1'': \theta < \theta_0$ 或 $H_0': \theta \leq \theta_0 \leftrightarrow H_1': \theta > \theta_0$ 的检验问题. 下面通过例子来说明.

例 5.5.1 设 X_1, \dots, X_n 为自正态总体 $N(\mu, \sigma^2)$ 抽取的随机样本. μ, σ^2 皆未知, 要分别求 μ 和 σ^2 的置信系数为 $1 - \alpha$ 的置信区间和置信上、下限.

解 先考虑 μ 的置信区间和置信上、下限问题. 5.2 节已给出检验问题

$$H_0: \mu = \mu_0 \leftrightarrow H_1: \mu \neq \mu_0$$

的水平为 α 的检验的接受域 $\bar{D} = \{(X_1, \dots, X_n) : |T| \leq t_{n-1}(\alpha/2)\}$, 其中 $T = \sqrt{n}(\bar{X} - \mu_0)/S$. 记 $\theta = (\mu, \sigma^2)$, 故有

$$P_\theta \left(\left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \right| \leq t_{n-1}(\alpha/2) \mid H_0 \right) = 1 - \alpha. \quad (5.5.2)$$

由于上述等式是在条件 H_0 成立, 即 $\mu = \mu_0$ 时获得的, 所以将下面出现的所有 μ_0 用 μ 代替是等价的. 解式 (5.5.2) 括号中的不等式得

$$\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2) \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2),$$

因此

$$\left[\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2), \bar{X} + \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2) \right]$$

即为 μ 的置信系数为 $1 - \alpha$ 的置信区间.

若要求 μ 的置信下限, 则考虑检验问题

$$H_0': \mu \leq \mu_0 \leftrightarrow H_1': \mu > \mu_0.$$

在 5.2 节中已给出水平为 α 的接受域 $\bar{D} = \{(X_1, \dots, X_n) : T \leq t_{n-1}(\alpha)\}$, 故有

$$P_\theta \left(\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \leq t_{n-1}(\alpha) \mid \mu = \mu_0 \right) = 1 - \alpha.$$

解括号中的不等式得

$$\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha) \leq \mu_0,$$

再改 μ_0 为 μ 得到 $\bar{X} - St_{n-1}(\alpha)/\sqrt{n} \leq \mu < \infty$. 因此 μ 的置信系数为 $1 - \alpha$ 的置信下限为 $\bar{X} - St_{n-1}(\alpha)/\sqrt{n}$. 同理, 可求 μ 的置信系数为 $1 - \alpha$ 的置信上限为 $\bar{X} + St_{n-1}(\alpha)/\sqrt{n}$.

关于正态总体方差 σ^2 的置信区间和置信上、下限留给读者作为练习.

例 5.5.2 设 X_1, \dots, X_m 和 Y_1, \dots, Y_n 分别自正态总体 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$ 中抽取的简单随机样本, 且样本 X_1, \dots, X_m 和 Y_1, \dots, Y_n 独立. 令 $\mu = \mu_2 - \mu_1$, 求 μ 的置信系数为 $1 - \alpha$ 的置信区间和置信上、下限.

解 5.2 节已求出了检验问题

$$H_0: \mu = \mu_0 \leftrightarrow H_1: \mu \neq \mu_0$$

的两样本 t 检验的接受域 $\bar{D} = \{(X, Y) : |T_w| \leq t_{n+m-2}(\alpha/2)\}$. 检验统计量为

$$T_w = \frac{\bar{Y} - \bar{X} - \mu_0}{S_w} \sqrt{\frac{mn}{m+n}},$$

此处 $S_w^2 = [(m-1)S_1^2 + (n-1)S_2^2]/(n+m-2)$, 而 S_1^2 和 S_2^2 分别为两组样本的样本方差. 若记 $\theta = (\mu_1, \mu_2, \sigma^2)$, 则有

$$P_\theta(|T_w| \leq t_{n+m-2}(\alpha/2) | H_0) = 1 - \alpha, \quad (5.5.3)$$

改 μ_0 为 μ , 解式 (5.5.3) 括号中的不等式得到

$$\bar{Y} - \bar{X} - S_w t_{n+m-2} \left(\frac{\alpha}{2} \right) \sqrt{\frac{1}{m} + \frac{1}{n}} \leq \mu \leq \bar{Y} - \bar{X} + S_w t_{n+m-2} \left(\frac{\alpha}{2} \right) \sqrt{\frac{1}{m} + \frac{1}{n}}.$$

因此 $\mu = \mu_2 - \mu_1$ 的置信系数为 $1 - \alpha$ 的置信区间为

$$\left[\bar{Y} - \bar{X} - S_w t_{n+m-2} \left(\frac{\alpha}{2} \right) \sqrt{\frac{1}{m} + \frac{1}{n}}, \bar{Y} - \bar{X} + S_w t_{n+m-2} \left(\frac{\alpha}{2} \right) \sqrt{\frac{1}{m} + \frac{1}{n}} \right].$$

类似方法求得 $\mu = \mu_2 - \mu_1$ 的置信系数为 $1 - \alpha$ 的置信下、上限分别为 $\bar{Y} - \bar{X} - S_w t_{n+m-2}(\alpha) \sqrt{1/m + 1/n}$ 和 $\bar{Y} - \bar{X} + S_w t_{n+m-2}(\alpha) \sqrt{1/m + 1/n}$.

这里假定了两总体有相同的方差 σ^2 . 若去掉这一假设, 假定两总体的方差分别为 σ_1^2 和 σ_2^2 , 则就得到著名的 Behrens-Fisher 问题, 由 5.2 节中给出的 Behrens-Fisher 问题的大样本检验方法和一个小样本的近似方法, 用类似的方法也同样可以得到一个近似的 Behrens-Fisher 问题的区间估计形式 (这已在 4.2 节中讨论过, 此处从略).

两正态总体方差比的置信区间和置信上、下限如何通过假设检验方法得到, 留给读者作练习.

5.5.2 如何由置信区间得到假设检验

若用某种方法建立了 θ 的置信水平为 $1 - \alpha$ 的区间估计 $[\hat{\theta}_1, \hat{\theta}_2]$, 对给定的 θ_0 不难求出检验问题 $H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$ 的一个水平为 α 的检验. 事实上, 一个简单方法就是若 $\theta_0 \in [\hat{\theta}_1, \hat{\theta}_2]$ 则接受 H_0 , 否则就拒绝 H_0 .

用类似方法可由置信系数为 $1 - \alpha$ 的置信上、下限求出检验问题 $H'_0 : \theta \geq \theta_0 \leftrightarrow H'_1 : \theta < \theta_0$ 和 $H''_0 : \theta \leq \theta_0 \leftrightarrow H''_1 : \theta > \theta_0$ 的水平为 α 的检验.

*5.5.3 一致最精确置信集与一致最优检验

1. 引言和定义

由 4.1 节可知, 评价一个区间估计的优劣有两个要素: 一是其可靠度, 即区间估计包含未知参数的概率有多大, 可靠度越大越好; 二是其精度, 衡量精度的明显指标是区间估计的长度. 长度越短越好. 如 4.1 节所述, 在固定样本大小的前提下, 这两个要素是彼此矛盾的.

长度作为精度的标准不适用于置信上、下限. 在满足一定可靠度的前提下, 置信上限越小精度越高; 置信下限越大精度越高.

如 4.1 节所述, Neyman 的置信区间理论是保证可靠度达到指定要求下, 尽量选择精度更高的置信集. 固定可靠度, 使精度达到最高的区间估计称为一致最精确区间估计 (Uniformly Most Accurate Confidence Interval), 其定义如下.

定义 5.5.1 称 $\hat{\theta}_U^*(\mathbf{X})$, $\hat{\theta}_L^*(\mathbf{X})$ 和 $[\hat{\theta}_1^*(\mathbf{X}), \hat{\theta}_2^*(\mathbf{X})]$ 分别是 θ 的置信水平为 $1 - \alpha$ 的一致最精确 (Uniformly Most Accurate, 简称 UMA) 置信上、下限和置信区间, 如果它们有置信水平 $1 - \alpha$, 而且

(1) 对任何置信水平为 $1 - \alpha$ 的置信上限 $\hat{\theta}_U(\mathbf{X})$, 以及任何 $\theta < \theta'$, 有

$$P_\theta(\hat{\theta}_U^*(\mathbf{X}) \geq \theta') \leq P_\theta(\hat{\theta}_U(\mathbf{X}) \geq \theta');$$

(2) 对任何置信水平为 $1 - \alpha$ 的置信下限 $\hat{\theta}_L(\mathbf{X})$, 以及任何 $\theta > \theta'$, 有

$$P_\theta(\hat{\theta}_L^*(\mathbf{X}) \leq \theta') \leq P_\theta(\hat{\theta}_L(\mathbf{X}) \leq \theta');$$

(3) 对任何置信水平为 $1 - \alpha$ 的置信区间 $[\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X})]$, 以及任何 $\theta \neq \theta'$, 有

$$P_\theta(\hat{\theta}_1^*(\mathbf{X}) \leq \theta' \leq \hat{\theta}_2^*(\mathbf{X})) \leq P_\theta(\hat{\theta}_1(\mathbf{X}) \leq \theta' \leq \hat{\theta}_2(\mathbf{X})).$$

注 5.5.1 在上述定义中 θ 为真参数, θ' 是非真参数. 由定义可知 UMA 置信集 (包括置信区间和置信上、下限) 的意义是: 它们包含非真参数 θ' 的概率是最小的.

2. 如何求 UMA 置信集

如本节 5.5.1 小节所述, 置信区间和置信上、下限与假设检验的接受域之间存在对应关系, 故可以通过一致最优检验 (UMPT) 的接受域来确定 UMA 置信区间和置信上、下限.

定理 5.5.1 由本节 5.5.1 小节中介绍过的由假设 $\theta = \theta_0$, $\theta \geq \theta_0$, $\theta \leq \theta_0$ 构造置信水平为 $1 - \alpha$ 的置信区间和置信上、下限的方法中, 若所用的检验是水平为 α 的 UMPT, 则所得到的置信区间和置信上、下限必是置信水平为 $1 - \alpha$ 的 UMA 置信区间和置信上、下限.

证 以置信下限为例证明, 其余两种情形的证明类似. 设 $\{X \in A^*(\theta_0)\}$ 为 $H : \theta \leq \theta_0 \leftrightarrow K : \theta > \theta_0$ 的检验水平为 α 的 UMPT 的接受域, 产生置信水平为 $1 - \alpha$ 的置信下限 $\hat{\theta}_L^*(\mathbf{X})$. 引进上述 $H \leftrightarrow K$ 水平为 α 的任一检验, 它有接受域 $\{\mathbf{X} : \hat{\theta}_L(\mathbf{X}) \leq \theta_0\} = A(\theta_0)$. 事实上若原假设 $H : \theta \leq \theta_0$ 成立, 则有

$$P_\theta(\text{原假设被接受}) = P_\theta(\hat{\theta}_L(\mathbf{X}) \leq \theta_0 | H) \geq P_\theta(\hat{\theta}_L(\mathbf{X}) \leq \theta_0 | H) \geq 1 - \alpha,$$

这证明它有水平 α .

但以 $A^*(\theta_0)$ 为接受域的检验是水平 α 的 UMPT, 按定义对 $\forall \theta_1 > \theta_0$, 有

$$\begin{aligned} P_{\theta_1}(\mathbf{X} \notin A^*(\theta_0)) &\geq P_{\theta_1}(\mathbf{X} \notin A(\theta_0)), \quad \text{即不犯 II 型错误的概率最大} \\ \iff P_{\theta_1}(\mathbf{X} \in A^*(\theta_0)) &\leq P_{\theta_1}(\mathbf{X} \in A(\theta_0)), \quad \text{即犯 II 型错误的概率最小} \\ \iff P_{\theta_1}(\hat{\theta}_L^*(\mathbf{X}) \leq \theta_0) &\leq P_{\theta_1}(\hat{\theta}_L(\mathbf{X}) \leq \theta_0). \end{aligned}$$

由于上式对满足 $\theta_1 > \theta_0$ 的任何 θ_1 成立, 故改 θ_1 为 θ , 改 θ_0 为 θ' 上式也成立. 即对 $\forall \theta > \theta'$, 有

$$P_\theta(\hat{\theta}_L^*(\mathbf{X}) \leq \theta') \leq P_\theta(\hat{\theta}_L(\mathbf{X}) \leq \theta').$$

按一致最精确置信集的定义, 可知 $\hat{\theta}_L^*(\mathbf{X})$ 是置信水平为 $1 - \alpha$ 的 UMA 置信下限.

例 5.5.3 设 X_1, \dots, X_n i.i.d. $\sim N(\theta, 1)$, 求 θ 的置信水平 $1 - \alpha$ 的 UMA 置信上、下限.

解 显然检验问题 $H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0$ 的水平为 α 的 UMPT 的接受域为 $\{\sqrt{n}(\bar{X} - \theta_0) \leq u_\alpha\} \iff \{\bar{X} - u_\alpha/\sqrt{n} \leq \theta_0\}$. 由定理 5.5.1 可知 $\bar{X} - u_\alpha/\sqrt{n}$ 就是 θ 的置信水平 $1 - \alpha$ 的 UMA 置信下限, 同理可知 $\bar{X} + u_\alpha/\sqrt{n}$ 是 θ 的置信水平 $1 - \alpha$ 的 UMA 置信上限.

5.5.4 假设检验和区间估计的比较

与点估计和假设检验比较, 区间估计这一推断形式有一个显著的特点, 即它的精度 (一般可用区间的长度刻画) 和可靠度 (用其置信系数刻画) 一目了然. 点估计

不具备这个特点, 才促使人们考虑区间估计. 而且区间估计可以在精度、可靠度和样本大小 n 之间进行调整, 以达到预先指定的要求. 而假设检验提供的信息不如区间估计确切, 请看下例:

设从正态总体 $N(\mu, \sigma^2)$ 中抽取一定大小的样本去检验假设 $H_0 : \mu = 0 \leftrightarrow H_1 : \mu \neq 0$. 结果假设被接受了. 如在 5.1 节中所述, 这并不意味着“证明”了 $\mu = 0$. 假如只知道 $\mu = 0$ 被接受了, 甚至无法估量真正的 μ 值与 0 相差有多大. 但如果被告知 μ 的置信系数 95% 的区间估计为 $[-0.05, 0.07]$ 或者是 $[-15, 20]$, 则在前一个场合, μ 与 0 相距最大不超过 0.07, 这么小一个值在实用上可能无关紧要. 这时就有一定的把握(概率 0.95)说 μ “事实上”可以认为是 0, 而不只是接受“ $\mu = 0$ ”了. 若在后一个场合, 虽然 $\mu = 0$ 这个假设也被接受(因为 0 这个点在区间 $[-15, 20]$ 内), 但因 μ 的可能范围很大, 实际上只能说对 μ “知之甚少”.

反之, 若得到“ $\mu = 0$ 被否定”. 从这句话也只知道有比较显著的证据认为 $\mu \neq 0$, 但还无法知道其实际意义如何. 但如果被告知: μ 的区间估计为 $[0.01, 0.02]$ 或 $[-40, -30]$. 在前一场合, 虽然 $\mu = 0$ 被否定(因为 0 不在区间 $[0.01, 0.02]$ 内), 但 μ 与 0 的最大差距不过 0.02, 这么小一个值可能实际上与 0 无异. 因此, 虽然在统计上否定了 $\mu = 0$, 但事实上可以认为 $\mu = 0$. 在后一个场合 μ 的值与 0 相距至少是 30, 不仅要否定 $\mu = 0$, 从实际上看 μ 也显著异于 0.

这些分析说明, 区间估计所提供的信息比假设检验更为确切. 这也提醒: ① 对假设检验结果的实际含义的解释要十分小心; ② 在得到假设检验结果时, 最好也将被检验参数的区间估计求出来作为参考.

5.5.5 检验的 p 值

假设检验的可能结论只有两个: 接受或是否定原假设. 作出这一结论的根据有多大, 则往往不易清楚地显示出来. 例如, 甲、乙两厂生产同一产品, 希望检验假设 H_0 : “甲不优于乙”. 当被告知这一假设应被接受时, 只知道作出的结论是“甲不优于乙”, 但作出这一结论的根据有多大? 不可能有一个数量的概念. 这是假设检验这种推断形式的一个缺点. 上一段曾将假设检验和区间估计作了一个比较, 并指出: 一般说来假设检验作出的结论, 不如区间估计那么精细, 其理由就在于检验这种形式固有的粗糙性.

但是, 对这一缺点可以作一些补救, 方法是引进“检验的 p 值”的概念. 这是基于下面直观上的想法.

例如, 设 X_1, \dots, X_{16} 为自正态总体 $N(\mu, 1)$ 中抽取的简单样本, 要检验假设 $H_0 : \mu = 0 \leftrightarrow H_1 : \mu \neq 0$, 取水平 $\alpha = 0.05$. 此检验的否定域为

$$\left\{ (X_1, \dots, X_{16}) : |\sqrt{16} \bar{X}| > 1.96 \right\} = \left\{ (X_1, \dots, X_n) : |\bar{X}| > 0.49 \right\}.$$

设对一组样本 X_1, \dots, X_{16} , 有 $\bar{X} = 0.48$, 则根据拒绝域, 应接受 $H_0 : \mu = 0$; 又设有另一组样本 X_1, \dots, X_{16} 算得 $\bar{X} = 0.12$, 当然也应接受 $H_0 : \mu = 0$. 对这两组样本而言, 结论一致, 都是接受 $H_0 : \mu = 0$. 然而, 会觉得后一场合, 作出 $H_0 : \mu = 0$ 的结论根据大一些, 而在前一场合, 根据就小一些. 为了反映这一点, 引进 检验的 p 值 进行定量的刻画. 其定义如下:

设对某一组如上所述的具体样本 X_1, \dots, X_n , 算出 \bar{X} 的具体值记为 \bar{x}_0 , 则这组样本的 p 值 定义为

$$p = P(|\bar{X}| > |\bar{x}_0| | H_0) = P(\sqrt{n}|\bar{X}| > \sqrt{n}|\bar{x}_0| | H_0) = P(|U| > \sqrt{n}|\bar{x}_0|),$$

此处 $U = \sqrt{2}|\bar{X}|$, 且 $U | H_0 \sim N(0, 1)$.

p 值的意义解释如下: 当获得一组具体样本, 算得样本均值 \bar{X} 的具体值为 \bar{x}_0 , 它与假设 $\mu = 0$ 的偏差 $|\bar{x}_0|$, 问: 达到 $|\bar{x}_0|$ 这么大或更大的偏离的机会有多大? 这就是上式定义的 p 值. 若概率 p 很大, 就证明在 $\mu = 0$ 之下, 得到这么大一个偏差很正常, 不值得奇怪, 因而认为 $\mu = 0$ 成立的根据很充分. 反之, 若 p 很小, 则在 $\mu = 0$ 之下得到这么大一个偏差很难得, 这很有可能意味着 μ 不为 0. 因此若 p 很小时, 认为 $\mu = 0$ 的根据很不足. 总之, p 越大 (小), 认为 $\mu = 0$ 的根据越充分 (不足). 当 $p < \alpha$ 时, 就要否定 $\mu = 0$ 了. 若 $p \geq \alpha$, 但离 α 很近, 虽然不能拒绝 H_0 , 但对它抱着很怀疑的态度. 拿上文的例子来说, 通过标准正态分布表, 查得对应于 $\bar{x}_0 = 0.48$ 和 $\bar{x}_0 = 0.12$ 两种情形的 p 值分别为

$$\begin{aligned} p &= P(|\bar{X}| \geq 0.48 | H_0) = P(|U| \geq 1.92) = 0.0548; \\ p &= P(|\bar{X}| \geq 0.12 | H_0) = P(|U| \geq 0.48) = 0.6312. \end{aligned}$$

前一情况离水平 α 很近, 虽然仍不能拒绝 $\mu = 0$, 但很值得怀疑. 后一场合表明: 出现像 0.12 或更大偏差的可能性在 $\mu = 0$ 之下为 0.6312, 这一可能性很大, 不足为奇. 故认为 $\mu = 0$ 的根据很充分.

上述分析可以推广到一般的情形, 对双边检验问题, 若原假设为 $H_0 : \theta = \theta_0$, 其否定域为 $|T| > c$, 设由样本算出检验统计量 T 之值为 t_0 , 则这组样本的 p 值为

$$p = P(|T| > |t_0| | H_0); \quad (5.5.4)$$

对单边检验, 若原假设为 $H_0 : \theta \leq \theta_0$, 其否定域为 $T > c$, 则 p 值为

$$p = P(T > t_0 | \theta = \theta_0); \quad (5.5.5)$$

对单边检验问题, 若原假设为 $H_0 : \theta \geq \theta_0$, 否定域为 $T < c$, 则 p 值为

$$p = P(T < t_0 | \theta = \theta_0). \quad (5.5.6)$$

将一样本问题中的 U 检验, t 检验和 χ^2 检验的 p 值公式及两样本问题中 t 检验, F 检验的部分 p 值列成表 5.5.1. 双边检验的 p 值公式可按式 (5.5.4) 推导, 单边检验的 p 值公式可按式 (5.5.5) 和式 (5.5.6) 推得.

表 5.5.1 p 值计算公式表

	参数	H_0	H_1	检验统计量 及其分布	p 值公式
一 样 本	σ^2 已知	$\mu = \mu_0$	$\mu \neq \mu_0$	$U = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0, 1)$	$p = P\left(U > \frac{\sqrt{n} \bar{x}_0 - \mu_0 }{\sigma}\right)$
	同上	$\mu \leq \mu_0$	$\mu > \mu_0$	同上	$p = P\left(U > \frac{\sqrt{n}(\bar{x}_0 - \mu_0)}{\sigma}\right)$
	σ^2 未知	$\mu = \mu_0$	$\mu \neq \mu_0$	$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim t_{n-1}$	$p = P\left(T > \frac{\sqrt{n} \bar{x}_0 - \mu_0 }{s_0}\right)$
	同上	$\mu \leq \mu_0$	$\mu > \mu_0$	同上	$p = P\left(T > \frac{\sqrt{n}(\bar{x}_0 - \mu_0)}{s_0}\right)$
	μ 未知	$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2$	$p = P\left(\chi^2 > \frac{(n-1)s_0^2}{\sigma_0^2}\right)$
	$\sigma_1^2 = \sigma_2^2$ 未知	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$T = \frac{\bar{Y} - \bar{X}}{S_w \sqrt{\frac{m+n}{mn}}} \sim t_{n+m-2}$	$p = P\left(T > \frac{ \bar{y}_0 - \bar{x}_0 }{S_{w0} \sqrt{\frac{m+n}{mn}}}\right)$
两 样 本	同上	$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$	同上	$p = P\left(T > \frac{ \bar{y}_0 - \bar{x}_0 }{S_{w0} \sqrt{\frac{m+n}{mn}}}\right)$
	μ_1, μ_2 未知	$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F = \frac{S_1^2}{S_2^2} \sim F_{m-1, n-1}$	$p = P(F > s_{10}^2 / s_{20}^2)$

表 5.5.1 一样本问题中 \bar{x}_0 , s_0^2 分别由 \bar{X} , S^2 的具体样本算得的值; 两样本问题中 \bar{y}_0 , \bar{x}_0 , s_{10}^2 , s_{20}^2 和 s_{w0} 分别由 \bar{Y} , \bar{X} , S_1^2 , S_2^2 及 S_{w0} 的具体样本算得的值. 一样本问题中的 \bar{X} , S^2 及两样本问题中的 \bar{Y} , \bar{X} , S_1^2 , S_2^2 和 S_w^2 的表达式与 5.2 节中相同.

例 5.5.4 从电信公司每月长途电话的账单中随机抽取 25 张, 算得月平均费用 $\bar{X} = 32.80$ 元, $S = 20.80$ 元. 假定每月长途电话费用服从正态分布 $N(\mu, \sigma^2)$, σ^2 未知, 要检验假设

$$H_0 : \mu = 30 \leftrightarrow H_1 : \mu \neq 30,$$

计算检验的 p 值.

解 按表 5.5.1 第三行中公式计算 p 值,

此处 $n = 25$, $\mu_0 = 30$, $\bar{x}_0 = 32.80$, $s_0 = 20.80$, 故有

$$p = P\left(|T| > \left|\frac{\sqrt{25}(32.80 - 30)}{20.80}\right|\right) = P(|T| > 0.673) \approx 0.53.$$

此 p 值较大, 表明数据与 $\mu = 30$ 相当符合, 即有足够的根据认为 $H_0 : \mu = 30$ 成立.

习题 5

1. 设 $\mathbf{X} = (X_1, \dots, X_{20})$ 为从两点分布总体 $B(1, p)$ 中抽取的简单样本, 对未知参数 p 的检验问题为 $H_0 : p = 0.2 \leftrightarrow H_1 : p \neq 0.2$, 取检验函数为

$$\phi(\mathbf{x}) = \begin{cases} 1, & \sum_{i=1}^{20} x_i \geq 7 \text{ 或 } \sum_{i=1}^{20} x_i \leq 1, \\ 0, & \text{其他.} \end{cases}$$

- (1) 求此检验函数在 $p = 0, 0.1, 0.2, \dots, 0.9, 1$ 时的功效函数并作图;
 (2) 求检验的水平 α 和 $p = 0.05$ 时犯第二类错误的概率.
2. 设 X_1, \dots, X_n i.i.d. $\sim N(\mu, 9)$, 其中 μ 为未知参数, \bar{X} 为样本均值. 设检验问题 $H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0$ 的否定域为

$$\{(X_1, \dots, X_n) : |\bar{X} - \mu_0| \geq c\}.$$

- (1) 定出常数 c , 使检验水平为 $\alpha = 0.05$;
 (2) 求此检验的功效函数 $\beta(\mu)$;
 (3) 固定样本容量 $n = 25$, 分析犯两种错误概率 α 和 β 之间的关系.
3. 设 X_1, \dots, X_n i.i.d. 服从均匀分布 $U(0, \theta)$, θ 为未知参数. $X_{(n)} = \max\{X_1, \dots, X_n\}$ 对检验问题 $H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0$ 取检验函数为

$$\varphi(\mathbf{x}) = \begin{cases} 1, & X_{(n)} \geq c, \\ 0, & \text{其他.} \end{cases}$$

- (1) 求检验函数 $\varphi(\mathbf{x})$ 的功效函数, 并证明它是 θ 的单调递增函数.
 (2) 在检验问题 $H_0 : \theta \leq 1/2 \leftrightarrow H_1 : \theta > 1/2$ 中, 选择什么样的 c 使检验水平恰好为 0.05?
 (3) 画出 $n = 20$ 时 (2) 中指定的 $\varphi(\mathbf{x})$ 的功效函数的粗略图形;
 (4) n 该多大, 能使 (2) 中检验问题的功效函数 $\varphi(\mathbf{x})$, 对于 $\theta = 3/4$ 功效 (即功效函数在 θ_1 中某一点的值) 为 0.98?
4. 设 X_1, \dots, X_n 取自均匀分布 $U(0, \theta)$ 的随机样本, 其中 $\theta > 0$ 为未知参数. 记 $X_{(n)} = \max\{X_1, \dots, X_n\}$. 如对检验问题

$$H_0 : \theta \geq 2 \leftrightarrow H_1 : \theta < 2$$

取检验的否定域为

$$W = \{(X_1, \dots, X_n) : X_{(n)} \leq 1.5\},$$

试求此检验犯第一类错误的概率的最大值.

5. 设 X_1, \dots, X_{10} 为取自 Poisson 分布 $P(\lambda)$ 的随机样本.

(1) 试用直观方法求单边假设检验问题 $H_0 : \lambda \leq 0.1 \leftrightarrow H_1 : \lambda > 0.1$ 的水平 $\alpha = 0.05$ 的检验.

(2) 求此检验的功效函数 $\beta(\lambda)$ 在 $\lambda = 0.05, 0.2, 0.3, \dots, 0.9$ 时的值, 并画出 $\beta(\lambda)$ 的图像.

6. 设 X_1, \dots, X_n 为取自两点分布 $b(1, p)$ 的随机样本.

(1) 试用直观方法求单边假设检验问题

$$H_0 : p \leq 0.01 \leftrightarrow H_1 : p > 0.01$$

的水平 $\alpha = 0.05$ 的检验.

(2) 若要求这个检验在 $p = 0.08$ 时犯第二类错误的概率不超过 0.10, 样本容量 n 应为多大?

7. 假定某切割机在正常工作时, 切割每段金属棒的长度 (单位: cm) 服从正态分布 $N(10.5, 0.15^2)$. 今从一批产品中随机抽取 15 段进行测量, 其结果如下:

$$\begin{aligned} & 10.4, 10.6, 10.1, 10.4, 10.5, 10.3, 10.3, 10.2, \\ & 10.9, 10.6, 10.8, 10.5, 10.7, 10.2, 10.7. \end{aligned}$$

试问该机工作是否正常 ($\alpha = 0.05$)?

8. 设某产品的指标服从正态分布, 已知它的标准差 $\sigma = 150$, 今抽了一个样本容量为 26 的样本, 计算得样本均值为 1637, 问在 5% 的显著水平下, 能否认为这批产品指标的期望值 μ 不大于 1600.

9. 某纺织厂在正常运行条件下, 平均每台布机每小时经纱断头为 0.973 根, 各台布机断头数的方差为 0.162 根. 该厂进行工艺改革, 减少经纱上浆率, 在 200 台布机上进行试验, 结果平均每台每小时经纱断头为 0.994 根. 假定每台布机断头数服从正态分布, 问新工艺上浆率能否推广 ($\alpha = 0.05$)?

10. 根据长期经验和资料分析, 某砖瓦厂生产的砖的抗断强度 X 服从方差为 σ^2 的正态分布, 今从该厂所生产的一批砖中随机抽取 6 块, 测得抗断强度 (单位: kg/cm^2) 如下:

$$32.56, 29.66, 31.64, 30.00, 31.87, 31.03.$$

问这一批砖的平均抗断强度可否认为是 $32.50 \text{ kg}/\text{cm}^2$? ($\alpha = 0.01$)

11. 测定某种溶液中的水分, 它的 10 个测定值给出 $\bar{X} = 0.452\%$, $S = 0.037\%$, 设测定值总体为正态分布 $N(a, \sigma^2)$. 试在水平 $\alpha = 0.05$ 下检验假设

(1) $H_0 : a \leq 0.5\% \leftrightarrow H_1 : a > 0.5\%$;

(2) $H'_0 : \sigma \geq 0.04\% \leftrightarrow H'_1 : \sigma < 0.04\%$.

12. 设有甲乙两个实验员, 对同样的试样进行分析, 各人试验分析结果如下 (分析结果服从正态分布):

实验员	1	2	3	4	5	6	7	8
甲	4.3	3.2	3.8	3.5	3.5	4.8	3.3	3.9
乙	3.7	4.1	3.8	3.8	4.6	3.9	2.8	4.4

试问甲乙两实验员试验分析结果之间有无显著差异 ($\alpha = 0.05$)?

13. 有一种新安眠药, 据说在一定剂量下, 能比某旧安眠药平均增加睡眠时间 3 小时. 根据以往资料, 用旧安眠药时平均睡眠时间为 20.8 小时. 为了检验新安眠药是否达到疗效, 收集到一组用新安眠药的睡眠时间, 分别为

$$26.7, 22.0, 24.1, 21.0, 27.2, 25.0, 23.4.$$

假定睡眠时间服从正态分布, 试在水平 $\alpha = 0.05$ 下, 检验假设: 新安眠药已达到疗效.

14. 某工厂生产的某种细纱支数服从正态分布, 其总体标准差为 1.2, 现从某日生产的一批产品中随机抽取 16 缕进行支数测量, 求得样本标准差为 2.1, 问纱的均匀度是否变劣 ($\alpha = 0.05$)?

15. 某厂的一批电子产品, 其寿命 T 服从指数分布, 其密度函数为

$$f(t, \theta) = \theta^{-1} \exp\{-t/\theta\} I_{(0, \infty)}(t).$$

从以往生产情况知平均寿命 $\theta = 2000$ 小时. 为检验当日生产是否稳定, 任取 10 个产品进行寿命试验, 到全部失效时试验停止. 试验得失效寿命数据之和为 30200. 试在显著水平 $\alpha = 0.05$ 下检验假设

$$H_0 : \theta = 2000 \leftrightarrow H_1 : \theta \neq 2000.$$

16. 设 X_1, \dots, X_9 i.i.d. $\sim N(a, 2.5^2)$; Y_1, \dots, Y_{16} i.i.d. $\sim N(b, 3.4^2)$, 算得 $\bar{x} = 49$, $\bar{y} = 44$, 在水平 $\alpha = 0.05$ 下检验

$$H_0 : a \geq b \leftrightarrow H_1 : a < b.$$

17. 用两种方法研究冰溶解时的潜热, 对冷却到 -0.72°C 的冰溶解为 0°C 的水的热量改变数据 (单位: cal/g):

方法 I: 79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04,
79.97, 80.05, 80.03, 80.02, 80.00, 80.02.

方法 II: 80.02, 79.94, 79.98, 79.97, 80.03, 79.95, 79.97, 79.97.

试问对水平 $\alpha = 0.05$ 这两种处理方法的平均性能有否显著差异? 假设这两种方法测定值都是服从正态分布且方差相同.

18. 有甲乙两台机床加工同样产品, 从这两台机床加工的产品中随机抽取若干产品, 测得产品直径 (单位: mm) 为

机床甲: 20.5, 29.8, 19.7, 20.4, 20.1, 20.0, 19.0, 19.9.

机床乙: 19.7, 20.8, 20.5, 19.8, 19.4, 20.6, 19.2.

假定产品直径服从正态分布, 且二者方差相同, 试比较甲乙两台机床加工的质量有无显著差异 ($\alpha = 0.05$)?

19. 假设甲乙两煤矿的含灰率分别服从正态分布 $N(a, \sigma^2)$ 和 $N(b, \sigma^2)$. 为检验这两个煤矿的煤的含灰率有无显著差异, 从两矿中各取若干份, 分析结果 (单位: %) 为

甲矿: 24.3, 18.8, 22.7, 19.3, 20.4.

乙矿: 25.2, 28.9, 24.2, 26.7, 22.3, 20.4.

试在水平 $\alpha = 0.05$ 之下, 检验 “含灰量无差异” 这个假设.

20. 两台机床加工同一零件, 分别取 6 个和 9 个零件, 量其长度 X_i ($i = 1, \dots, 6$) 和 Y_j ($j = 1, \dots, 9$) 后算得 $Q_1^2 = \sum_{i=1}^6 (X_i - \bar{X})^2 = 1.725$, $Q_2^2 = \sum_{j=1}^9 (Y_j - \bar{Y})^2 = 2.856$. 假定零件长度服从正态分布, 问是否可认为两台机床加工的零件的方差无显著差异 ($\alpha = 0.02$)?

21. 两位化验员 A, B 对一种矿砂的含铁量各自独立地用同一种方法作了 5 次分析, 得到样本方差分别为 0.4322 与 0.5006, 若假定 A, B 测定值的总体都是正态分布, 其方差分别为 σ_A^2 , σ_B^2 , 试在水平 $\alpha = 0.10$ 下检验 $H_0: \sigma_A^2 = \sigma_B^2$.

22. 为研究矽肺患者功能的变化情况, 某医院对 I, II 期矽肺患者各 33 名测定其肺活量, 得到 I 期患者的平均数为 2710 ml, 标准差为 147 ml, II 期患者的平均数为 2830 ml, 标准差为 118 ml, 对水平 $\alpha = 0.10$, 试问第 I, II 期患者的肺活量有无显著差异 (假定肺活量服从正态分布)?

23. 为了比较两种枪弹的速度 (单位: 米/秒), 在相同条件下进行速度测定. 算得样本均值和样本标准差如下:

枪弹甲: $n_1 = 120$, $\bar{x} = 2805$, $s_1 = 120.41$;

枪弹乙: $n_2 = 60$, $\bar{y} = 2680$, $s_2 = 90.00$.

问在水平 $\alpha = 0.05$ 下, 两种枪弹的速度和均匀性方面是否有显著差异?

24. 在 10 块土地上同时试种甲、乙两个品种的农作物, 假定每种作物的产量服从正态分布, 并算得样本均值和样本标准差如下: $\bar{X} = 30.97$, $\bar{Y} = 21.79$, $S_x = 26.7$, $S_y = 10.1$, 问这两种作物的产量有无显著差异? ($\alpha = 0.02$, 提示: 用 Behrens-Fisher 小样本方法的近似解)

25. 为确定肥料的效果, 取 1000 株植物做试验, 在没有施肥的 100 株植物中有 53 株长势良好, 在已施肥的 900 株中, 则有 783 株长势良好, 问施肥的效果是否显著 ($\alpha = 0.01$)?

26. 设有两工厂生产的一种产品, 要检验假设 H : 它们的废品率 p_1, p_2 相同, 在第一、二工厂的产品中各自独立抽取 $n_1 = 1500$ 个及 $n_2 = 1800$ 个, 分别有废品 300 个及 320 个, 问在 5% 水平上应接受还是拒绝 H .

27. 设 X_1, \dots, X_m i.i.d. \sim Poisson 分布 $P(\lambda_1)$, Y_1, \dots, Y_n i.i.d. \sim Poisson 分布 $P(\lambda_2)$, 且样本 X_1, \dots, X_m 和 Y_1, \dots, Y_n 独立. 用大样本方法求列检验问题 $H_0: \lambda_2 - \lambda_1 = 0 \leftrightarrow H_1: \lambda_2 - \lambda_1 \neq 0$, 检验水平 α 给定.

28. X_1, \dots, X_m i.i.d. $\sim b(1, p_1)$, Y_1, \dots, Y_n i.i.d. $\sim b(1, p_2)$, 两组样本独立, 用大样本方法求检验

$$H_0 : p_2 - p_1 = p_0 \leftrightarrow H : p_2 - p_1 \neq p_0$$

检验水平 α 和 p_0 给定.

29. 设 X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma^2)$, 其中 σ^2 已知, 求检验问题: $H_0 : \mu = 0 \leftrightarrow H_1 : \mu \neq 0$ 的似然比检验.

30. 设 X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma^2)$, $-\infty < \mu < +\infty$, $\sigma^2 > 0$ 都是未知参数, 试求检验问题 $H_0 : \sigma^2 \leq \sigma_0^2 \leftrightarrow H_1 : \sigma^2 > \sigma_0^2$ 水平为 α 的似然比检验.

31. 设 X_1, \dots, X_m i.i.d. $\sim N(\mu_1, \sigma^2)$, Y_1, \dots, Y_n i.i.d. $\sim N(\mu_2, \sigma^2)$, 且样本 X_1, \dots, X_m , 和 Y_1, \dots, Y_n 独立, 求下列检验问题水平为 α 的似然比检验:

$$H_0 : \mu_2 - \mu_1 = 0 \leftrightarrow H_1 : \mu_2 - \mu_1 \neq 0.$$

32. 设 X_1, \dots, X_m i.i.d. $\sim N(\mu_1, \sigma_1^2)$, Y_1, \dots, Y_n i.i.d. $\sim N(\mu_2, \sigma_2^2)$, 且 X_1, \dots, X_m 与 Y_1, \dots, Y_n 独立, 试求 $H : \sigma_1^2 = \sigma_2^2 \leftrightarrow K : \sigma_1^2 \neq \sigma_2^2$ 的似然比检验 (取检验水平为 α).

33. 从 k 个具有同方差的正态总体中各取一容量为 n 的样本, 导出检验假设“所有平均数都为 0”的似然比判别式, 证明此判别式是具有 F 分布比的函数.

34. 在正态两样本情况下, 给定 $c > 0$, 找出 $H_0 : \sigma_1^2 = c\sigma_2^2 \leftrightarrow H_1 : \sigma_1^2 \neq c\sigma_2^2$ 的水平为 α 的似然比检验.

35. 设 X_1, \dots, X_n i.i.d. \sim 指数分布 $\text{Exp}(\lambda)$, 求 (1) $H_0 : \lambda = \lambda_0 \leftrightarrow H_1 : \lambda \neq \lambda_0$ 和 (2) $H_0 : \lambda \leq \lambda_0 \leftrightarrow H_1 : \lambda > \lambda_0$ 的水平为 α 的似然比检验.

36. 设 X_1, \dots, X_n 为取自下列指数分布总体的样本:

$$f(x, \mu) = \exp\{-(x - \mu)\}, \quad x \geq \mu, \quad -\infty < \mu < \infty.$$

求检验问题 $H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0$ 的水平为 α 的似然比检验.

37. 设 X_1, \dots, X_m 为从指数分布 $f(x, \theta_1) = \theta_1^{-1} \exp\{-x/\theta_1\}$ 抽取的简单样本, Y_1, \dots, Y_n 为从指数分布 $f(x, \theta_2) = \theta_2^{-1} \exp\{-x/\theta_2\}$ 抽取的简单样本, 且两组样本独立, θ_1 和 θ_2 为未知参数. 求检验问题 $H_0 : \theta_1 = \theta_2 \leftrightarrow H_1 : \theta_1 \neq \theta_2$ 的似然比检验.

38. 设 X_1, X_2, \dots, X_n i.i.d. $\sim B(1, p)$, 用大样本方法求检验问题 $H : p = p_0 \leftrightarrow K : p \neq p_0$ 的似然比检验 (取水平为 α).

39. 设 X_1, X_2, \dots, X_n i.i.d. 服从成功概率为 p 的几何分布, 用大样本方法求检验问题 $H_0 : p = p_0 \leftrightarrow H_1 : p \neq p_0$ 的检验水平为 α 的似然比检验.

40. 设 $0 < \alpha < 1$, φ 是检验问题 $H : \theta = \theta_0 \leftrightarrow K : \theta = \theta_1$ 的水平为 α 的 UMPT, 且假定 $\beta = E_K[\varphi(X)] < 1$, 则 $1 - \varphi$ 是检验问题 K 对 H 的水平为 $1 - \beta$ 的 UMPT.

41. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为自总体分布族 $\{f(x, \theta), \theta \in \Theta\}$ 中抽取的随机样本, 且 $T = T(X_1, \dots, X_n)$ 为 θ 的充分统计量, 则由 NP 引理确定的检验函数 $\varphi(x)$ 为充分统计量 T 的函数.

42. 设 X_1, \dots, X_n i.i.d. 服从均匀分布 $U(0, \theta)$, 试求
 (1) 检验问题 $H_0 : \theta = 1 \leftrightarrow H_1 : \theta = 2$ 检验水平为 α 的 UMPT;
 (2) 检验问题 $H'_0 : \theta = 1 \leftrightarrow H'_1 : \theta = 1/2$ 检验水平为 α 的 UMPT.
43. 设 X_1, \dots, X_n i.i.d. $\sim N(\mu, 1)$, 对水平 α , 试求检验问题 $H_0 : \mu \geq \mu_0 \leftrightarrow H_1 : \mu < \mu_0$ 的 UMPT.
44. 设 X_1, \dots, X_n i.i.d. $\sim \text{Poisson}$ 分布 $P(\lambda)$, 对水平 α , 试求检验问题 $H_0 : \lambda \geq \lambda_0 \leftrightarrow H_1 : \lambda < \lambda_0$ 的 UMPT.
45. 设 X_1, \dots, X_n i.i.d. $\sim N(0, \sigma^2)$, 对水平 α , 试求检验问题 $H_0 : \sigma^2 \leq \sigma_0^2 \leftrightarrow H_1 : \sigma^2 > \sigma_0^2$ 的 UMPT.
46. 设 X_1, \dots, X_n i.i.d. $\sim b(1, p)$, 对水平 α , 试求检验问题 $H_0 : p \geq 1/2 \leftrightarrow H_1 : p < 1/2$ 的 UMPT.
47. 设 k 为已知的自然数. 为检验一事件的概率 $p \leq p_0$ 是否成立, 将试验独立地重复下去, 直到该事件发生 k 次为止. 以 X 记到那时为止的试验次数.
 (1) 证明 X 的分布为
- $$P(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \quad x = k, k+1, \dots;$$
- (2) 求 $H_0 : p \leq p_0 \leftrightarrow H_1 : p > p_0$ 水平为 α 的 UMPT, 其中 p_0 和 α 给定.
48. 设 X_1, \dots, X_n 为自指数分布总体 $\text{Exp}(\lambda)$ 中抽取的随机样本, $\lambda > 0$ 为未知参数. 求 $H_0 : \lambda \geq \lambda_0 \leftrightarrow H_1 : \lambda < \lambda_0$ 的水为 α 的 UMPT, 其中 λ_0 和 α 给定.
49. 设 X_1, \dots, X_n 表示受试的 n 个元件寿命, 它们皆服从指数分布 $\text{Exp}(\lambda)$. 而实际上只试验到有 r 个失效时即停止. 这 r 个寿命从小到大排列为 $X_{(1)} \leq \dots \leq X_{(r)}$. 试求 $H_0 : \lambda \leq \lambda_0 \leftrightarrow H_1 : \lambda > \lambda_0$ 的基于 $X_{(1)} \dots X_{(r)}$ 的 UMPT (提示: 由例 3.3.5 可证 $T = \sum_{i=1}^r X_{(i)} + (n-r)X_{(r)}$ 为充分统计量, 且 $2\lambda T \sim \chi^2_{2r}$).
50. 设 $\varphi(\mathbf{x})$ 为一检验函数, 而 $T = T(\mathbf{X})$ 为充分统计量. 证明 $E_\theta(\varphi(\mathbf{X}) | T)$ 也是检验函数, 且其功效函数与 φ 同. 这个结果说明了什么问题? (提示: 由充分性知 $E_\theta(\varphi(\mathbf{X}) | T)$ 与 θ 无关, 再由 φ 是检验函数, 证明 $E_\theta(\varphi(\mathbf{X}) | T)$ 也是检验函数.)
51. 设 X_1, \dots, X_m i.i.d. $\sim N(a, 1)$, Y_1, \dots, Y_n i.i.d. $\sim N(b, 1)$, 且样本 X_1, \dots, X_m 和 Y_1, \dots, Y_n 独立, 证明 $H_0 : b \leq a \leftrightarrow H_1 : b > a$ 的 UMPT 存在, 否定域为 $\bar{Y} - \bar{X} > c$ (提示: 利用上题结果).
52. 设 X_1, \dots, X_n i.i.d. \sim 均匀分布 $U(0, \theta)$, 对水平 α , 试求检验问题 $H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0$ 的 UMPT, 且求出该 UMPT 的功效函数.
- *53. 设 X_1, \dots, X_n i.i.d. \sim 均匀分布 $U(\theta, 2\theta)$, $\theta > 0$. 指定 $\theta_0 > 0$, 问 $H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0$ 的 UMPT 是否存在? (提示: 取 $\theta_1 > \theta_0$, 考虑检验问题 $H'_0 : \theta = \theta_0 \leftrightarrow H'_1 : \theta = \theta_1$, 证明所得水平 α 的 UMPT 与 θ_1 无关, 故它也是 $H'_0 \leftrightarrow H'_1$ 的 UMPT. 再证明功效函数 $\beta(\theta)$ 为单调增函数.)
- *54. 设 X_1, \dots, X_n i.i.d. 服从下列分布: $f(x) = e^{-(x-\theta)} I_{(\theta, \infty)}(x)$, 试求检验问题 $H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$ 水平 α 的 UMPT.

- *55. 设 X_1, \dots, X_n i.i.d. \sim 均匀分布 $U(0, \theta)$, 指定 θ_0 . 证明 $H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$ 的 UMPT 存在 (提示: 取 $\theta_1 > \theta_0$ 和 $\theta_1 < \theta_0$ 分别考虑检验问题 $H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta = \theta_1$).
56. 设 X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma^2)$, μ 和 σ^2 皆未知, 利用假设检验方法导出 σ^2 的置信系数为 $1 - \alpha$ 的置信区间和置信上、下限.
57. 设 X_1, \dots, X_m i.i.d. $\sim N(\mu_1, \sigma_1^2)$, Y_1, \dots, Y_n i.i.d. $\sim N(\mu_2, \sigma_2^2)$, 且样本 X_1, \dots, X_m 和 Y_1, \dots, Y_n 独立. 利用假设检验方法求出 σ_1^2/σ_2^2 的置信系数为 $1 - \alpha$ 的置信区间和置信上、下限.

第6章 非参数假设检验

6.1 引言

在前几章讨论的统计问题大多属于这样一种情形：样本分布族的数学形式已知（如正态、指数、Poisson 分布等），但其中包含有限个未知的实参数。最典型的情况是关于正态分布族的均值、方差的估计和检验问题，这种统计问题称为参数型的。

然而，在实际问题中也有这样的情况：对样本分布族并未给出其数学形式，而只作了某些一般的假定，如只假定总体分布的连续性或对称性，而对其分布的数学形式则一无所知。类似这种情形的统计问题称为非参数型的。

如果关于总体分布族事先有比较多的信息，那么把统计推断问题放在参数型的范围内讨论是恰当的。如果事先关于总体分布族只有很少信息，此时就需要一种与总体分布族的具体数学形式无关的统计推断方法。这种方法称为非参数方法。

非参数方法具有以下几个特点：① 适用面广而针对性较差。这是由于非参数方法对总体分布族没有提出太多的假设。但是，如果对特定的统计推断问题，参数模型适合该问题，且存在一个具有优良性质的参数型统计方法，则二者相比，非参数统计方法一般效率较低。② 大样本理论在非参数统计中起着极其重要的作用。这是因为对总体分布族知之甚少，很难指望导出有关统计量的精确分布，迫使考虑用统计量的极限分布近似代替其精确分布。③ 只能使用样本中的“一般”信息，如位置、次序关系等。这也是由于对总体分布族所知甚少之故。④ 非参数方法具有稳健性，即当真实模型与设定的理论模型有一定偏离时，能维持较好的性质。这是由于非参数方法对模型限制小，故自然就具有稳健性（robustness）。

参数型统计方法往往较易使用，针对性强，常常具有更好的性质。所以参数和非参数方法是针对不同情况提出来的，它们各有优点，是互为补充的。

6.2 节和 6.3 节将讨论一样本和两样本问题中常用的几种非参数假设检验方法，如符号检验、符号秩和检验、Wilcoxon 秩和检验。拟合优度检验问题也是非参数型的。因为在此问题中，理论分布是一个待检验的假设而并非假定。6.4 节将讨论拟合优度检验（也称为 Pearson χ^2 检验）。作为其应用，6.5 节讨论列联表的独立性和齐一性检验。本章最后一节简要介绍有关分布检验的其他几种方法，如柯尔莫哥洛夫（Kolmogorov）检验、斯米尔诺夫（Smirnov）检验和正态性检验方法等。

6.2 符号检验及符号秩和检验

第 5 章讨论了当总体分布族是正态情形, 关于总体均值的一样本 t 检验方法. 但是, 当无把握认为总体分布族为正态模型时, 则必须用其他方法来检验. 下面介绍两种常用的非参数检验方法, 即符号检验法、符号秩和检验法.

6.2.1 符号检验法

例 6.2.1 为比较甲乙两种酒的优劣, 找了 N 个人去品尝. 同一个人品尝两种酒后, 请他们分别给两种酒评分. 这里, 每一个品酒人对甲乙两种酒的评分结果构成一个对子, 正好是一个成对比较的模型.

以 X_i 记第 i 个品酒人对甲酒的评分, Y_i 记第 i 个品酒人对乙酒的评分. 记 $Z_i = X_i - Y_i$, $i = 1, \dots, N$. 如果假定 $Z_i \sim N(\mu, \sigma^2)$, 则甲、乙两酒是否有优劣的问题将转化为原假设 $H_0: \mu = 0 \leftrightarrow H_1: \mu \neq 0$ 的检验问题, 这就是在 5.2 节讨论过的一样本 t 检验问题. 可是在一些情况下, 不见得有根据假定 Z_i 服从正态分布. 这时上述方法就失效了. 下面是一个替代方法: 对每一个评酒人的评分给出一个符号

$$S_i = \begin{cases} +, & Z_i > 0, \\ -, & Z_i < 0, \\ 0, & Z_i = 0, \end{cases} \quad (6.2.1)$$

即品酒人给以“+”号表示他认为“甲酒优于乙酒”, 另两个符号的意义类推. 如此, 得到 N 个符号 S_1, \dots, S_N . 检验问题

$$H_0: \text{甲乙两种酒一样好} \leftrightarrow H_1: \text{甲乙两种酒不一样} \quad (6.2.2)$$

的检验就建立在试验结果的这 N 个符号的基础上, 故称为符号检验 (sign test). 下面将会看到: 从统计模型而言, 符号检验不过是二项分布参数检验的一个特例. 符号检验的具体方法如下.

1. 小样本方法

记 N 个试验结果 S_1, \dots, S_N 中“+”号的个数有 n_+ 个, 出现“-”号的个数有 n_- 个, 其余为 0. 记 $n = n_+ + n_-$. 如果 H_0 成立, 即甲乙两种酒一样好, 则在 n 个非 0 结果中出现“+”或“-”的机会相同, 即每个非 0 试验结果中出现“+”号的概率 $\theta = 1/2$; 故在这个情况下, n_+ 的分布服从 $b(n, 1/2)$. 若甲乙两种酒确有优劣之分, 则每个结果出现“+”号的概率 $\theta \neq 1/2$. 若记 $X = n_+$, 则所提问题转化为检验问题 $X \sim b(n, \theta)$, $0 \leq \theta \leq 1$, 要检验

$$H_0: \theta = \frac{1}{2} \leftrightarrow H_1: \theta \neq \frac{1}{2}. \quad (6.2.3)$$

一个水平为 α 的检验的否定域为

$$\left\{ X = n_+ \geq c \text{ 或 } X \leq d \right\},$$

其中 c 和 d 的值由下式确定:

$$\sum_{i=c}^n \binom{n}{i} \left(\frac{1}{2}\right)^n \leq \frac{\alpha}{2}, \quad d = n - c,$$

其中 c 的值已制定成表, 见附表 10. 可以通过查附表 10 直接获得 c 的值, 再按公式计算得到 d 的值.

一个更恰当的方法是计算检验的 p 值 (见 5.5.5 小节). 在此, 令由符号 S_1, \dots, S_N 算得的 $X = n_+$ 的具体值为 x_0 , 记 $x'_0 = \min\{x_0, n - x_0\}$, 则检验的 p 值为

$$p = \sum_{i=0}^{x'_0} \binom{n}{i} \left(\frac{1}{2}\right)^n + \sum_{i=n-x'_0}^n \binom{n}{i} \left(\frac{1}{2}\right)^n. \quad (6.2.4)$$

若 n 为偶数, 而 $x_0 = n/2$, 则取 p 值为 $p = 1$. p 值越接近 1, 则 H_0 越可信. 例如, 给定检验水平 α , 则当 $p < \alpha$ 时否定 H_0 , 当 $p \geq \alpha$ 时接受 H_0 .

在例 6.2.1 中, 令 $N = 13$, S_1, \dots, S_{13} 中 + 号和 - 号的个数分别是 $n_+ = 2$, $n_- = 10$, 因此 $n = n_+ + n_- = 12$. 取检验水平 $\alpha = 0.10$, 查附表 10 “符号检验临界值表”得 $c = 10$, 故 $d = n - c = 2$. 故检验的否定域 $D = \{X = n_+ \geq 10 \text{ 或 } X \leq 2\}$. 检验统计量 $X = n_+ = 2$, 因此否定原假设, 即认为甲、乙两酒不一样.

对本例中的检验问题, 也可通过计算检验的 p 值来作出结论. 此处, $n = 12$, $x_0 = n_+ = 2$, 按式 (6.2.4), $x'_0 = \min(2, 12 - 2) = 2$, 查二项分布表得

$$p = \sum_{i=0}^2 \binom{12}{i} \left(\frac{1}{2}\right)^{12} + \sum_{i=10}^{12} \binom{12}{i} \left(\frac{1}{2}\right)^{12} = 0.0384 < 0.10,$$

故在 0.10 显著性水平下应否定 H_0 .

2. 大样本方法

对检验问题 (6.2.2) 或等价地对检验问题 (6.2.3), 若 n 很大, 则可用大样本方法: 由二项分布的中心极限定理可知当 H_0 成立且 $n \rightarrow \infty$ 时有

$$U = \frac{X - E(X)}{\sqrt{D(X)}} = \frac{X - n/2}{\sqrt{n/4}} = \frac{2X - n}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0, 1).$$

因此检验问题 (6.2.2) 水平近似为 α 的检验的否定域是

$$\left\{ X : |U| > u_{\alpha/2} \right\}, \quad (6.2.5)$$

其中 $u_{\alpha/2}$ 为标准正态分布的上侧 $\alpha/2$ 分位数.

有时检验的目的是从“甲不优于乙”和“甲优于乙”中选择一个, 以前者为原假设, 则检验问题可表示为 $X \sim b(n, \theta)$, $0 \leq \theta \leq 1$, 而

$$H'_0 : \theta \leq \frac{1}{2} \leftrightarrow H'_1 : \theta > \frac{1}{2}.$$

这种问题在例 5.4.5 中已给出其 UMPT. 当 n 充分大时, 可用大样本方法, 其检验水平近似为 α 的检验的否定域是

$$\left\{ X : U > u_\alpha \right\}.$$

例 6.2.2 一种饮料有传统配方 (甲) 及修改配方 (乙) 两个品种, 都在市场上出售. 制造商为了解公众的反映, 寄出大量征求意见函, 结果回收 10000 份, 其中认为传统配方优于修改配方的有 5150 人, 认为修改配方优于传统配方的有 4850 人. 根据这一调查结果, 可作出怎样的结论?

解 设检验问题为

$$H_0 : \text{甲乙两种配方一样} \leftrightarrow H_1 : \text{甲乙两种配方不一样},$$

此处 $N = n = 10000$, $X = n_+ = 5150$. 由于 n 很大, 适合用大样本方法. 由正态近似计算出

$$U = \frac{2X - n}{\sqrt{n}} = \frac{10300 - 10000}{100} = 3.$$

按式 (6.2.4) 计算 p 值, 查标准正态表得

$$p = P(|U| \geq 3 | H_0) = 2(1 - 0.9987) = 0.0026.$$

若取检验水平 $\alpha = 0.01$, 则 $p = 0.0026 < 0.01$, 故在 0.01 显著性水平下也要否定 H_0 , 即认为调查结果充分显示两种配方有显著差异, 传统配方较优.

由观察数据, 估计出传统配方的支持率为 $5150/10000 = 0.515$, 比 0.5 大得并不多, 可是由于样本容量 n 很大, 这个 0.015 的差异就有机会显示出来.

例 6.2.3 工厂的两个化验室, 每天同时从工厂的冷却水中取样, 测量水中的含氯量一次. 表 6.2.1 是 $n = 11$ 天的记录:

表 6.2.1

i	1	2	3	4	5	6	7	8	9	10	11
x_i	1.15	1.86	0.76	1.82	1.14	1.65	1.92	1.01	1.12	0.90	1.40
y_i	1.00	1.90	0.90	1.80	1.20	1.70	1.95	1.02	1.23	0.97	1.52

其中 x_i 表示化验室 A 的测量记录, y_i 表示化验室 B 的测量记录. 问两个化验室测定的结果之间有无显著差异 (取 $\alpha = 0.10$)?

解 分别记化验室 A 和 B 的测量误差为 ξ 和 η . 设 ξ 和 η 为连续随机变量, 其分布函数分别为 F 和 G . 检验问题是

$$H_0 : F = G \leftrightarrow H_1 : F \neq G. \quad (6.2.6)$$

显然含氯量的测定值, 除了与化验室的不同有关外, 还与当日水中含氯量的多少有关. 可以认为 X_i 和 Y_i 具有数据结构

$$X_i = \mu_i + \xi_i, \quad Y_i = \mu_i + \eta_i, \quad i = 1, 2, \dots, n,$$

其中 μ_i 为第 i 天水中的含氯量, ξ_i 和 η_i 分别表示第 i 天化验室 A, B 的测量误差. 显然 ξ_1, \dots, ξ_n 和 η_1, \dots, η_n 都是不可观察的独立同分布的随机变量. 前者与 $\xi \sim F$ 同分布, 后者与 $\eta \sim G$ 同分布.

不同日的两个数据 X_i 与 Y_i 显然不一定是同分布的, 而且 X_i 与 X_j 以及 Y_i 与 Y_j 也不一定是同分布的. 它们之间的差异不但与测量误差有关, 而且也与 μ_i 和 μ_j 的差异有关. 因此虽然 X_1, \dots, X_n 相互独立, 但不能假定它们同分布, Y_1, \dots, Y_n 也是如此. 所以两样本的统计比较方法, 如两样本的 t 检验方法以及后面要介绍的两样本非参数检验方法都不能用于这类数据的检验工作. 在 5.2 节中也提到过成对数据的上述特点.

处理成对数据检验问题, 很自然地想到如何把 μ_i 的影响消除掉. 由于对每个 i , X_i 与 Y_i 之间可比, 若将同一天的两个数据相减, 从而把 μ_i 的影响消除掉. 令

$$Z_i = X_i - Y_i = \xi_i - \eta_i, \quad i = 1, 2, \dots, n. \quad (6.2.7)$$

显然 Z_i 仅与化验室 A, B 在第 i 日的测量误差之差有关. 记 $Z = \xi - \eta$, 则 Z_1, \dots, Z_n 可看成来自总体 Z 的随机样本, 即 Z_1, \dots, Z_n 是独立同分布的样本. 由于 Z 是两个测量误差之差, 所以 Z 的均值为 0, 且可证明它是关于原点对称的.

令 n_+ 为 Z_1, \dots, Z_n 中取正值的个数, n_- 为 Z_1, \dots, Z_n 中取负值的个数, 它们都是随机变量. 由于假定了 ξ 和 η 是连续随机变量, 故 Z_1, \dots, Z_n 这些变量以概率为 1 取非 0 值. 因此可记 $n = n_+ + n_-$. 当 H_0 , 即式 (6.2.6) 成立时, 则在 n 个试验单元中 Z_i 取 “+” 和取 “-” 的可能性皆为 $1/2$. 因此检验问题转化为 $X = n_+ \sim b(n, \theta)$, $0 \leq \theta \leq 1$, 检验问题为

$$H'_0 : \theta = \frac{1}{2} \leftrightarrow H'_1 : \theta \neq \frac{1}{2},$$

否定域 $D = \{n_+ \geq c \text{ 或 } n_+ \leq d\}$.

因此, 在给定显著性水平 α 之后, c 和 d 的值由

$$\sum_{k=c}^n \binom{n}{k} \left(\frac{1}{2}\right)^n \leq \frac{\alpha}{2}, \quad d = n - c$$

所确定. 也可以通过查附表 10 直接获得 c 的值, 再按公式计算得到 d 的值.

在本例中 $n = 11$, $\alpha = 0.10$, 查附表 10 得 $c = 9$, $d = n - c = 2$. 故水平 $\alpha = 0.10$ 的符号检验的否定域为

$$\{n_+ \leq 2 \text{ 或 } n_+ \geq 9\}.$$

作差值 $z_i = x_i - y_i$, 得

$$\begin{aligned} & 0.15, -0.04, -0.14, 0.02, -0.06, -0.05, \\ & -0.03, -0.01, -0.11, -0.07, -0.12, \end{aligned}$$

其中取正数的个数为 $n_+ = 2$, 因此在水平 $\alpha = 0.10$ 下否定 H_0 , 即认为化验室 A, B 测定结果之间有显著差异.

对这一检验问题, 也可以通过计算检验的 p 值来解决:

$$p = \sum_{i=0}^2 \binom{11}{i} \left(\frac{1}{2}\right)^{11} + \sum_{i=9}^{11} \binom{11}{i} \left(\frac{1}{2}\right)^{11} = 0.0654 < \alpha = 0.10,$$

故否定 H_0 .

符号检验的另一个重要应用是分位数(特别是中位数)检验. 请看下例.

例 6.2.4 检验某种维尼纶的纤度, 测得 100 个数据见表 6.2.2.

表 6.2.2

编号	1	2	3	4	5	6	7	8	9	10
纤度	1.26	1.29	1.32	1.35	1.38	1.41	1.44	1.47	1.50	1.53
频数	1	4	7	22	23	25	10	6	1	1

试问该维尼纶纤度的中位数 m_e 是否为 1.40 ($\alpha = 0.05$)?

解 本题在显著水平 $\alpha = 0.05$ 下, 检验假设

$$H_0 : m_e = 1.40 \leftrightarrow H_1 : m_e \neq 1.40.$$

若令表中所列 100 个数据的纤度值为 X_i , $i = 1, \dots, 100$, 令 $Y_i = X_i - 1.40$, $i = 1, \dots, 100$. 计算 Y_i 取正值的个数 n_+ 和取负值的个数 n_- , 取值为 0 的个数为 0, 因此 $n_+ + n_- = 100$. 在 H_0 成立的前提下, 则每个 Y_i 为正或负的可能性皆为 $1/2$, 故 100 个数据中 n_+ 和 n_- 应差别不大, 若记 $X = n_+$, 易见 $X \sim b(100, 1/2)$, 因此检验问题转化为 $X \sim b(100, \theta)$, $0 \leq \theta \leq 1$, 要检验

$$H_0: \theta = \frac{1}{2} \leftrightarrow H_1: \theta \neq \frac{1}{2}, \quad \alpha = 0.05.$$

由于样本容量 n 较大, 采用大样本方法. 由式 (6.2.5) 可知检验水平近似为 α 的否定域是

$$\left\{ X : |U| = \left| \frac{2X - n}{\sqrt{n}} \right| > u_{\alpha/2} \right\}.$$

取 $n = 100$, $\alpha = 0.05$, $u_{0.025} = 1.96$, $X = n_+ = 43$, 可算得

$$\left| \frac{2X - n}{\sqrt{n}} \right| = 1.4 < 1.96.$$

故不足以否定 H_0 , 即认为该维尼纶的纤维度的中位数是 1.40.

本例中也可以通过计算大样本检验的 p 值作出结论. 由样本算得 $|U|$ 的具体值 1.4, 查标准正态分布表得

$$P(|U| > 1.4 | H_0) = 0.1616 > \alpha,$$

不足以否定 H_0 , 故接受 H_0 .

6.2.2 符号秩和检验

再回顾一下符号检验, 仍就例 6.2.1 中品酒的问题来说明. 在计算 $Z_i = X_i - Y_i$ 后, 放弃 Z_i 的具体数值而取其符号 S_i 时, 丢失了一些信息. 这种信息的丢失, 使符号检验的效率有所降低. 为此提出了符号秩和检验, 它是符号检验的改进.

例 6.2.5 (续例 6.2.1) 仍看例 6.2.1, 设想请了 13 个人品尝甲、乙两种酒, 评分结果见表 6.2.3.

表 6.2.3

品酒人	1	2	3	4	5	6	7	8	9	10	11	12	13
甲 (x_i)	55	32	41	50.5	60	48	39	45	48	46	52.2	45	44
乙 (y_i)	35	37	43.1	55	34	50.3	43	46.1	51	47.3	55	46.5	44
符号 (z_i)	+	-	-	-	+	-	-	-	-	-	-	-	0

此处 $z_i = x_i - y_i$. 试问甲乙两种酒是否一样好? 一共 12 个非 0 符号中, 有两个“+”号, 显示多数品酒人认为乙酒好. 在符号检验中就只能根据“+”“-”号的数目去下结论. 但细看一下结果, 发现, 在认为“乙酒比甲酒优”的 10 人中, 乙酒的得分比甲酒高得不多, 而在认为“甲酒优于乙酒”的 2 人中, 甲的得分远远高于乙. 这个事实给 2:10 这个表面结果打了一个折扣, 它启示: 除了考虑符号外, 还应当把这一点考虑进来. 符号秩和的概念为此提供了一种有效的方法.

定义 6.2.1 设 X_1, \dots, X_n 为两两不相等的一组样本, 将其按大小排列为 $X_{(1)} < \dots < X_{(n)}$, 若 $X_i = X_{(R_i)}$, 则称 X_i 在样本 (X_1, \dots, X_n) 中的秩为 R_i .

显然, 若 X_1, \dots, X_n 为来自连续分布 $F(x)$ 的样本, 则以概率为 1 保证 X_1, \dots, X_n 中两两互不相等.

定义 6.2.2 设 X_1, \dots, X_n 为来自单个总体的样本, 或来自多个总体样本的合样本, 则 $\mathbf{R} = (R_1, R_2, \dots, R_n)$ 称为 (X_1, \dots, X_n) 的秩统计量 (rank statistics), 其中 R_i 为 X_i 的秩. 由 \mathbf{R} 导出的统计量也称为秩统计量. 基于秩统计量的检验方法称为秩检验 (rank test).

1. 小样本方法

现在仍回到例 6.2.5, 把表 6.2.3 扩充成表 6.2.4, 把符号为“+”的那两个秩 (即 11 和 12) 括起来, 它们的和是 $W^+ = 11 + 12 = 23$, 称为“符号秩和”. 一般它可以用下列方式来定义: 记 $Z_i = X_i - Y_i$, 令

$$\bar{V}_i = \begin{cases} 1, & Z_i > 0, \\ 0, & \text{其他.} \end{cases}$$

R_i 为 $|Z_i|$ 在 $(|Z_1|, \dots, |Z_n|)$ 中的秩, 则 Wilcoxon 符号秩和 (the sum of Wilcoxon signed rank) 检验统计量定义为

$$W^+ = \sum_{i=1}^n \bar{V}_i R_i. \quad (6.2.8)$$

表 6.2.4

品酒人 (i)	甲 (x_i)	乙 (y_i)	符号 (z_i)	$ Z_i = x_i - y_i $	秩
1	55	35	+	20	[11]
2	32	37	-	5	10
3	41	43.1	-	2.1	4
4	50.5	55	-	4.5	9
5	60	34	+	26	[12]
6	48	50.3	-	2.3	5
7	39	43	-	4	8
8	45	46.1	-	1.1	1
9	48	51	-	3	7
10	46	47.3	-	1.3	2
11	52.2	55	-	2.8	6
12	45	46.5	-	1.5	3
13	44	44	0	0	不定秩

容易理解: 在例 6.2.5 中, 若甲优于乙, 则不仅“+”号会多, 且“+”号观察值相应的秩, 一般也偏大, 故总的效果是 W^+ 应偏大. 反之, 若乙优于甲, 则 W^+ 将偏小. 因此检验问题 (6.2.2), 即

$$H_0: \text{甲、乙两种酒一样好} \leftrightarrow H_1: \text{甲、乙两种酒不一样}$$

成立时, W^+ 应当不大不小. 检验的否定域是

$$\{W^+ \leq d \text{ 或 } W^+ \geq c\}, \quad (6.2.9)$$

此处 d 和 c 取决于 n (本例中 $n = 12$) 及指定的检验水平 α , 即当给定 α 时, c, d 分别由下列两式决定:

$$P(W^+ \leq d | H_0) \leq \alpha/2, \quad P(W^+ \geq c | H_0) \leq \alpha/2.$$

H_0 为真时 W^+ 的分布见文献 [4] 第 246 页. 对某些特定的 α 及不大的 n , c 和 d 可以查附表 11 求得, 在附表 11 中仅可查到 c , 而 $d = n(n+1)/2 - c$.

在例 6.2.5 中, 由表 6.2.2 可知本题中 $n = 12$, $W^+ = 23$. 取 $\alpha = 0.10$, 查附表 11 中 $\alpha/2$ 那一栏, 在 $n = 12$ 处得 $c = 61$, 算得 $d = 17$, 按式 (6.2.9) 得否定域为

$$\{W^+ \leq 17 \text{ 或 } W^+ \geq 61\}.$$

而 $17 < W^+ = 23 < 61$, 故应接受 H_0 , 即所得观察结果不构成甲、乙有优劣之分的充分证据.

这个检验称为 Wilcoxon 双侧符号秩和检验 (以下简称双侧 W^+ 检验), 之所以取 $\alpha/2$, 也是由于这个“双侧”而来.

注 6.2.1 例 6.2.1 和例 6.2.5 中的同一个检验问题用符号检验和符号秩和检验得到两种不同的结论. 按符号检验否定 H_0 , 即认为甲、乙两酒有优劣之分且乙优于甲. 按符号秩和检验的方法, 接受 H_0 , 即表明无充分证据否定“甲、乙两酒一样好”. 这里看到: 同一个问题, 同一批数据, 用不同方法, 检验结果不同, 这不足为怪. 正如用同一批数据去估计正态总体的数学期望值, 用样本均值估计与用中位数估计, 两者结果不同. 这就产生了一个问题: 这两种检验法哪一种好? 这个问题不能一概而论. 要回答这个问题, 必须给出一种准则, 根据它去判定何者为优, 这属于数理统计理论问题. 可以指出的是: 符号检验全然不看数值而只看符号; 基于正态假定的 t 检验则要看数值, W^+ 检验介于二者之间: 它既不忽视数值, 也不全看数值 (数值只用于决定秩, 而不用其本身值), 应当注意这一点.

2. 大样本方法

可以证明

$$E(W^+) = \frac{n(n+1)}{4}, \quad D(W^+) = \frac{1}{24}n(n+1)(2n+1).$$

与 6.3 节的秩和统计量 W 类似, 当 H_0 成立且 $n \rightarrow \infty$ 时, W^+ 的标准化随机变量

$$W_*^+ = \frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \xrightarrow{\mathcal{L}} N(0, 1), \quad (6.2.10)$$

故水平近似为 α 的双侧 W^+ 检验的否定域为

$$\{|W_*^+| > u_{\alpha/2}\}.$$

例 6.2.6 设某工厂甲乙两种不同工艺生产的啤酒, 找了 100 名品酒师品尝, 将两种啤酒得分之差的绝对值按大小排序, 获得乙种啤酒的符号秩和 $W^+ = 2200$. 问这两种啤酒是否有显著差异? ($\alpha = 0.10$)

解 由题意可知检验问题为

$$H_0: \text{甲乙两种啤酒一样好} \leftrightarrow H_1: \text{甲乙两种啤酒不一样好}.$$

由于 n 较大, 可采用大样本方法. 由符号秩和检验的大样本方法可知, 水平近似为 $\alpha = 0.10$ 的否定域由式 (6.2.10) 给出, 即

$$\left\{ |W_*^+| = \left| \frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \right| \geq u_{\alpha/2} \right\}.$$

此处 $W^+ = 2200$, $n = 100$, $u_{0.05} = 1.65$, 因而有

$$|W_*^+| = \left| \frac{2200 - 100 \times 101/4}{\sqrt{100 \times 101 \times 201/24}} \right| = \frac{325}{290.84} = 1.12 < 1.65,$$

故不足以否定 H_0 , 因此接受 H_0 , 即认为两种啤酒无显著差异.

6.3 Wilcoxon 两样本秩和检验

在两样本的比较问题中, 当样本的随机误差不服从正态分布时, 就需要更一般的提法, 并使用相应的非参数检验方法. 这方面的理论和方法较多, 但大都很专业, 这里只对 Wilcoxon 两样本秩和检验作一简略介绍.

6.3.1 引言及定义

首先来看一看这一检验的实际背景. 两样本检验问题的一般提法如下: 设 X_1, \dots, X_m 和 Y_1, \dots, Y_n 分别是从具有分布为 F_1 和 F_2 的一维总体中抽取的简单样本, 且假定样本 X_1, \dots, X_m 和 Y_1, \dots, Y_n 独立. 要检验下列假设:

$$H_0: F_1 = F_2 \leftrightarrow H_1: F_1 \neq F_2. \quad (6.3.1)$$

在数理统计学中, 习惯上称这个检验问题为“两样本问题”. 下面来分别考虑下列几种情况.

(1) 设根据问题的实际背景, 如果有理由假定 F_1 和 F_2 为具有相同方差的正态分布, 即假定

$$F_1 \sim N(a, \sigma^2), \quad F_2 \sim N(b, \sigma^2),$$

其中 a, b 和 σ^2 皆未知, $-\infty < a, b < +\infty$, $\sigma^2 > 0$, 这时检验问题转化为

$$H_0': a = b \leftrightarrow H_1': a \neq b. \quad (6.3.2)$$

在这个假定下, 总体分布 F_1 和 F_2 只依赖于 3 个未知参数 a, b 和 σ^2 , 检验问题 (6.3.1) 归结为检验这些未知参数是否满足式 (6.3.2). 按 5.1 节所述, 这属于“参数型假设检验问题”. 这就是在 5.2 节中讨论的两样本 t 检验.

(2) 如果对问题的实际背景所知甚少, 只好认为对 F_1 和 F_2 完全未知. 在这样宽广的假定下, 再不能使用通常的两样本 t 检验. 处理这个问题的一种方法是斯米尔诺夫检验, 这将在 6.6 节中讨论.

在这一情形, 总体分布 F_1 和 F_2 不能用有限个实参数去刻画, 因此称为非参数检验问题.

(3) 现在讨论一种中间情况. 设 X 是一种产品在某种生产工艺下的质量指标, 而 Y 是该产品在另一种生产工艺下的质量指标. 有理由认为, 改变生产工艺不影响产品质量指标的概率分布, 而只能使此分布发生一些平移. 也就是说, 若以 $F(x)$ 记 X 的分布, 则 Y 分布为 $F(x - \theta)$, 其中 θ 是一个未知的位置参数. 在这个假定下, “ X, Y 同分布”的假设相当“ $\theta = 0$ ”, 而对立假设为“ $\theta \neq 0$ ”. 因此检验式 (6.3.1) 归结为检验

$$H_0: \theta = 0 \leftrightarrow H_1: \theta \neq 0. \quad (6.3.3)$$

式 (6.3.3) 是一个很重要的假设检验问题. 在这一模型中, 假定 F 未知, 因而比正态模型要广. 另外这一模型又比“斯米尔诺夫检验”中的模型窄一些, 因为对后者而言, 两分布 F_1 和 F_2 毫无关系, 而在此 F_1 和 F_2 之间有 $F_2(x) = F_1(x - \theta)$.

虽然表面上看式 (6.3.3) 像一个参数检验问题: 假设中只涉及 θ , 而它是一个实参数. 其实不然, 因为总体的分布与 F 和 θ 都有关, 而 F 的分布未知, 所以按非参数统计问题的定义, 式 (6.3.3) 应视为非参数检验问题.

一般地, 两样本问题 (6.3.1) 还有一些具有实际背景的中间情况. 例如, $F_2(x) = F_1(x/\sigma)$, 此 $\sigma > 0$ 为未知的刻度参数, 分布 F_i , $i = 1, 2$ 也未知. 检验问题 (6.3.1) 在此情况下转化为

$$H_0^*: \sigma = 1 \leftrightarrow H_1^*: \sigma \neq 1. \quad (6.3.4)$$

Wilcoxon 两样本秩和检验就是考虑式 (6.3.3) 的假设检验问题. 下面首先给出 Wilcoxon 两样本秩和统计量的定义.

定义 6.3.1 设 $X_1, \dots, X_m, Y_1, \dots, Y_n$ 这 $n + m$ 个值两两不相同, 把它们按大小排列, 结果为

$$Z_1 < Z_2 < \dots < Z_N, \quad N = m + n. \quad (6.3.5)$$

显然, 每个 Y_i 必为式 (6.3.5) 中的某一个. 若 $Y_i = Z_{R_i}$, 则 Y_i 在合样本 $X_1, \dots, X_m, Y_1, \dots, Y_n$ 中的秩为 R_i . 而 Y_1, \dots, Y_n 的秩和为

$$W = R_1 + \dots + R_n, \quad (6.3.6)$$

它称为 Wilcoxon 两样本秩和统计量. 这是 Wilcoxon 在 1945 年的一项工作中引进的.

6.3.2 Wilcoxon 两样本秩和检验 —— 小样本方法

Wilcoxon 两样本秩和检验就是考虑式 (6.3.3) 的假设检验问题, 即设 X_1, \dots, X_m i.i.d. $\sim F(x)$, Y_1, \dots, Y_n i.i.d. $\sim F(x - \theta)$, 且合样本独立. 要检验式 (6.3.3), 即

$$H_0 : \theta = 0 \leftrightarrow H_1 : \theta \neq 0.$$

设样本 Y_1, \dots, Y_n 的秩和 W 由式 (6.3.6) 给出. 现在这样推理: 每个 R_i 都可取 $1, 2, \dots, N$ 之一为值. 若原假设 H_0 成立, 则全部样本来自同一总体, 每个都不占特殊位置, 不会取较小或较大的值, W 所取之值应集中在平均数 $n \cdot (N + 1)/2$ 附近. 故得到下列检验:

$$\text{当 } W \leq d \text{ 或 } W \geq c \text{ 时, 否定 } H_0. \quad (6.3.7)$$

如何确定 c 和 d ? 它们的确在原则上可以解决: 当 H_0 成立时, 合样本独立同分布, 由此根据对称性的考虑, 易知 (R_1, \dots, R_n) 的联合分布为

$$P(R_1 = r_1, \dots, R_n = r_n) = \begin{cases} \frac{1}{N(N-1)\cdots(N-n+1)}, & r_1, \dots, r_n \leq N \text{ 为互不相同的自然数,} \\ 0, & \text{其他.} \end{cases}$$

由此不难形式地写出 W 的分布. 从而由

$$\alpha = P(W \leq d \text{ 或 } W \geq c | H_0)$$

定出 c 和 d . 对较小的 m, n 已制成表, 见附表 12.

如果假设检验是单边的, 即要检验

$$H'_0 : \theta \leq 0 \leftrightarrow H'_1 : \theta > 0, \quad (6.3.8)$$

由于 $W = \sum_{i=1}^n R_i$ 是 Y_1, \dots, Y_n 在合样本中的秩和. 若 $\theta > 0$ 则因每个 Y_i 的分布与 $X_i + \theta$ 的分布相同, Y_i 与 X_i 相比较倾向于取更大的值, 即 Y_i 取值大于 X_i 的“机会”更多, 而小于它的机会则少. 这样一来, R_1, \dots, R_n 当 $\theta > 0$ 时倾向于取集合

$\{1, 2, \dots, N\}$ 中较大的值 (此处 $N = m + n$), 同样 $\theta < 0$, 则 R_1, \dots, R_n 倾向于取集合 $\{1, 2, \dots, N\}$ 中较小的值. 因为 W 在 $\theta > 0$ 时倾向于取较大的值, 在 $\theta < 0$ 时倾向于取较小的值, 故检验问题 (6.3.8) 的检验为

当 $W \geq c$ 时, 否定 H'_0 ;

同样, 检验问题

$$H''_0 : \theta \geq 0 \leftrightarrow H''_1 : \theta < 0 \quad (6.3.9)$$

的检验为

当 $W \leq d$ 时, 否定 H''_0 .

将上述三类检验列成表 6.3.1.

表 6.3.1 Wilcoxon 两样本秩和检验(小样本情形)

H_0	H_1	否定域
$\theta = 0$	$\theta \neq 0$	$W \leq d$ 或 $W \geq c$
$\theta \leq 0$	$\theta > 0$	$W \geq c$
$\theta \geq 0$	$\theta < 0$	$W \leq d$

如何确定临界值 c 和 d ? 对较小的 m, n ($n \leq m$) 已经制成表, 见附表 12. 关于此表, 作以下两点说明:

(1) 分别记 X_1, \dots, X_m 和 Y_1, \dots, Y_n 在合样本中的秩和为 W_1 和 W_2 , 则

$$W_1 + W_2 = 1 + 2 + \dots + (m+n) = \frac{(m+n)(m+n+1)}{2}$$

是一个常数. 因此, 在使用 Wilcoxon 秩和检验法比较两总体分布时, 用 W_1 作为检验统计量与用 W_2 作为检验统计量是一回事. 故不失一般性, 假设 $n \leq m$.

(2) 在附表 12 中只给出了秩和检验的临界值 $P(W \geq c) \leq \alpha$ 中 c 的值, 对 $P(W \leq d) \leq \alpha$ 的临界值 d 如何利用此表示出? 可以证明 $P(W \leq d) = P(W \geq n(m+n+1)-d)$, 即若记 $c = n(m+n+1)-d$, 先对给定的 α, m, n 求出 $P(W \geq c) \leq \alpha$ 的临界值 c , 然后 d 由公式 $d = n(m+n+1) - c$ 算出.

例 6.3.1 某种羊毛在进行某种工艺处理之前和处理之后, 各随机抽取一个容量为 5 的样本, 测得其含脂率如下:

处理前: 0.20, 0.24, 0.66, 0.42, 0.12;

处理后: 0.13, 0.07, 0.21, 0.08, 0.19.

问处理后含脂率是否下降 ($\alpha = 0.05$).

解 设 X 和 Y 分别表示处理前、后羊毛的含脂率, 它们的分布函数分别为 $F(x)$ 和 $F(x - \theta)$, 则检验问题

$$H_0 : \theta \geq 0 \leftrightarrow H_1 : \theta < 0,$$

即将“处理后羊毛含脂率没有下降”作为原假设.

由表 6.3.1 可知: 当 $W \leq d$ 时否定原假设. 由前面关于附表 12 的使用说明(2), 先从附表中由 $P(W \geq c) \leq \alpha$ 查出 c , 则 $d = n(m+n+1) - c$.

本题中 $m = n = 5$, $\alpha = 0.05$, 由附表 12, 查出 $c = 36$, 故有

$$d = n(m+n+1) - c = 5 \times 11 - 36 = 19.$$

这表明

$$P(W \leq 19) = P(W \geq 36) \leq 0.05.$$

因此上述检验问题的否定域

$$D = \{(X, Y) : W \leq 19\}.$$

现将两组样本观察值按从小到大排成一列成表 6.3.2.

表 6.3.2

0.07	0.08	0.12	0.13	0.19	0.20	0.21	0.24	0.42	0.66
1	2	3	4	5	6	7	8	9	10

下划线的数是处理后羊毛含脂率 (Y) 的观察值的秩. 故 Y 的观察值的秩和为

$$W = 1 + 2 + 4 + 5 + 7 = 19.$$

将其与否定域中临界值比较 $W = 19 \leq d$ ($d = 19$), 因此否定 H_0 , 即认为处理后羊毛含脂率下降了.

6.3.3 Wilcoxon 两样本秩和检验 —— 大样本方法

前面讨论了当 m, n 较小时 Wilcoxon 两样本秩和检验的小样本方法. 当 m, n 较大时检验统计量 W 的分布的计算很复杂, 对给定的 α 没有现成的表可以查到否定域的临界值. 因此得依靠极限定理, 方法如下: 容易求得

$$E(W) = \frac{n(m+n+1)}{2} = \frac{n(N+1)}{2},$$

$$D(W) = \frac{mn(n+m+1)}{12} = \frac{mn(N+1)}{12}.$$

此处 W 由式 (6.3.6) 给出, $N = m+n$. 记

$$W^* = \frac{W - E(W)}{\sqrt{D(W)}} = \frac{W - n(N+1)/2}{\sqrt{mn(N+1)/12}}.$$

可以证明在原假设 $H_0 : \theta = 0$ 成立之下, 当 $m, n \rightarrow \infty$ 时有

$$W^* = \frac{W - n(n+1)/2}{\sqrt{mn(n+1)/12}} \xrightarrow{\mathcal{L}} N(0, 1).$$

因此可得检验问题 (6.3.3) 的水平近似为 α 的否定域

$$D = \left\{ (X, Y) : |W^*| > u_{\alpha/2} \right\}.$$

类似可得检验问题 (6.3.8) 和问题 (6.3.9) 的水平近似为 α 的否定域, 详见表 6.3.3.

表 6.3.3 Wilcoxon 两样本秩和检验(大样本情形)

H_0	H_1	否定域
$\theta = 0$	$\theta \neq 0$	$ W^* \geq u_{\alpha/2}$
$\theta \leq 0$	$\theta > 0$	$W^* \geq u_\alpha$
$\theta \geq 0$	$\theta < 0$	$W^* \leq -u_\alpha$

例 6.3.2 设甲、乙两台机床精度相同, 用它们加工同样的产品. 从这两台机床加工的产品中分别随机地抽取 100 件和 80 件, 测得这些产品的直径(单位: mm), 将合样本按大小排序, 获得机床乙加工的 80 件产品的秩和为 $W = 6540$, 试问甲、乙两台机床加工产品的直径有无显著差异 ($\alpha = 0.01$).

解 设 X_1, \dots, X_m 和 Y_1, \dots, Y_n 分别表示甲、乙两台车床加工产品的直径, 假定这两组样本的分布函数分别为 $F(x)$ 和 $F(x - \theta)$, 则检验问题为

$$H_0 : \theta = 0 \leftrightarrow H_1 : \theta \neq 0.$$

此例中 $m = 100$, $n = 80$, $W = 6540$. 由于 m, n 较大, 故可采用大样本方法. 由表 6.3.3 可知水平近似 $\alpha = 0.05$ 的大样本否定域为

$$D = \left\{ (X, Y) : |W^*| \geq u_{0.005} = 2.58 \right\},$$

其中

$$|W^*| = \left| \frac{W - n(n+m+1)/2}{\sqrt{mn(m+n+1)/12}} \right| = \left| \frac{6540 - 7240}{347, 37} \right| = 2.02 < 2.58,$$

不足以否定 H_0 , 故可接受 H_0 , 即认为这两台机床加工的直径无显著差异.

有人可能会觉得, 这种秩和检验的效率不会高, 因为它只利用了样本中的大小关系而完全忽略了其具体数值, 其实不然. 近代关于秩检验的大样本理论证明了, 一般秩检验至少在样本容量较大时, 与传统的参数检验相比并不逊色. 拿 Wilcoxon 两样本秩和检验与两样本 t 检验相比较, 即使在随机误差分布 F 为正态时, Wilcoxon 检验的效率相对于 t 检验也达到 $3/\pi \approx 0.95$ (当样本容量较大时). 对别的分布, 这个相对效率可任意大, 且总不会低于 0.864.

以上假定了合样本 $X_1, \dots, X_m, Y_1, \dots, Y_n$ 彼此不同, 因而 Y_i 的秩 R_i 可以唯一确定. 若假定 F 处处连续, 这一事实将以概率为 1 成立, 这时不存在问题. 当 F 不连续时, 合样本中可能出现相同的, 即所谓“结”的问题. 例如,

$$x_2 < x_1 = x_4 < x_5 < x_3,$$

习惯上把取相同值的几个变量称为一个“结”. 结外的 x_i 的秩唯一确定. 如此处 x_2, x_5 和 x_3 的秩分别为 1, 4 和 5. 结内的 x_i 的秩就不清楚了. 此时取这些相继秩数的平均值作为结内各个 x_i 的秩, 如此处 x_1 和 x_4 占有秩 2 和 3, 取 2.5 分别作为 x_1 和 x_4 的秩. 对所有的“结”作处理之后, 按前面所述的方法讨论两样本秩和检验问题.

6.4 拟合优度检验

6.4.1 引言

拟合优度检验问题的提法如下: 设有一个一维或多维随机变量 X , 令 X_1, \dots, X_n 为从总体 X 中抽取的简单样本, F 是一已知的分布函数. 要利用样本 X_1, \dots, X_n 去检验假设

$$H_0: \text{r.v. } X \text{ 的分布为 } F, \quad (6.4.1)$$

其中 F 常称为理论分布.

导出这种假设检验的想法大致如下: 设法提出一个反映实际数据 X_1, \dots, X_n 与理论分布 F 偏差的量 $D = D(X_1, \dots, X_n; F)$. 如果 D 较大, 如 $D \geq c$, 则认为理论分布 F 与数据 X_1, \dots, X_n 不符, 因而否定 H_0 . 然而这种“非此即彼”的提法常显得有些牵强. 因为一般来说, 理论和实际没有截然的符合或不符合. 更恰当的提法是实际数据与理论分布符合的程度如何? 因此通常对检验问题 (6.4.1) 不是以“是”或“否”来回答, 而是提供一个介于 0 和 1 之间的数字作为回答, 即用此数作为符合程度的度量刻画. 就具体样本算出 D 之值, 记为 d_0 . 称下列的条件概率:

$$p(d_0) = P(D \geq d_0 | H_0)$$

因为在选定的偏离指标 D 之下, 样本与理论分布的拟合优度 (goodness of fit). $p(d_0)$ 越接近 1, 表示样本与理论分布拟合的越好, 因而原假设 (6.4.1) 越可信. 反之, 它越接近 0, 则原假设 H_0 越不可信. 如果它低到指定的水平 α 之下, 则就要否定 H_0 了. 这种思想其实很简单, 比如说, 某个学生考试及格了, 这给出了一定的信息. 但 100 分也及格, 60 分也及格, 如果给出他的具体分数是 90 分, 就知道他不但及格了, 而且学得很好. 因此, 这个信息就比“及格”更充分了.

因此, 在给定检验水平 $0 < \alpha < 1$ 后, 根据拟合优度可以给出检验问题 (6.4.1) 的一个检验如下:

$$\text{当 } p(d_0) < \alpha \text{ 时否定 } H_0, \text{ 当 } p(d_0) \geq \alpha \text{ 时接受 } H_0. \quad (6.4.2)$$

拟合优度 $p(d_0)$ 越大, 则接受 H_0 的结论越可靠. 这种类型的检验称为拟合优度检验 (goodness of fit test).

由于 D 可以用种种不同的方式定义, 可以有种种不同的拟合优度检验, 其中最著名的是 K. Pearson 在 1900 年提出的 χ^2 检验 (称为 Pearson χ^2 检验) 和柯尔莫哥洛夫在 1933 年提出的一种检验. 本节讨论前者, 后者将在 6.6 节介绍.

6.4.2 Pearson χ^2 检验: 理论分布完全已知的情况

1. 随机变量 X 为离散型, 且只取有限个不同值 a_1, \dots, a_r 的情形

设 X_1, \dots, X_n 为从总体 X 中抽取的简单样本, 理论分布为

$$F: \begin{pmatrix} a_1, & a_2, & \cdots, & a_r \\ p_1, & p_2, & \cdots, & p_r \end{pmatrix},$$

其中 p_1, p_2, \dots, p_r 已知, 且 $\sum_{i=1}^r p_i = 1$. 检验问题 (6.4.1) 写成

$$H_0: P(X = a_i) = p_i, \quad i = 1, 2, \dots, r. \quad (6.4.3)$$

设样本 X_1, \dots, X_n 中等于 a_i 的个数记为 ν_i , $i = 1, 2, \dots, r$, 则 ν_i 称为 a_i 的观察频数, 显然有 $\sum_{i=1}^r \nu_i = n$; 而 np_i 称为 a_i 的理论频数 (因为 ν_i/n 为 X_1, \dots, X_n 中取值为 a_i 的频率, 频率的极限为 p_i , 故当 n 充分大时有 $\nu_i/n \approx p_i$, 因此极限情形的频数为 np_i , 称之为理论频数). 基于上述解释, 可见 $\sum_{i=1}^r c_i (\nu_i/n - p_i)^2$ 可作为样本与理论分布偏差的一种度量, 取平方是防止正负抵消. c_i 取何值好呢? K. Pearson 在 1900 年证明了: 若取 $c_i = n/p_i$, 则在 H_0 成立的前提下,

$$K_n = K_n(X_1, \dots, X_n; F) = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i} \quad (6.4.4)$$

的极限分布 (当 $n \rightarrow \infty$ 时) 为 χ^2_{r-1} . 因此有如下定理.

定理 6.4.1 (K. Pearson) 设 K_n 由式 (6.4.4) 给出, 则在 H_0 成立的条件下, 当样本容量 $n \rightarrow \infty$ 时有

$$K_n \xrightarrow{\mathcal{L}} \chi^2_{r-1},$$

即 K_n 的分布收敛于自由度为 $r-1$ 的 χ^2 分布.

这一定理的证明放在本节最后.

按照这一定理可以提出如下的检验方法: 当 n 充分大时, 可以近似地认为: 检验统计量 K_n 的分布就是 χ_{r-1}^2 . 于是得到检验问题 (6.4.3) 的水平近似为 α 的检验

$$\text{当 } K_n > \chi_{r-1}^2(\alpha) \text{ 时否定 } H_0, \text{ 否则就接受 } H_0. \quad (6.4.5)$$

正如在引言中所述, 对检验问题 (6.4.3) 只给出一个“是”或“否”的结论, 有时太牵强. 常求出一个拟合优度, 具体方法如下:

记 k_0 为对一组具体样本按式 (6.4.4) 算出的 K_n 之值. 计算概率

$$p(k_0) = P(K_n \geq k_0 | H_0) \approx P(\chi_{r-1}^2 \geq k_0), \quad (6.4.6)$$

其中 χ_{r-1}^2 表示自由度为 $r-1$ 的 χ^2 变量. 上式中的近似等号, 是由于当 n 充分大时 $K_n \sim \chi_{r-1}^2$ 近似成立之故. $p(k_0)$ 称为拟合优度, 它是度量样本与理论分布偏离程度的量. 若 $p(k_0)$ 较大, 如为 0.80, 它表明 H_0 成立时, 检验统计量 K_n 落在 $[k_0, +\infty)$ 中的可能性有 80%. 这表明在一次抽样中出现像 k_0 这么大或比 k_0 更大的偏差不足为奇, 因此可以认为样本数据与理论分布拟合较好. 反之, 若 $p(k_0)$ 较小, 如为 0.01, 它表明在 H_0 成立的前提下, 产生 k_0 这么大或比 k_0 更大的偏差的概率只有 1%, 这是一个小概率事件, 在一次抽样中发生, 这就产生了异常 (一般认为小概率事件, 在一次抽样中不应该发生). 因此, 这表明样本与理论分布不一致, 拟合的不好.

在获得拟合优度后, 若检验水平 α 给定了, 可按式 (6.4.2) 来对检验问题作出结论, 这与式 (6.4.5) 给出的检验是一致的. 但拟合优度除了能作出检验来, 还能给出更多的信息. 例如, 若 $p(k_0)$ 较大, 则就知道接受 H_0 的把握较大.

例 6.4.1 孟德尔 (Mendel) 豌豆杂交试验. 在试验中孟德尔按颜色和形状把豌豆分成四类: 黄而圆的, 青而圆的, 黄而有角的, 青而有角的 (分别记这些类为 $A_i, i = 1, 2, 3, 4$). 按孟得尔理论, 这四类豌豆个数之比为 9:3:3:1. 他对杂交试验的豌豆进行了 556 次观察, 观察到这四类豌豆频数分别为 315, 108, 101, 32, 要检验以上 9:3:3:1 规律是否成立.

解 任取一粒豌豆, 由 9:3:3:1 可知它属于这四类的概率为 $9/16, 3/16, 3/16$ 和 $1/16$. 若记 $X = i$ 表示任一粒豌豆属于类 $A_i, i = 1, 2, 3, 4$, 则检验问题可表示为

$$H_0 : P(X = 1) = \frac{9}{16}, \quad P(X = 2) = \frac{3}{16}, \\ P(X = 3) = \frac{3}{16}, \quad P(X = 4) = \frac{1}{16}.$$

令 $\nu_1 = 315, \nu_2 = 108, \nu_3 = 101, \nu_4 = 32$, 则

$$k_0 = \frac{(315 - 556 \times 9/16)^2}{556 \times 9/16} + \frac{(108 - 556 \times 3/16)^2}{556 \times 3/16} \\ + \frac{(101 - 556 \times 3/16)^2}{556 \times 3/16} + \frac{(32 - 556 \times 1/16)^2}{556 \times 1/16} = 0.47,$$

拟合优度为

$$p(k_0) = P(K_n \geq k_0 | H_0) \approx P(\chi^2_3 \geq 0.47) > 0.90,$$

可见数据与理论分布拟合很好. 所以有较大的把握认为 9:3:3:1 的规律可信.

2. 理论分布为任一确定分布的情形

这一情形包括理论分布为离散型随机变量但取可列个值的情形, 或理论分布为连续分布的情形.

设 X_1, \dots, X_n 是从总体 X 中抽取的简单样本, 要检验

$$H_0: \text{r.v. } X \text{ 的分布为 } F, \quad (6.4.7)$$

其中 F 为一已知分布. 具体做法如下:

设 X 为一维. 第一步, 取 $r-1$ 个常数 a_1, \dots, a_{r-1} , 满足 $a_0 = -\infty < a_1 < a_2 < \dots < a_{r-1} < +\infty = a_r$, 将实数轴 $(-\infty, +\infty)$ 分成 r 个子区间

$$I_1 = (-\infty, a_1), \quad I_2 = [a_1, a_2), \quad \dots, \\ I_j = [a_{j-1}, a_j), \quad \dots, \quad I_r = [a_{r-1}, +\infty). \quad (6.4.8)$$

若 r.v. X 是 m 维的, $m > 1$, 则将 R_m 分解为 r 个彼此无公共点的区域 I_1, \dots, I_r .

第二步, 计算 r 个事件在 H_0 成立下的概率

$$p_j = P_F(X \in I_j) = F(a_j) - F(a_{j-1}), \quad j = 1, 2, \dots, r, \quad (6.4.9)$$

此处易见 $F(a_0) = 0, F(a_r) = 1$, 则检验问题 (6.4.7) 转化为

$$H_0: P(X \in I_j) = p_j, \quad j = 1, 2, \dots, r. \quad (6.4.10)$$

第三步, 令 ν_j 为 X_1, \dots, X_n 落入式 (6.4.8) 中的区间 I_j 的观察频数, $j = 1, 2, \dots, r$. 计算检验统计量

$$K_n = \sum_{j=1}^r \frac{(\nu_j - np_j)^2}{np_j} \quad (6.4.11)$$

的值. 在 H_0 成立时, 当 $n \rightarrow \infty$, 定理 6.4.1 仍成立.

最后, 若记 k_0 为由 K_n 算出的具体值, 按式 (6.4.6) 算出拟合优度 $p(k_0)$. 根据 $p(k_0)$ 之值和检验水平 α 对检验问题 (6.4.7) 作出结论.

但要注意两点: ① 在第一步中分组的大小, 即 r 选多大才好, 取决于 n 的大小, 一种经验法则认为 a_1, \dots, a_{r-1} 的选择应使理论频数 np_i 和观察频数 ν_i ($i = 1, 2, \dots, r$) 都不小于 5 为宜. 否则, 将相邻子区间合并, 直到满足上述要求为止. ② 另一点要注意的是式 (6.4.8) 中的 a_1, \dots, a_r 必须不依赖于样本. 就是说不能根据样本 X_1, \dots, X_n 的位置去选择它们, 而必须事先定好, 只有这样定理 6.4.1 的结论才有效.

6.4.3 Pearson χ^2 检验: 理论分布带有未知参数的情形

此时要检验的假设是: r.v. X 的分布属于一个确定的分布族 $\{F(x; \theta_1, \dots, \theta_s) : (\theta_1, \dots, \theta_s) \in \Theta\}$, 其中 Θ 为参数空间. 令 X_1, \dots, X_n 为自总体 X 中抽取的简单样本, 要检验假设

$$H_0: \text{存在 } (\theta_1^0, \dots, \theta_s^0) \in \Theta, \text{ 使 } X \text{ 的分布为 } F(x; \theta_1^0, \dots, \theta_s^0). \quad (6.4.12)$$

例如, 有一批数据 X_1, \dots, X_n , 要检验它是否来自 $N(a, \sigma^2)$, 其中 a, σ^2 未知, 就属于这一情形.

求式 (6.4.12) 的检验方法是前一段所讨论的理论分布完全已知情形的直接推广. 取 I_1, \dots, I_r 如式 (6.4.8) 所示, 改式 (6.4.9) 为下列形式:

$$\begin{aligned} p_j(\theta_1, \dots, \theta_s) &= P(X \in I_j) = F(a_j; \theta_1, \dots, \theta_s) - F(a_{j-1}; \theta_1, \dots, \theta_s), \\ j &= 1, 2, \dots, r, \quad a_0 = -\infty, \quad a_r = +\infty. \end{aligned} \quad (6.4.13)$$

因此检验问题 (6.4.12) 转化为

$$\begin{aligned} H_0: \text{存在 } (\theta_1^0, \dots, \theta_s^0) \in \Theta, \text{ 使得 } P(X \in I_j) &= p_j(\theta_1^0, \dots, \theta_s^0), \\ j &= 1, 2, \dots, r, \end{aligned} \quad (6.4.14)$$

其中对 $p_j(\theta_1, \dots, \theta_s)$ 作如下假定:

- (1) 对任何 $(\theta_1, \dots, \theta_s) \in \Theta$, $p_j = p_j(\theta_1, \dots, \theta_s) > 0$, $j = 1, \dots, r$ 且 $\sum_{j=1}^r p_j = 1$;
- (2) 对任何 $1 \leq j \leq r$, $p_j(\theta_1, \dots, \theta_s)$ 对 $\theta_1, \dots, \theta_s$ 有一阶连续偏导数.

算出样本 X_1, \dots, X_n 落入 I_j 的观察频数 ν_j , 求出类似于式 (6.4.11) 的表达式

$$K_n(\theta_1, \dots, \theta_s) = \sum_{j=1}^r \frac{(\nu_j - np_j(\theta_1, \dots, \theta_s))^2}{np_j(\theta_1, \dots, \theta_s)}. \quad (6.4.15)$$

由于 $\theta_1, \dots, \theta_s$ 未知, 故式 (6.4.15) 中 $K_n(\theta_1, \dots, \theta_s)$ 还不能作为检验统计量. 因此要按某种估计方法, 将 $\theta_1, \dots, \theta_s$ 用样本 X_1, \dots, X_n 估出. 若令 $\hat{\theta}_j = \hat{\theta}_j(X_1, \dots, X_n)$ 为 θ_j 的估计, 以之取代式 (6.4.15) 中的 θ_j , $j = 1, 2, \dots, s$, 则检验问题 (6.4.12) 可表为 H_0 : r.v. X 的分布为 $F(x; \hat{\theta}_1, \dots, \hat{\theta}_s)$, 得检验统计量

$$K_n^* = K_n^*(\hat{\theta}_1, \dots, \hat{\theta}_s) = \sum_{j=1}^r \frac{(\nu_j - np_j(\hat{\theta}_1, \dots, \hat{\theta}_s))^2}{np_j(\hat{\theta}_1, \dots, \hat{\theta}_s)}. \quad (6.4.16)$$

然后, 通过 K_n^* 的极限分布作出拟合优度检验.

Pearson 在 1900 年的工作中认为, 在理论分布带未知参数的情形下定理 6.4.1 的结论仍成立, 即若假设 (6.4.12) 成立, 则当 $n \rightarrow \infty$ 时, K_n^* 的分布收敛于 χ_{r-1}^2 . Fisher 在 20 世纪 20 年代发现 Pearson 推理中的疏忽, 指出 K_n^* 极限分布的自由度不是 $r-1$, 而应为 $r-1-s$, 而且对 θ_j ($j = 1, \dots, s$) 的估计量的大样本性质还有所要求. 下面叙述 Fisher 的方法及主要结果, 而略去其证明.

首先, 通过简单样本 X_1, \dots, X_n , 用极大似然方法对 $\theta_1, \dots, \theta_s$ 作一估计. 如前, 设 ν_j 为样本 X_1, \dots, X_n 中落入 I_j 的个数, $j = 1, 2, \dots, r$, 则似然函数为

$$L = \frac{n!}{\nu_1! \dots \nu_r!} (p_1(\theta_1, \dots, \theta_s))^{\nu_1} \cdots (p_r(\theta_1, \dots, \theta_s))^{\nu_r},$$

对 $\log L$ 关于 θ_i 求导, 得方程组

$$\sum_{j=1}^r \frac{\nu_j}{p_j(\theta_1, \dots, \theta_s)} \cdot \frac{\partial p_j(\theta_1, \dots, \theta_s)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, s. \quad (6.4.17)$$

解这一方程组得

$$\hat{\theta}_i = \hat{\theta}_i(X_1, \dots, X_n), \quad i = 1, 2, \dots, s. \quad (6.4.18)$$

按式 (6.4.16) 算出检验统计量 K_n^* . R. A. Fisher 证明了如下定理.

定理 6.4.2 (R.A. Fisher) 设 $(\theta_1^0, \dots, \theta_s^0)$ 为参数空间 Θ 的内点, 使得 X 的分布为 $F(x; \theta_1^0, \dots, \theta_s^0)$. $(\hat{\theta}_1, \dots, \hat{\theta}_s)$ 为式 (6.4.17) 的相合解, 检验统计量 K_n^* 由式 (6.4.16) 给出, 则在原假设式 (6.4.12) 成立下, 当 $n \rightarrow \infty$ 时有

$$K_n^* \xrightarrow{\mathcal{L}} \chi_{r-1-s}^2.$$

定理 6.4.2 的证明超出本课程范围, 故从略. 定理的证明可见文献 [1] 第 302 页或文献 [3] 第 409 页.

利用上述定理可得检验问题 (6.4.12) 的水平近似为 α 的大样本检验如下:

$$\text{若 } K_n^* > \chi_{r-1-s}^2(\alpha) \text{ 则否定 } H_0, \text{ 否则接受 } H_0. \quad (6.4.19)$$

类似地, 可以计算拟合优度. 设有了样本后, 记 k_0^* 为按式 (6.4.16) 算出的 K_n^* 的具体值, 计算

$$p(k_0^*) = P(K_n^* > k_0^* | H_0) \approx P(\chi_{r-1-s}^2 > k_0^*), \quad (6.4.20)$$

其中 χ^2_{r-1-s} 表示自由度为 $r-1-s$ 的 χ^2 变量. 对给定的检验水平 α , 当 $p(k_0^*) < \alpha$ 时否定 H_0 , 当 $p(k_0^*) \geq \alpha$ 时接受 H_0 , 拟合优度检验能比检验式 (6.4.19) 给出了更多的信息, 它可以告诉我们接受 (或拒绝) H_0 的把握有多大.

注 6.4.1 解方程组 (6.4.17) 并不容易. 一个常用的做法是: 对分布 $F(x; \theta_1, \dots, \theta_s)$, 直接由样本 X_1, \dots, X_n (而不是由 ν_1, \dots, ν_r) 求出参数 $\theta_1, \dots, \theta_s$ 的 MLE. 例如, F 为正态分布 $N(a, \sigma^2)$, 参数 a 和 σ^2 的 MLE 为 \bar{X} 和 $S_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$, 用 $\hat{\theta}_1^* = \bar{X}$, $\hat{\theta}_2^* = S_n^2$ 代替式 (6.4.18) 中的 $\hat{\theta}_i$, 可以大大地简化计算. 在下面的例 6.4.2 和例 6.4.3 中将会看到这一点. 然而, 在理论上可以证明: 若用 $\theta_1, \dots, \theta_s$ 的基于样本 X_1, \dots, X_n 的 MLE, $\hat{\theta}_1^*, \dots, \hat{\theta}_s^*$, 代替式 (6.4.15) 中的 $\theta_1, \dots, \theta_s$, 而不用式 (6.4.17) 之解 $\hat{\theta}_1, \dots, \hat{\theta}_s$. 记 $\hat{p}_j^* = \hat{p}_j^*(\hat{\theta}_1^*, \dots, \hat{\theta}_s^*)$, 则所算出的检验统计量为

$$\tilde{K}_n^* = \sum_{j=1}^r \frac{(\nu_j - np_j^*)^2}{np_j^*}.$$

此检验统计量 \tilde{K}_n^* 不一定有极限分布 χ^2_{r-1-s} . 更确切地说, 可以证明: 当 n 充分大时, 真正的拟合优度值 $p(\tilde{k}_0^*) = P(\tilde{K}_n^* > \tilde{k}_0^* | H_0)$ 介于 $P(\chi^2_{r-1-s} \geq \tilde{k}_0^*)$ 和 $P(\chi^2_{r-1} \geq \tilde{k}_0^*)$ 之间 (后者较大). 按定理 6.4.2 算出的值为前者, 而当无待估参数时, 按定理 6.4.1 算出的值为后者. 因此这个结果可以形象地解释为: 用极大似然估计 $(\hat{\theta}_1^*, \dots, \hat{\theta}_s^*)$ 而不用式 (6.4.17) 之解 $(\hat{\theta}_1, \dots, \hat{\theta}_s)$, 相当于把失掉的自由度 s 挽回一部分. 问题在于 \tilde{K}_n^* 的极限分布不再是 χ^2_{r-s-1} 分布了.

在下面的例子中, 用极大似然估计 $(\hat{\theta}_1^*, \dots, \hat{\theta}_s^*)$ 代替式 (6.4.17) 的解 $(\hat{\theta}_1, \dots, \hat{\theta}_s)$, 仍用定理 6.4.2 中的结果作为其极限分布, 因为这是一个近似结果, 虽有差别但差别不大.

例 6.4.2 在某细纱机上进行断头测定, 试验锭子总数为 443, 测得断头总次数为 308 次, 各锭子的断头次数记录见表 6.4.1.

表 6.4.1

每锭断头数	0	1	2	3	4	5	6	7	8
锭数 (实测)	263	112	38	19	5	1	1	0	4

问锭子的断头数是否服从 Poisson 分布 ($\alpha = 0.05$).

解 令锭子的断头数为 r.v. X , 则要检验的假设为

$$H_0: \text{存在 } \lambda_0 \text{ 使得 } X \sim \text{Poisson 分布 } P(\lambda_0). \quad (6.4.21)$$

(1) 将数据重新分组, 使得每组断头的锭子数不少于 5, 见表 6.4.2.

表 6.4.2

每锭断头数	0	1	2	3—8
锭数	263	112	38	30

(2) 求未知参数的 MLE. Poisson 分布中参数 λ 的 MLE 为

$$\hat{\lambda}^* = \bar{X} = \frac{308}{443} = 0.7,$$

检验问题可视为在检验问题 (6.4.21) 中取 $\lambda_0 = \hat{\lambda}^* = 0.7$, 即

$$H_0: \text{r.v. } X \sim \text{Poisson 分布 } P(0.7).$$

这就转化为与理论分布完全已知情形.

(3) 求 \tilde{K}_n^* . 此处 $n = 443$,

$$n\hat{p}_0^* = nP_{\hat{\lambda}^*}(X=0) = 443 \times \frac{(0.7)^0 e^{-0.7}}{0!} = 443 \times 0.496585 = 219.99,$$

$$n\hat{p}_1^* = 443 \times \frac{(0.7) \times e^{-0.7}}{1!} = 153.99,$$

$$n\hat{p}_2^* = 443 \times \frac{(0.7)^2 \times e^{-0.7}}{2!} = 53.90,$$

$$n\hat{p}_3^* = 443 \times P(3 \leq X \leq 8) = 443 \cdot \sum_{i=3}^8 \frac{(0.7)^i e^{-0.7}}{i!} = 15.13.$$

从而有

$$\begin{aligned} \tilde{K}_n^* &= \frac{(263 - 219.99)^2}{219.99} + \frac{(112 - 153.99)^2}{153.99} + \frac{(38 - 53.90)^2}{53.90} + \frac{(30 - 15.13)^2}{15.13} \\ &= 8.41 + 11.45 + 4.69 + 14.61 = 39.16. \end{aligned}$$

(4) 计算拟合优度. 记 $\tilde{k}_0^* = 39.16$, $r = 4$, $s = 1$ 故 $r - s - 1 = 2$, 则拟合优度为

$$p(\tilde{k}_0^*) = P(\tilde{K}_n^* > \tilde{k}_0^* | H_0) \approx P(\chi_2^2 > 39.16) < 0.005 < \alpha,$$

故否定 H_0 , 即认为锭子的断头数不能服从 Poisson 分布.

例 6.4.3 研究混凝土抗压强度的分布, 200 件混凝土制件的抗压强度 (单位: kg/cm^2) 分组见表 6.4.3.

表 6.4.3

压强区间	190—200	200—210	210—220	220—230	230—240	240—250
频数 ν_i	10	26	56	64	30	14

设 X_1, \dots, X_{200} 为从总体 $F(x)$ 中抽取的简单样本, $F(x)$ 为压强分布函数, 问 $F(x)$ 是否为正态分布? (取 $\alpha = 0.05$)

解 设 r.v. X 表示混凝土制件的抗压强度, 记 $\theta = (\mu, \sigma^2)$ 检验问题为

$$H_0: \text{存在 } \mu_0, \sigma_0^2 \text{ 使得 } X \text{ 服从正态分布 } N(\mu_0, \sigma_0^2). \quad (6.4.22)$$

此题不需要重新分组. 取每个区间的中点作为代表值, 则 μ 和 σ^2 的 MLE 为

$$\hat{\mu}^* = \bar{X} = \frac{1}{200}(195 \times 10 + 205 \times 26 + 215 \times 56 + 225 \times 64)$$

$$+ 235 \times 30 + 245 \times 14) = 221 \text{ (kg/cm}^2\text{)},$$

$$\hat{\sigma}_*^2 = \frac{1}{200} \sum_{i=1}^{200} (X_i - \bar{X})^2 = \frac{1}{200} [(195 - 221)^2 \times 10 + (205 - 221)^2 \times 26$$

$$+ \cdots + (245 - 221)^2 \times 14] = 12.33^2,$$

故 $\hat{\sigma}_* = 12.33$. 检验问题转化为在检验问题 (6.4.22) 中取 $\mu_0 = 221, \sigma_0^2 = 12.33^2$, 即

$$H_0: \text{r.v. } X \sim N(221, 12.33^2).$$

为计算检验统计量 \tilde{K}_n^* , 先计算 r.v. X 在每个区间中的概率. 区间的个数 $r = 6$,

$$\hat{p}_i^* = P(a_{i-1} \leq X \leq a_i) = P(u_{i-1} \leq U \leq u_i) = \Phi(u_i) - \Phi(u_{i-1}),$$

此处 $i = 1, \dots, 6$, $U = (X - 221)/12.33$, 显然当 H_0 成立时 $U \sim N(0, 1)$;

$$\Phi(u_i) = \int_{-\infty}^{u_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt,$$

其中 $u_i = (a_i - 221)/12.33$, $a_0 = -\infty$, $a_6 = +\infty$, 故有

$$\begin{aligned} \hat{p}_1^* &= P(-\infty < X < 200) = P\left(-\infty < U < \frac{200 - 221}{12.33}\right) \\ &= P(U < -1.70) = 0.045, \end{aligned}$$

$$\hat{p}_2^* = P(200 < X < 210) = P(-1.70 < U < -0.89) = 0.142,$$

类似可求出 $\hat{p}_3^*, \dots, \hat{p}_6^*$, 可按表 6.4.4 计算 K_n^* 的值.

表 6.4.4 χ^2 检验统计量的计算

i	区间	ν_i	\hat{p}_i^*	$n\hat{p}_i^*$	$(\nu_i - n\hat{p}_i^*)^2$	$\frac{(\nu_i - n\hat{p}_i^*)^2}{n\hat{p}_i^*}$
1	$(-\infty, 200)$	10	0.045	9	1	0.11
2	$[200, 210)$	26	0.142	28.4	5.76	0.20
3	$[210, 220)$	56	0.281	56.2	0.04	0.00
4	$[220, 230)$	64	0.299	59.8	17.64	0.29
5	$[230, 240)$	30	0.171	34.2	17.64	0.52
6	$[240, +\infty)$	14	0.062	12.4	2.56	0.21
\sum	—	200	1	200	—	1.33

即 \tilde{K}_n^* 的具体值 $\tilde{k}_0^* = 1.33$, 拟合优度为

$$p(\tilde{k}_0^*) = P(\tilde{K}_n^* > \tilde{k}_0^* | H_0) \approx P(\chi_3^2 > 1.33) > 0.70 > \alpha.$$

这个概率较大, 在 H_0 成立时出现像 1.33 或比它更大的值很正常, 因此接受 H_0 . 由于拟合优度较高, 这一结论比较可靠.

*6.4.4 定理 6.4.1 的证明

注意在假设 (6.4.3) 成立时, $(\nu_1, \nu_2, \dots, \nu_r)$ 有多项分布

$$P(\nu_1 = n_1, \dots, \nu_r = n_r) = \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r},$$

其中 n_i ($i = 1, \dots, r$) 为非负整数, 且 $\sum_{i=1}^r n_i = n$.

在 $r = 2$ 时, (ν_1, ν_2) 服从二项分布, 且

$$K_n = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(\nu_2 - np_2)^2}{np_2} = \left(\frac{\nu_1 - np_1}{\sqrt{np_1 p_2}} \right)^2.$$

由棣莫弗-拉普拉斯极限定理可知

$$\frac{\nu_1 - np_1}{\sqrt{np_1 p_2}} \xrightarrow{\mathcal{L}} N(0, 1),$$

故有 $K_n \xrightarrow{\mathcal{L}} \chi_1^2$. 因此, 在 $r = 2$ 时定理成立. 由此可以看到 χ^2 统计量的极限定理是二项分布的中心极限定理在多项分布中的推广. 下面利用特征函数 (c.f.) 这一有用工具来证明 χ^2 统计量的极限定理.

令

$$Y_j = \frac{\nu_j - np_j}{\sqrt{np_j}}, \quad j = 1, 2, \dots, r,$$

则有 $K_n = \sum_{j=1}^r Y_j^2$, 且有 $\sum_{j=1}^r \sqrt{p_j} Y_j = 0$. 由于 (ν_1, \dots, ν_r) 的 c.f. 为 (以下 $i^2 = -1$)

$$\begin{aligned} g(t_1, \dots, t_r) &= E(e^{it_1 \nu_1 + \dots + it_r \nu_r}) \\ &= \sum_{n_1 + \dots + n_r = n} \frac{n!}{n_1! \cdots n_r!} (p_1 e^{it_1})^{n_1} \cdots (p_r e^{it_r})^{n_r} \\ &= (p_1 e^{it_1} + \dots + p_r e^{it_r})^n, \end{aligned}$$

所以 (Y_1, \dots, Y_r) 的 c.f. 为

$$\begin{aligned} \varphi(t_1, \dots, t_r) &= E(e^{it_1 Y_1 + \dots + it_r Y_r}) \\ &= \exp \left\{ -i\sqrt{n} \sum_{j=1}^r \sqrt{p_j} t_j \right\} \cdot g \left(\frac{t_1}{\sqrt{np_1}}, \dots, \frac{t_r}{\sqrt{np_r}} \right). \end{aligned}$$

两边取对数得

$$\log \varphi(t_1, \dots, t_r) = -i\sqrt{n} \sum_{j=1}^r \sqrt{p_j} t_j + n \log \left[\sum_{j=1}^r p_j \exp \left(\frac{it_j}{\sqrt{np_j}} \right) \right].$$

利用泰勒展开

$$e^x = 1 + x + \frac{x^2}{2} + o(x^2), \quad \log(1+x) = x - \frac{x^2}{2} + o(x^2),$$

有

$$\log \varphi(t_1, \dots, t_r) = -\frac{1}{2} \sum_{j=1}^r t_j^2 + \frac{1}{2} \left(\sum_{j=1}^r t_j \sqrt{p_j} \right)^2 + o(n^{-\frac{1}{2}}).$$

设 $\mathbf{Y} = (Y_1, \dots, Y_r)', \mathbf{T} = (t_1, \dots, t_r)', \boldsymbol{\alpha} = (\sqrt{p_1}, \dots, \sqrt{p_r})'$, 则 $\boldsymbol{\alpha}' \boldsymbol{\alpha} = 1$. 于是有

$$\begin{aligned} \lim_{n \rightarrow \infty} \varphi(t_1, \dots, t_r) &= \exp \left\{ -\frac{1}{2} \sum_{j=1}^r t_j^2 + \frac{1}{2} \left(\sum_{j=1}^r t_j \sqrt{p_j} \right)^2 \right\} \\ &= \exp \left\{ -\frac{1}{2} Q(t_1, \dots, t_r) \right\}, \end{aligned} \quad (6.4.23)$$

其中二次型

$$Q(t_1, \dots, t_r) = \sum_{j=1}^r t_j^2 - \left(\sum_{j=1}^r t_j \sqrt{p_j} \right)^2 = \mathbf{T}' \mathbf{T} - (\mathbf{T}' \boldsymbol{\alpha})^2, \quad (6.4.24)$$

此处二次型的矩阵为 $\mathbf{A} = \mathbf{I} - \boldsymbol{\alpha} \boldsymbol{\alpha}'$, 其中 \mathbf{I} 为 r 阶单位阵.

作正交变换

$$\mathbf{Z} = (Z_1, \dots, Z_r)' = \mathbf{B} \mathbf{Y}, \quad (6.4.25)$$

其中 \mathbf{B} 为正交阵, 其第一行为 $\boldsymbol{\alpha}'$, 则有

$$Z_1 = \boldsymbol{\alpha}' \mathbf{Y} = \sum_{j=1}^r \sqrt{p_j} Y_j = 0,$$

故由变换的正交性, 有

$$K_n = \sum_{j=1}^r Y_j^2 = \mathbf{Y}' \mathbf{Y} = \mathbf{Z}' \mathbf{Z} = \sum_{i=2}^r Z_i^2. \quad (6.4.26)$$

\mathbf{Z} 的 c.f. 记为 $\psi(\mathbf{u})$, 则由式 (6.4.25) 可知 $\psi(\mathbf{u}) = \varphi(\mathbf{B}' \mathbf{u})$, 其中 $\mathbf{u} = (u_1, \dots, u_r)'$. 故由式 (6.4.23) 有

$$\lim_{n \rightarrow \infty} \psi(\mathbf{u}) = \lim_{n \rightarrow \infty} \varphi(\mathbf{B}' \mathbf{u}) = \exp \left\{ -\frac{1}{2} Q(\mathbf{B}' \mathbf{u}) \right\}. \quad (6.4.27)$$

注意到 $B\alpha = (1, 0, \dots, 0)'$, 故由式 (6.4.24) 可知

$$\begin{aligned} Q(B'u) &= (B'u)'(B'u) - (B'u)'\alpha\alpha'(B'u) \\ &= u'BB'u - u'(B\alpha)(\alpha'B')u = \sum_{j=1}^r u_j^2 - u_1^2 = \sum_{j=2}^r u_j^2. \end{aligned} \quad (6.4.28)$$

由式 (6.4.27) 和式 (6.4.28) 可知

$$\lim_{n \rightarrow \infty} \psi(u) = \exp \left\{ -\frac{1}{2}(u_2^2 + \dots + u_r^2) \right\}.$$

由此及特征函数的连续性定理, 可知 $(Z_2, \dots, Z_r)'$ 的分布弱收敛到相互独立的标准正态分布 $N(0, 1)$. 因此, 由式 (6.4.26) 可知

$$K_n = \sum_{j=2}^r Z_j^2 \xrightarrow{\mathcal{L}} \chi_{r-1}^2.$$

定理 6.4.1 得证.

6.5 列联表中的独立性和齐一性检验

6.5.1 列联表中的独立性检验 —— χ^2 检验的应用

1. 问题的提法

设某总体内的每一个体可根据两个属性 A 和 B 分类. 目的是考察这两个属性是否具有联带关系. 例如, 设总体为某地区特定的一群人. 每个人可按他是否吸烟分类, 也可按他是否患肺癌分类. 目的是要说明患肺癌是否与吸烟有关. 为了研究这一问题, 从整个总体中随机地抽出 n 个人作调查, 结果见表 6.5.1.

表 6.5.1

		B		
			患肺癌	未患肺癌
		A		
A	吸烟		n_{11}	n_{12}
	不吸烟		n_{21}	n_{22}

其中 n_{11} 为这 n 个人中既吸烟又患肺癌的人数等. 这种表称为 2×2 列联表 (contingency table). 要根据这表中的资料来推断患肺癌是否与吸烟有关.

将这一想法推广, 得出下面的一般问题: 设总体中每一个体按 A, B 两种属性分类, 属性 A, B 分别有 r 和 s 个水平, 分别记为 A_1, \dots, A_r 和 B_1, \dots, B_s (例如, 若 A 代表“吸烟”这个属性, 它可分为下列 4 个水平, A_1 : 每日吸 5 支烟以下 (包

括不吸烟者); A_2 : 每日吸烟 5-10 支; A_3 : 每日吸烟 10-20 支; A_4 : 每日吸烟 20 支以上。而若 B 代表“患肺癌”这个属性, 它也可分为以下几个水平, B_1 : 不患肺癌; B_2, B_3, B_4 分别表示肺癌的早、中、晚期)。从总体中抽取容量为 n 的随机样本, 测得第 i 个个体上指标状况为 $(A_{r_i}, B_{s_i}), i = 1, 2, \dots, n$, 要依据这些资料判断 A, B 两个属性是否独立。

形式地引进随机向量 $\mathbf{X} = (X^{(1)}, X^{(2)})$, $X^{(1)}$ 和 $X^{(2)}$ 分别记同一个人体上的 A, B 指标的级 (或称水平), 而第 i 个个体的观察结果记为 $X_i = (X_i^{(1)}, X_i^{(2)})$, 按上文的记号, $X_i = (r_i, s_i)$, $i = 1, 2, \dots, n$ 。将 n 个观察结果列成表 6.5.2.

表 6.5.2

$X^{(1)}$	$X^{(2)}$	1	...	j	...	s	\sum
1		n_{11}	...	n_{1j}	...	n_{1s}	$n_{1\cdot}$
\vdots		\vdots		\vdots		\vdots	\vdots
i		n_{i1}	...	n_{ij}	...	n_{is}	$n_{i\cdot}$
\vdots		\vdots		\vdots		\vdots	\vdots
r		n_{r1}	...	n_{rj}	...	n_{rs}	$n_{r\cdot}$
\sum		$n_{\cdot 1}$...	$n_{\cdot j}$...	$n_{\cdot s}$	n

以 n_{ij} 记 X_1, \dots, X_n 中取 (i, j) 为值的个数。这表称为 $r \times s$ 列联表。上表中

$$n_{i\cdot} = \sum_j n_{ij}, \quad n_{\cdot j} = \sum_i n_{ij}, \quad n = \sum_i n_{i\cdot} = \sum_j n_{\cdot j},$$

要检验

$$H_0: X^{(1)} \text{ 和 } X^{(2)} \text{ 独立.} \quad (6.5.1)$$

2. 检验方法

若记

$$P(X^{(1)} = i, X^{(2)} = j) = p_{ij}, \quad i = 1, \dots, r; \quad j = 1, \dots, s,$$

满足 $\sum_i \sum_j p_{ij} = 1$ 。因此如果 $X^{(1)}$ 和 $X^{(2)}$ 独立, 则对一切 (i, j) 有

$$P(X^{(1)} = i, X^{(2)} = j) = P(X^{(1)} = i) P(X^{(2)} = j) = p_{i\cdot} p_{\cdot j},$$

其中 $p_{i\cdot} = \sum_j p_{ij}$, $p_{\cdot j} = \sum_i p_{ij}$ 且有 $\sum_i p_{i\cdot} = 1$, $\sum_j p_{\cdot j} = 1$ 。

因此检验问题 (6.5.1) 转化为

$$H_0: P(X^{(1)} = i, X^{(2)} = j) = p_{i\cdot} p_{\cdot j}, \quad i = 1, \dots, r; \quad j = 1, \dots, s, \quad (6.5.2)$$

满足限制条件

$$\sum_i p_{i \cdot} = 1, \quad \sum_j p_{\cdot j} = 1. \quad (6.5.3)$$

将 $p_{i \cdot}$ 和 $p_{\cdot j}$ 视为参数, 由式 (6.5.3) 可知独立的未知参数为 $s + r - 2$ 个. 因此检验问题变成 6.4.4 小节中讨论过的情形, 即理论分布含有未知参数的情形.

首先对未知参数求其 MLE. 若 H_0 成立, 则似然函数为

$$L = c \prod_{i=1}^r \prod_{j=1}^s (p_{i \cdot} p_{\cdot j})^{n_{ij}} = c \left(\prod_{i=1}^r p_{i \cdot}^{n_{i \cdot}} \right) \left(\prod_{j=1}^s p_{\cdot j}^{n_{\cdot j}} \right),$$

其中 $c = n! / \left(\prod_i \prod_j n_{ij}! \right)$. 取对数得 $\log L$, 并在约束条件 (6.5.3) 下求极值, 用与例 3.3.8 完全类似方法求得 $p_{i \cdot}$ 和 $p_{\cdot j}$ 的 MLE 如下:

$$\hat{p}_{i \cdot}^* = \frac{n_{i \cdot}}{n}, \quad i = 1, 2, \dots, r; \quad \hat{p}_{\cdot j}^* = \frac{n_{\cdot j}}{n}, \quad j = 1, 2, \dots, s.$$

算出 χ^2 统计量之值

$$\begin{aligned} K_n^* &= \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{i \cdot} \hat{p}_{\cdot j}^*)^2}{n \hat{p}_{i \cdot} \hat{p}_{\cdot j}} = n \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{i \cdot} n_{\cdot j} / n)^2}{n_{i \cdot} n_{\cdot j}} \\ &= n \left(\sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_{i \cdot} n_{\cdot j}} - 1 \right). \end{aligned} \quad (6.5.4)$$

由定理 6.4.2 可知, 当 H_0 成立且 $n \rightarrow \infty$ 时有

$$K_n^* \xrightarrow{\mathcal{L}} \chi_{rs-1-(r+s-2)}^2 = \chi_{(r-1)(s-1)}^2.$$

对检验水平 α , 查表求出 $\chi_{(r-1)(s-1)}^2(\alpha)$. 得水平近似为 α 的检验如下:

$$\text{若 } K_n^* > \chi_{(r-1)(s-1)}^2(\alpha) \text{ 时, 否定 } H_0, \text{ 否则就接受 } H_0. \quad (6.5.5)$$

若记 k_0^* 为由样本算得 K_n^* 的具体值, 则检验的拟合优度为

$$p(k_0^*) = P(K_n^* \geq k_0^* | H_0) \approx P(\chi_{(r-1)(s-1)}^2 \geq k_0^*),$$

对给定的检验水平 α , 当 $p(k_0^*) < \alpha$ 时否定 H_0 , 当 $p(k_0^*) \geq \alpha$ 时接受 H_0 , $p(k_0^*)$ 越大, 则接受 H_0 的结论越可靠.

注 6.5.1 在 $r = s = 2$ 这个特例情形, 即 2×2 列联表情形, 简单的代数计算证明, 式 (6.5.4) 可简化为

$$K_n^* = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_1 n_2 n_{1 \cdot} n_{\cdot 2}}. \quad (6.5.6)$$

也可能指标是连续取值, 而不是分成 n 个离散的级别. 例如, A 指标是一个人每日的运动量, B 指标是其体重. 这时类似于式 (6.4.8), 可以把指标的取值范围分成若干个区间, 然后按离散型的情形处理.

所考虑的指标也可以多于 2, 这时每个个体按 3 个或 3 个以上属性进行分类, 就会有三重 (亦称三维) 或三重以上列联表, 统称为多重列联表. 多重列联表独立性的检验方法与二重列联表是类似的, 只是更复杂一些.

例 6.5.1 为研究色盲是否与性别有关, 调查 1000 人, 结果见表 6.5.3.

表 6.5.3

	男	女	\sum
正常	442	514	956
色盲	38	6	44
\sum	480	520	1000

问色盲是否与性别有关 ($\alpha = 0.01$).

解 检验问题为

$$H_0: \text{色盲和性别独立}.$$

此例中 $n = 1000$, $r = s = 2$, 查表 $\chi^2_{(r-1)(s-1)}(\alpha) = \chi^2_1(0.01) = 6.64$. 由式 (6.5.6) 可知检验统计量

$$K_n^* = \frac{1000(442 \times 6 - 514 \times 38)^2}{956 \times 44 \times 480 \times 520} = 27.14 > \chi^2_1(0.01) = 6.64,$$

检验的拟合优度为

$$p(k_0^*) = P(K_n^* \geq k_0^* | H_0) \approx P(\chi^2_1 \geq 27.14) < 0.005.$$

拒绝原假设. 这表明色盲和性别有非常密切的关系.

6.5.2 列联表中的齐一性检验

1. 问题的提法

设有 r 个生产同一产品的工厂, 产品分为 s 个等级. 第 i 个工厂的 j 等品率为 $p_i(j)$, $j = 1, 2, \dots, s$; $i = 1, 2, \dots, r$. “ r 个工厂产品质量相同”, 理解为等级品率都相同. 于是“ r 个工厂产品质量齐一”这个假设, 可表示为

$$H_0: p_1(j) = p_2(j) = \dots = p_r(j), \quad j = 1, 2, \dots, s. \quad (6.5.7)$$

现从第 i 个工厂的产品中抽出 n_i 个, 记录其中的 j 等品有 n_{ij} 个, $j = 1, 2, \dots, s$; $i = 1, 2, \dots, r$. 要依据观察结果 n_{ij} 检验假设 (6.5.7).

把这个问题一般化: 设有 r 个总体 $X^{(1)}, \dots, X^{(r)}$, 它们可能的取值都相同, 为 a_1, \dots, a_s , 且第 i 个总体 $X^{(i)}$ 取值为 a_j 的概率记为 $p_i(j)$, $j = 1, 2, \dots, s$; $i = 1, 2, \dots, r$, 则要求检验假设: “这 r 个总体具有相同分布”, 可写成如式 (6.5.7) 的形式. 其次, 从第 i 个总体中抽取大小为 $n_{i\cdot}$ 的样本, 其中取值为 a_j 的有 n_{ij} 个, $j = 1, \dots, s$; $i = 1, \dots, r$, 可得到表 6.5.4.

此表与表 6.5.2 完全相似, 因此也称为 $r \times s$ 列联表. 但要注意到二者有一显著的不同, 即在表 6.5.4 中诸 $n_{i\cdot}$ 是事先选定之数, 它没有随机性. 而表 6.5.2 中它与抽样结果有关, 因而 $n_{i\cdot}$ 是随机的.

表 6.5.4

X	a	a_1	\cdots	a_j	\cdots	a_s	\sum
1		n_{11}	\cdots	n_{1j}	\cdots	n_{1s}	$n_{1\cdot}$
\vdots		\vdots		\vdots		\vdots	\vdots
i		n_{i1}	\cdots	n_{ij}	\cdots	n_{is}	$n_{i\cdot}$
\vdots		\vdots		\vdots		\vdots	\vdots
r		n_{r1}	\cdots	n_{rj}	\cdots	n_{rs}	$n_{r\cdot}$
\sum		$n_{\cdot 1}$	\cdots	$n_{\cdot j}$	\cdots	$n_{\cdot s}$	n

2. 检验的制定

(1) 若分布完全已知, 即

$$p_1(j) = \cdots = p_r(j) = p_j^0, \quad j = 1, 2, \dots, s,$$

而 p_1^0, \dots, p_s^0 皆已知且满足 $\sum_{j=1}^s p_j^0 = 1$, 这时

$$K = K_n = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{i\cdot} p_j^0)^2}{n_{i\cdot} p_j^0},$$

则由定理 6.4.1 可知, 当 H_0 成立且 $n_{i\cdot} \rightarrow \infty$ ($i = 1, \dots, r$) 时有

$$K_n \xrightarrow{\mathcal{L}} \chi_{(s-1)r}^2.$$

检验问题 (6.5.7) 的水平近似为 α 的检验为

$$\text{当 } K > \chi_{(s-1)r}^2(\alpha) \text{ 否定 } H_0, \text{ 否则接受 } H_0. \quad (6.5.8)$$

也可计算检验的拟合优度为 $p(k_0) = P(K \geq k_0 | H_0) \approx P(\chi_{(s-1)r}^2 \geq k_0)$, 按拟合优度 $p(k_0)$ 的大小对检验作出结论, 此处 k_0 是由样本算得 K 的具体值.

(2) 若分布未知, 即

$$p_1(j) = \cdots = p_r(j) = p_j, \quad j = 1, 2, \dots, s,$$

而 p_1, \dots, p_s 皆未知. 这时首先对 p_1, \dots, p_s 求出其 MLE, 方法如下: 若 H_0 成立, 似然函数为

$$L = c \cdot p_1^{n_{\cdot 1}} p_2^{n_{\cdot 2}} \cdots p_s^{n_{\cdot s}}, \quad \sum_{j=1}^s p_j = 1.$$

通过 $\frac{\partial \log L}{\partial p_j} = 0, j = 1, 2, \dots, s$, 类似于例 3.3.8 可解得

$$\hat{p}_j^* = \frac{n_{\cdot j}}{n}, \quad j = 1, 2, \dots, s,$$

算出

$$\hat{K}_{ni}^* = \sum_{j=1}^s \frac{(n_{ij} - n_{\cdot i} \cdot \hat{p}_j^*)^2}{n_{\cdot i} \cdot \hat{p}_j^*} = n \sum_{j=1}^s \frac{(n_{ij} - n_{\cdot i} \cdot n_{\cdot j} / n)^2}{n_{\cdot i} \cdot n_{\cdot j}}, \quad i = 1, \dots, r,$$

故总的 \hat{K}_n^* 为

$$\begin{aligned} \hat{K}^* &= \hat{K}_n^* = \sum_{i=1}^r \hat{K}_{ni}^* = n \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{\cdot i} \cdot n_{\cdot j} / n)^2}{n_{\cdot i} \cdot n_{\cdot j}} \\ &= n \left(\sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_{\cdot i} \cdot n_{\cdot j}} - 1 \right). \end{aligned} \quad (6.5.9)$$

在齐一性问题中, 第 i 个总体中抽取的样本 $n_{\cdot i}$ 是事先给定的, 它没有随机性. 而在独立性问题中 $n_{\cdot i}$ 并未事先指定, 它是随机变量. 有了这个重要差别, 定理 6.4.2 就不能直接用于齐一性检验问题中, 必须从头研究 \hat{K}_n^* 的分布, 但经过复杂的论证 (见文献 [1] 第 310 页), 证明了在齐一性问题中若假设 (6.5.7) 成立, 则当 $n_{\cdot i} \rightarrow \infty (i = 1, 2, \dots, r)$ 时, 则 \hat{K}_n^* 的极限分布仍为 $\chi_{(r-1)(s-1)}^2$. 故知检验问题 (6.5.7) 的水平近似为 α 检验为

$$\text{当 } \hat{K}^* > \chi_{(r-1)(s-1)}^2(\alpha) \text{ 时否定 } H_0, \text{ 否则就接受 } H_0. \quad (6.5.10)$$

若记 \hat{k}^* 为由样本算得 \hat{K}^* 的具体值, 则检验的拟合优度为

$$p(\hat{k}^*) = P(\hat{K}^* \geq \hat{k}^* | H_0) \approx P(\chi_{(r-1)(s-1)}^2 \geq \hat{k}^*),$$

对给定的检验水平 α , 当 $p(\hat{k}^*) < \alpha$ 时否定 H_0 , 当 $p(\hat{k}^*) \geq \alpha$ 时接受 H_0 , $p(\hat{k}^*)$ 越大, 则接受 H_0 的结论越可靠.

因此齐一性和独立性检验, 虽然在计算方法上是一样的, 但所依据的统计模型和极限定理是有差别的.

如果指标取值不是离散的而是连续的, 若记 F_i 为第 i 个总体中指标的分布, 则式 (6.5.7) 可写成

$$H_0 : F_1 \equiv F_2 \equiv \cdots \equiv F_r.$$

这种问题称为多样本问题, 有多种检验方法, χ^2 方法是其中一种. 为此, 像式 (6.4.8) 那样, 把指标的值域分组以实现离散化, 并将 n_{ij} 定义为: 第 i 个总体所抽出的样本中, 其指标值落在 I_j 内的个数, 然后用列联表的方式去处理.

例 6.5.2 设一工厂分三班生产. 产品分成 $1, 2, \dots, 6$ 等 6 个等级. 要检验三个班产品质量是否相同. 为此在甲、乙和丙三班中分别抽取 83, 95 和 83 个产品来检查, 结果见表 6.5.5 (取 $\alpha = 0.01$).

表 6.5.5

	1	2	3	4	5	6	\sum
甲	11	23	8	5	18	18	83
乙	17	29	10	17	7	15	95
丙	6	21	8	24	15	9	83
\sum	34	73	26	46	40	42	261

解 检验问题为

$$H_0 : \text{甲、乙、丙三班质量齐一}.$$

此例中 $r = 3, s = 6$, 而 $(r - 1)(s - 1) = 2 \times 5 = 10$, 查表得 $\chi^2_{10}(0.01) = 23.21$. 按式 (6.5.9) 计算检验统计量

$$\begin{aligned}\hat{K}^* &= 261 \left[\left(\frac{11^2}{34 \times 83} + \frac{23^2}{73 \times 83} + \cdots + \frac{9^2}{42 \times 83} \right) - 1 \right] \\ &= 26.3 > \chi^2_{10}(0.01) = 23.21,\end{aligned}$$

检验的拟合优度为

$$p(\hat{k}^*) = P(\hat{K}^* \geq \hat{k}^* | H_0) \approx P(\chi^2_{10} \geq 26.3) < 0.01.$$

因此否定 H_0 , 即认为三个班次产品质量同一的假设不成立, 因此可认为产品质量分布与班次有关, 此处 \hat{k}^* 为由样本算得 \hat{K}^* 的具体值.

*6.6 其他的非参数检验方法

关于总体分布的检验, 除了 Pearson 的 χ^2 检验方法外, 柯尔莫哥洛夫在 1933 年还提出的一种检验方法, 本节将作一简要介绍.

关于多样本检验问题, 即检验多组样本是否来自同一总体的假设检验问题, 在 6.5 节列联表的齐一性问题中已作介绍. 这一问题的另一种检验方法, 就是著名的关于两个总体情况下的斯米尔诺夫检验. 本节也将作一扼要的介绍.

检验一组样本是否服从正态分布在应用上具有十分重要的意义. 除了可以使用 Pearson χ^2 检验或柯尔莫哥洛夫检验外, 本节还将简要介绍关于总体分布正态性的 W 检验和 D 检验.

6.6.1 柯尔莫哥洛夫检验

如 6.4 节所述, 不管总体分布是什么类型, Pearson χ^2 检验都可以用. 不过对于理论分布为连续分布时, 本段介绍的柯尔莫哥洛夫检验效果更好些. 这是因为 Pearson χ^2 检验需要按式 (6.4.8) 的方法分组, 因此检验统计量之值依赖于把 $(-\infty, +\infty)$ 分成 r 个区间的具体划分方法, 包括 r 的选择和区间的位置. 苏联著名数学家柯尔莫哥洛夫在 1933 年提出了一种新的关于总体分布的拟合优度检验方法——柯尔莫哥洛夫检验 (简称柯氏检验法).

设 r.v. X 的分布函数 $F(x)$ 未知, X_1, \dots, X_n 为从 F 中抽取的简单随机样本, $F_0(x)$ 为给定的某个分布函数. 来研究下列检验问题:

$$H_0 : F(x) = F_0(x). \quad (6.6.1)$$

首先, 从样本出发求出 $F(x)$ 的经验分布函数如下:

$$F_n(x) = \begin{cases} 0, & x \leq X_{(1)}, \\ \frac{k}{n}, & X_{(k)} < x \leq X_{(k+1)}, \quad k = 1, 2, \dots, n-1, \\ 1, & x > X_{(n)}, \end{cases} \quad (6.6.2)$$

其中 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ 是样本 X_1, \dots, X_n 的次序统计量. $F_n(x)$ 的性质见 1.3.3 小节.

令检验统计量为

$$D_n = \sup_{-\infty < x < +\infty} |F_n(x) - F_0(x)|, \quad (6.6.3)$$

D_n 常称为 F_n 与 F_0 之间的柯氏距离. 由定理 1.3.1 (Glivenko-Cantelli 定理) 可知: 如果 H_0 成立, 则 $P(\lim_{n \rightarrow \infty} D_n = 0) = 1$. 换言之, 如果 H_0 成立, n 又较大, D_n 的值

倾向于取小值. 如果 D_n 值太大, 倾向于否定 H_0 . 即检验可叙述为: 当 $D_n \geq c$ 时否定 H_0 , c 为临界值, 待定. 其拟合优度的计算公式如下: 在有了具体样本后, 计算出 D_n 的具体值 D_0 , 则概率

$$p(D_0) = P(D_n > D_0 | H_0) \quad (6.6.4)$$

就是在柯氏距离下, 样本 X_1, \dots, X_n 与理论分布 $F_0(x)$ 的拟合优度. 若指定一个阈值 α (亦称检验水平), 则需定出一个常数 $D_{n,\alpha}$, 使得

$$p(D_{n,\alpha}) = P(D_n > D_{n,\alpha} | H_0) = \alpha, \quad (6.6.5)$$

则当 $D_n > D_{n,\alpha}$ 时否定 H_0 , 不然就接受 H_0 , 这就是柯氏拟合优度检验. 当 n 较小时, $D_{n,\alpha}$ 已制成表, 见附表 13.

为了具体实施这个检验, 或计算拟合优度 $p(D_0)$, 需要知道在 H_0 成立的条件下, 检验统计量 D_n 的精确分布, 这个分布的形式极为复杂, 不便应用. 1933 年柯尔莫哥洛夫 (A. N. Kolmogorov) 证明了下列著名的极限定理.

定理 6.6.1 若理论分布 $F_0(x)$ 在 $-\infty < x < +\infty$ 处处连续, 则当原假设 H_0 成立时有

$$\lim_{n \rightarrow \infty} P\left(D_n \leq \frac{\lambda}{\sqrt{n}}\right) = K(\lambda) = \begin{cases} \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 \lambda^2}, & \lambda > 0, \\ 0, & \lambda \leq 0. \end{cases} \quad (6.6.6)$$

由这个定理, 在 n 较大时可近似地决定检验的临界值 $D_{n,\alpha} = \lambda/\sqrt{n}$. 由式 (6.6.5) 可知当 $D_n > D_{n,\alpha}$ 时否定 H_0 . 一般当给定 α 时查附表 14 获得 λ , 然后由公式 $D_{n,\alpha} = \lambda/\sqrt{n}$ 求出临界值 $D_{n,\alpha}$. 例如, $\alpha = 0.05$, 由 $K(\lambda) = 1 - \alpha = 0.95$ 查附表 14 得到 $\lambda = 1.358$, 故 $D_{n,\alpha} = 1.358/\sqrt{n}$; 若 $\alpha = 0.01$, 由 $K(\lambda) = 1 - \alpha = 0.99$, 从附表 14 中查出 $\lambda = 1.628$, 故 $D_{n,\alpha} = 1.628/\sqrt{n}$ 等.

综上所述, 柯尔莫哥洛夫检验适用于理论分布 $F_0(x)$ 为完全已知的连续分布的情形. 这个检验方法的步骤如下:

- (1) 将样本观察值 x_1, \dots, x_n 按大小排列为 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$;
- (2) 计算 D_n 的值

$$D_n = \max_i \{ \delta_i, i = 1, 2, \dots, n \},$$

其中

$$\delta_i = \max \left\{ \left| F_0(x_{(i)}) - \frac{i-1}{n} \right|, \left| F_0(x_{(i)}) - \frac{i}{n} \right| \right\}, \quad i = 1, 2, \dots, n; \quad (6.6.7)$$

(3) 给出检验水平 α , 在 $n \leq 100$ 时查附表 13, 在 $n > 100$ 时查附表 14, 得到检验的临界值 $D_{n,\alpha}$;

(4) 若 $D_n > D_{n,\alpha}$ 则否定 H_0 , 认为数据与已知的理论分布 $F_0(x)$ 不符; 若 $D_n \leq D_{n,\alpha}$ 则不足以否定 H_0 .

需要说明的一点是 δ_i 由式 (6.6.7) 计算的理由. 由于 $F_0(x)$ 是单调非降函数, $F_n(x)$ 是单调非降的阶梯函数, 所以偏差 $|F_n(x) - F_0(x)|$ 的上确界可在 n 个点 $x^{(i)}$ 处找到. 因此 δ_i ($i = 1, 2, \dots, n$) 可由式 (6.6.7) 获得.

例 6.6.1 在水平 0.10 下, 是否可以认为下列 10 个数:

0.034, 0.0437, 0.863, 0.964, 0.366, 0.469, 0.637, 0.632, 0.804, 0.261

是来自均匀分布 $U(0, 1)$ 的随机数 (取 $\alpha = 0.10$).

解 检验问题是

$$H_0: \text{样本来自的总体分布 } F(x) = F_0(x),$$

其中 $F_0(x)$ 为 $(0, 1)$ 上的均匀分布.

用柯尔莫哥洛夫检验法, 计算可在表 6.6.1 中进行. 因此 $D_n = \max_i \{\delta_i, i = 1, 2, \dots, 10\} = 0.166$. 由 $n = 10$, $\alpha = 0.10$ 查附表 13 得 $D_{10, 0.10} = 0.37 > 0.166 = D_n$, 不足以否定 H_0 , 故接受 H_0 , 即认为数据与理论分布 $U(0, 1)$ 相符.

Pearson χ^2 检验与柯尔莫哥洛夫检验的比较: 大体上可以这样说: 在总体 X 为一维且理论分布为完全已知的连续分布时, 柯尔莫哥洛夫检验优于 χ^2 检验. 这是因为: ① Pearson χ^2 统计量之值依赖于把 $(-\infty, +\infty)$ 分为 r 个区间的具体分法, 包括 r 的选取和区间的位置, 柯氏距离 $\sup |F_n - F_0|$ 则没有这个依赖性. ② 一般说来柯氏方法鉴别力强. 也就是说, 在 F_0 不是总体 X 的分布时, 用柯氏检验法较容易发现.

表 6.6.1 柯尔莫哥洛夫检验的计算

i	$x^{(i)}$	$F_0(x^{(i)})$	$(i-1)/n$	i/n	δ_i
1	0.034	0.034	0	0.1	0.066
2	0.261	0.261	0.1	0.2	0.161
3	0.366	0.366	0.2	0.3	0.166
4	0.437	0.437	0.3	0.4	0.137
5	0.469	0.469	0.4	0.5	0.069
6	0.623	0.623	0.5	0.6	0.123
7	0.637	0.637	0.6	0.7	0.063
8	0.804	0.804	0.7	0.8	0.104
9	0.863	0.863	0.8	0.9	0.063
10	0.964	0.964	0.9	1.0	0.064

另一方面, Pearson χ^2 检验也有它的优点: ① 当总体 X 是多维时, 处理方法与一维一样, 极限分布的形式也与维数无关; ② 尤其重要的是: 对于理论分布包含未知参数时, χ^2 检验容易处理, 但柯氏方法处理起来很难.

6.6.2 斯米尔诺夫检验

设 X_{i1}, \dots, X_{in_i} 为抽自具有一维连续分布总体 F_i 的简单随机样本, $i = 1, 2$ 且合样本独立. 设 $F_1(x)$, $F_2(x)$ 是未知的两个连续函数. 考虑检验问题

$$H_0 : F_1(x) = F_2(x), \quad -\infty < x < +\infty. \quad (6.6.8)$$

设 $F_{1n_1}(x)$ 和 $F_{2n_2}(x)$ 分别记这两组样本的经验分布函数, 令

$$D_{n_1, n_2}^+ = \sup_{-\infty < x < +\infty} (F_{1n_1}(x) - F_{2n_2}(x)),$$

$$D_{n_1, n_2} = \sup_{-\infty < x < +\infty} |F_{1n_1}(x) - F_{2n_2}(x)|.$$

苏联数学家斯米尔诺夫 (N. Smirnov) 于 1936 年证明了下列结果.

定理 6.6.2 设原假设 (6.6.8) 成立, 则有

$$\lim_{\substack{n_1 \rightarrow \infty \\ n_2 \rightarrow \infty}} P \left(\sqrt{\frac{n_1 n_2}{n_1 + n_2}} D_{n_1, n_2}^+ \leq x \right) = \begin{cases} 1 - e^{-2x^2}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

$$\lim_{\substack{n_1 \rightarrow \infty \\ n_2 \rightarrow \infty}} P \left(\sqrt{\frac{n_1 n_2}{n_1 + n_2}} D_{n_1, n_2} \leq x \right) = K(x),$$

其中 $K(x)$ 与式 (6.6.6) 同. D_{n_1, n_2}^+ 和 D_{n_1, n_2} 分别称为单边和双边的斯米尔诺夫检验统计量.

如果要检验的原假设是式 (6.6.8), 取 D_{n_1, n_2} 作为检验统计量, 则当 $D_{n_1, n_2} > D_{n_1, n_2; \alpha}$ 时否定 H_0 . 临界值

$$D_{n_1, n_2; \alpha} = \frac{\lambda}{\sqrt{n_1 n_2 / (n_1 + n_2)}},$$

其中 λ 的值可由附表 14 查出. 这就是斯米尔诺夫检验.

若假设检验问题为

$$H_0 : F_1(x) \leq F_2(x) \leftrightarrow K : F_1(x) > F_2(x), \quad x \in (-\infty, \infty),$$

则用 D_{n_1, n_2}^+ 作为检验统计量.

*6.6.3 正态性检验

在实际工作中常常要检验一个随机变量是否服从正态分布, 这称为正态性检验。前面介绍的 Pearson χ^2 检验、柯氏检验法等当然可以使用。但是由于上述方法是通用的, 适用面广, 故有针对性不强的缺点。这些方法都没有充分利用原假设成立时的信息, 检验功效不高。对正态分布往往可以找到针对这类特定分布功效较高的检验。下面介绍的两种基于次序统计量的正态性检验: 小样本 (样本大小为 30–50) 的 W 检验和大样本 (样本大小为 50–1000) 的 D 检验可以克服上述缺点, 提高检验的功效。这两个方法已被列入我国统计方法的国家标准 GB4882–85 之中, 见文献 [9]。

1. W 检验 (Wilk 检验)

考虑检验问题

$$H_0: X \text{ 服从正态分布} \leftrightarrow H_1: X \text{ 不服从正态分布}. \quad (6.6.9)$$

设 X_1, \dots, X_n 为来自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本, $X_{(1)} \leq \dots \leq X_{(n)}$ 为其次序统计量。设 $Y_i = (X_i - \mu)/\sigma$, $i = 1, \dots, n$, 则 Y_1, \dots, Y_n i.i.d. $\sim N(0, 1)$ 。令

$$\begin{aligned} Y_{(i)} &= \frac{X_{(i)} - \mu}{\sigma}, \quad e_i = X_{(i)} - E(X_{(i)}), \\ m_i &= E(Y_{(i)}), \quad i = 1, 2, \dots, n. \end{aligned}$$

注意 m_1, \dots, m_n 是与 μ, σ^2 无关的确定的数。显然有

$$X_{(i)} = \mu + \sigma m_i + e_i, \quad i = 1, 2, \dots, n, \quad (6.6.10)$$

其中 $e = (e_1, \dots, e_n)'$ 是均值为 0, 协方差阵为 V 的 n 维向量。

作一直角坐标系, 横轴表示 $X_{(i)}$, 纵轴表示 m_i 。由式 (6.6.10) 可见, 在这个坐标系中 $(X_{(1)}, m_1), (X_{(2)}, m_2), \dots, (X_{(n)}, m_n)$ 应该大致成一条直线, 微小的差别是由随机误差 e_i 造成的。怎样判别 n 个点是否近似在一条直线上呢? 可以计算一下 $\mathbf{X} = (X_{(1)}, \dots, X_{(n)})'$ 和 $\mathbf{m} = (m_1, \dots, m_n)'$ 之间的相关系数 R ,

$$R^2 = \frac{\left(\sum_{i=1}^n (X_{(i)} - \bar{X})(m_i - \bar{m}) \right)^2}{\sum_{i=1}^n (X_{(i)} - \bar{X})^2 \sum_{i=1}^n (m_{(i)} - \bar{m})^2}.$$

显然 $0 \leq R^2 \leq 1$, 当 R^2 越接近 1, \mathbf{X} 与 \mathbf{m} 的线性关系越明显。因此当 H_0 成立, 诸 X_i 服从 $N(\mu, \sigma^2)$ 时, R^2 接近 1。可见当 $R^2 < c$ (c 为较小的正数, 待定) 时倾向于否定 H_0 。

由于 $N(0, 1)$ 是对称分布, 所以 $(Y_{(1)}, \dots, Y_{(n)})$ 与 $(-Y_{(n)}, \dots, -Y_{(1)})$ 有相同的联合分布, 从而 $Y_{(k)}$ 与 $-Y_{(n+1-k)}$ 同分布, 故 $m_k = -m_{n+1-k}$, $k = 1, \dots, n$, $\bar{m} = \sum_{i=1}^n m_i/n = 0$, 于是

$$R^2 = \frac{\left(\sum_{i=1}^n m_i X_{(i)}\right)^2}{\sum_{i=1}^n (X_{(i)} - \bar{X})^2 \sum_{i=1}^n m_i^2} = \frac{\left[\sum_{i=1}^{[n/2]} b_i (X_{(n+1-i)} - X_{(i)})\right]^2}{\sum_{i=1}^n (X_{(i)} - \bar{X})^2}, \quad (6.6.11)$$

其中 $b_i = m_{n+1-i}/||\mathbf{m}||$, 此处 $||\mathbf{m}||^2 = \sum_{i=1}^n m_i^2$. 因此 $W = R^2$ 可作为检验统计量.

夏皮洛 (Shapiro) 和威尔克 (Wilk) 对式 (6.6.11) 作了修正得到检验统计量 (详见文献 [4] 第 294 页)

$$W = \frac{\left\{ \sum_{i=1}^{[n/2]} a_i (X_{(n+1-i)} - X_{(i)}) \right\}^2}{\sum_{i=1}^n (X_{(i)} - \bar{X})^2}. \quad (6.6.12)$$

在 $n \leq 50$ 时, $\{a_i : i \leq [n/2]\}$ 的值已制成表, 详见附表 15. 式 (6.6.12) 可以用来简化统计量 W 的计算.

可以证明, 检验统计量 W 的一个重要性质: 即在正态假设 H_0 成立时, W 的分布仅与样本容量 n 有关 (详见文献 [4] 中引理 5.5.4). 因而在讨论有关统计量 W 的问题时无妨假定样本来自 $N(0, 1)$ 分布.

如前所述, W 是 n 个数对之间的相关系数的平方, 因此 $0 \leq W \leq 1$. 由线性模型理论可知在正态假设 H_0 下, 这 n 个数对之间基本上存在线性关系, 故 W 取值应接近于 1. 因此, 给定检验水平 α 后, 检验问题 (6.6.9) 的 W 检验是

$$\text{当 } W \leq W_\alpha \text{ 时否定 } H_0, \text{ 否则接受 } H_0, \quad (6.6.13)$$

其中 W 按式 (6.6.12) 计算, 临界值 W_α 可由附表 16 查出. 附表 16 是根据 W 的分布仅与样本容量 n 有关的这个性质, 利用随机模拟法编制而成的.

例 6.6.2 为了检验一批煤灰砖中各块砖的抗压强度是否服从正态分布, 从中随机取 10 块测得抗压强度 (由小到大排列) 为

$$57, 66, 74, 77, 81, 87, 91, 95, 97, 109.$$

试检验这些数据是否与正态分布相符? ($\alpha = 0.05$)

解 将数据填入表 6.6.2.

表 6.6.2

i	$x_{(i)}$	$x_{(11-i)}$	$x_{(11-i)} - x_{(i)}$	a_i
1	57	109	52	0.5739
2	66	97	31	0.3291
3	74	95	21	0.2141
4	77	91	14	0.1224
5	81	87	6	0.0399

其中 a_i 这一列的值由附表 15 根据 $n = 10$ 查得. 经计算得

$$\sum_{i=1}^{10} x_{(i)} = 834, \quad \frac{1}{10} \sum_{i=1}^{10} x_{(i)} = \bar{x} = \frac{1}{10} 834 = 83.4,$$

$$\sum_{i=1}^{10} (x_{(i)} - \bar{x})^2 = \sum_{i=1}^{10} x_{(i)}^2 - 10\bar{x}^2 = 71736 - 10 \times 6955.56 = 2180.4,$$

$$\sum_{i=1}^5 a_i (x_{(11-i)} - x_{(i)}) = 46.494, \quad \left[\sum_{i=1}^5 a_i (x_{(11-i)} - x_{(i)}) \right]^2 = 2161.692.$$

于是有

$$W = \frac{\left[\sum_{i=1}^5 a_i (x_{(11-i)} - x_{(i)}) \right]^2}{\sum_{i=1}^{10} (x_{(i)} - \bar{x})^2} = \frac{2161.7}{2180.4} = 0.99.$$

由 $\alpha = 0.05$, $n = 10$, 查附表 16 得 $W_{0.05} = 0.842 < 0.99 = W$, 所以不能拒绝正态性假定.

2. D 检验

检验问题和样本仍如 W 检验中所述. W 检验是有效的, 可惜它只适用于样本容量 $3 \leq n \leq 50$ 的样本. 当 $n > 50$ 时很难计算附表 15 中的相应的值. 为此人们提出了 D 检验. 达戈斯蒂诺 (D'Agostino) 建议在 $n > 50$ 时用

$$D = \sum_{i=1}^n \left(i - \frac{n+1}{2} \right) X_{(i)} \left/ \left[n^3 \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{1/2} \right. \quad (6.6.14)$$

作为检验统计量. 由此导出的检验方法, 称为 D 检验.

可以证明, 在正态假设 H_0 成立时, D 的分布仅与样本容量 n 有关, 且

$$E(D) \approx 0.28209479, \quad \sqrt{\text{Var}(D)} \approx \frac{0.02998598}{\sqrt{n}},$$

其中 $\text{Var}(D)$ 表示 D 的方差. 将 D 标准化得

$$Y = \frac{\sqrt{n}(D - 0.28209479)}{0.02998598}.$$

可以证明: 当正态性假定 H_0 成立且 $n \rightarrow \infty$ 时有

$$Y \xrightarrow{\mathcal{L}} N(0, 1).$$

但是统计量 Y 趋向于标准正态分布的速度很慢, 以致于 $n = 1000$ 时, Y 的分布与标准正态分布仍有不可忽略的偏差. 故 D'Agostino 用随机模拟法获得 Y 的分位数值 (见附表 17).

大量模拟表明, 在 H_0 成立时, Y 的值集中在零左右, 在正态性假定不成立时, Y 的值不是偏小就是偏大, 因此检验问题 (6.6.9) 水平为 α 的 D 检验是

当 $Y \leq Y_{\alpha/2}$ 或 $Y \geq Y_{1-\alpha/2}$ 时, 否定 H_0 ; 否则就接受 H_0 , (6.6.15)

其中 $Y_{\alpha/2}$ 和 $Y_{1-\alpha/2}$ 分别是 Y 的 $\alpha/2$ 和 $1 - \alpha/2$ 分位数, 其值可从附表 17 查出.

注 6.6.1 (正态概率纸检验法) 用正态概率纸检验正态性是简便易行的方法, 所以不少实际工作者喜欢使用这个方法. 但是, 由于这种方法靠的是人的视觉, 不少场合下得出的结论会因人而异, 是一种定性的方法. 因此, 在可能的情形下应尽量采用定量的方法, 如前面介绍的 W 检验和 D 检验方法来克服这一缺点. 用正态概率纸检验正态性的具体步骤如下:

(1) 把从总体中获得的容量为 n 的样本按大小排列成次序统计量

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}.$$

(2) 将数对 $(x_{(i)}, i/(n+1))$ ($i = 1, \dots, n$) 点在正态概率纸上.

(3) 目测这些点的位置, 看它们是否近似成一条直线. 如果这些点与直线偏离较小, 则接受正态性假设, 否则就拒绝正态性假设.

关于正态概率纸的原理见文献 [4] 第 267 页.

现在统计软件的使用已很方便. 可以用统计软件 (如 R), 将数据输入计算机后作出正态性检验的 Q-Q, 如果 Q-Q 图上的点近似在一条直线附近, 则可认为样本来自正态总体.

习题 6

- 在 8 块土地上同时试种甲、乙两种品种作物, 每块分成大小、形状一样的两小块, 随机地将其中一小块派给甲, 另一小块给乙. 它们的产量列于下表:

编号	1	2	3	4	5	6	7	8
品种甲	209	200	177	169	159	187	169	198
品种乙	151	168	147	164	166	176	169	188

- (1) 假定甲、乙两种作物产量的差服从正态分布, 试在 $\alpha = 0.05$ 下判定, 甲品种是否是对乙品种的改良 (取 $\alpha = 0.05$).
- (2) 去掉 (1) 中正态假定, 分别用符号检验法、符号秩和检验法判别甲品种是否是对乙品种的改良 (取 $\alpha = 0.05$).
2. 从某铜矿东西两段各抽取容量为 10 的样本, 随机配成 10 对, 数据如下表:

对子	1	2	3	4	5	6	7	8	9	10
东段含铜量	28	20	4	32	8	12	16	48	8	20
西段含铜量	20	11	13	10	45	15	11	13	25	8

- 用符号检验法, 在水平 $\alpha = 0.10$ 下, 检验“东西段含铜量无显著差异”之假设.
3. 比较新旧两种饲料对猪的催肥效果 (以饲养一段时间后猪的增重表示). 由于猪的体质, 特别猪的初始重量对猪的催肥有相当重要的影响, 所以取体质尽可能相同的两头猪做对比试验. 下表是 9 个对比试验的记录 (增重 x_i 和 y_i 以斤为单位):

试验号	1	2	3	4	5	6	7	8	9
x_i (新饲料)	35	24	24	33	30	24	29	37	21
y_i (旧饲料)	30	26	24	29	28	25	26	30	27

- 问新饲料能否推广 (用符号检验和符号秩和检验, $\alpha = 0.05$)?
4. 甲、乙两厂都生产电视机显像管. 从甲厂抽取 8 只, 从乙厂抽取 10 只, 测出其寿命 (单位: 月) 如下:

甲	32	25	40	31	35	29	37	39	
乙	41	39	36	47	45	34	48	44	30

- 试在水平 $\alpha = 0.10$ 下用 Wilcoxon 秩和检验法, 检验“两厂显像管平均寿命相同”之假设.
5. 用两种材料的灯丝制造灯泡, 设甲材料灯泡的寿命分布为 $F(x)$, 乙材料灯泡寿命为 $F(x-\theta)$. 今分别就两种材料灯泡随机抽取若干个进行试验, 得到寿命 (单位: h) 数据如下表:

甲材料	1610	1650	1680	1710	1750	1720	1800
乙材料	1580	1600	1640	1630	1700		

- 试用 Wilcoxon 秩和检验法检验两种材料对灯泡寿命的影响有无显著差异 ($\alpha = 0.05$)?
6. 分别在甲、乙两厂生产的电子元器件中抽取 150 个和 130 个进行寿命试验. 算出乙工厂样本秩和 $W = 20251$. 试在水平 $\alpha = 0.05$ 下检验假设

H : 这两厂所生产的电子元器件的寿命无显著差异.

7. 掷骰子 300 次, 结果如下表:

点数	1	2	3	4	5	6
频数	43	49	56	45	66	41

试在水平 $\alpha = 0.05$ 下检验假设: 此骰子是均匀的.

8. 64 只某种杂交的几内亚猪后代, 34 只红的, 10 只黑的, 20 只白的. 根据遗传学的模型, 它们之间的比例应为 9:3:4. 问以上数据是否在 $\alpha = 0.05$ 水平下与模型相符.
9. 在生产同一产品的甲、乙两厂中各抽出 n 个产品组成 n 对, 让 n 个人每人评判一对. 各人评判是“甲好”“乙好”“甲乙一样”这三种之一. 试根据评判结果, 用 χ^2 检验法检验“甲、乙两厂产品质量无差别”这一假设.
10. 某种配偶的后代按体格的属性分为三类, 各类的个数为 10, 53, 46. 根据某种遗传学模型, 其相对频率的比应为 $p^2:2p(1-p):(1-p)^2$. 问以上数据是否在 $\alpha = 0.05$ 水平下与模型相符.
11. 用运动手枪对 100 个靶子各射击 10 发子弹, 只记录命中与否, 射击结果的记录如下表:

命中数	0	1	2	3	4	5	6	7	8	9	10
靶数	0	2	4	10	22	26	18	12	4	2	0

用 χ^2 检验法检验射击结果是否服从二项分布 (取 $\alpha = 0.05$).

12. 观察某城市每日交通情况, 210 天的纪录如下表:

事故数	0	1	2	3	4	≥ 5
天数	109	65	22	3	4	7

初步推测每日发生事故数服从 Poisson 分布, 试用 χ^2 检验法检验之 ($\alpha = 0.05$).

13. 卢瑟福观察了每 $1/8$ 分钟内一放射性物质放射的粒子数. 他观察了 2612 次, 结果如下表:

粒子数	0	1	2	3	4	5	6	7	8	9	10	11
频数	57	203	383	525	532	408	273	139	49	27	10	6

问以上数据是否在 $\alpha = 0.10$ 水平下与 Poisson 分布相符.

14. 下表列出了某型号标准容量为 100μ 的电介容器 100 只的实际电容量的数据, 问该电介容器的电容量的分布是否为正态分布 ($\alpha = 0.05$)?

电容量	101	102	103	104	105	106	107
电容个数	1	2	3	3	7	16	13
电容量	108	109	110	111	112	113	114
电容个数	17	11	9	10	3	4	1

15. 试证明 2×2 列联表的检验统计量为

$$K^* = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_1 \cdot n_2 \cdot n_{11} \cdot n_{22}}.$$

16. 为研究慢性气管炎与吸烟量的关系, 调查 272 人结果如下:

吸烟量/(支/日)	0-9	10-19	20 以上
患者人数	22	98	25
健康者人数	22	89	16

试问慢性气管炎与吸烟量是否有关 ($\alpha = 0.05$).

17. 在 500 人身上的试验某种血清预防感冒的作用. 把他们在一年中的记录与另外 500 名未用血清处理的人作比较, 结果如下表:

	未感冒	感冒一次	感冒两次以上
处理	252	145	103
未处理	224	136	140

未感冒、感冒一次, 感冒两次以上的人所占的比例, 是否与在未用血清处理的人中, 具有这些情况的人所占比例相等 ($\alpha = 0.05$)?

18. 使用柯尔莫哥洛夫检验法检验下述 25 个数据 (已按大小排列) :

-2.46, -2.11, -1.23, -0.99, -0.42, -0.39, -0.21, -0.15, -0.10
-0.07, -0.02, 0.27, 0.40, 0.42, 0.44, 0.70, 0.81, 0.88,
1.07, 1.39, 1.40, 1.47, 1.62, 1.64, 1.76.

试问上述 25 个数据是否与 $N(0, 1)$ 相符 ($\alpha = 0.05$)?

19. 在 π 的前 800 位小数的数字中, 0, 1, 2, ⋯, 9 相应出现了 74, 92, 83, 79, 80, 73, 77, 75, 76, 91 次. 试用柯氏检验法检验这些数据与 $[0, 10]$ 上的均匀分布相适合的假设 ($\alpha = 0.05$).

20. 抽查用克矽平治疗的矽肺患者 10 名, 得到他们治疗前后血红蛋白的差 (单位: g %) 如下:

2.7, -1.2, -1.0, 0, 0.07, 2.0, 3.7, -0.6, 0.8, -0.3.

用 W 检验法检验“治疗前后血红蛋白差服从正态分布”这一假设 (取 $\alpha = 0.05$).

21. 为检验一批地砖的抗压强度是否服从正态分布, 从中抽取 20 块, 得抗压强度 (由小到大排列) 为

57, 62, 66, 67, 74, 76, 77, 80, 81, 86,
87, 89, 91, 94, 95, 96, 97, 103, 109, 112.

试用 W 检验检验“这些数据是否与正态分布相符”(取 $\alpha = 0.05$).

22. 下表数据是 200 个零件的直径 (单位: cm):

直径	2.25	2.35	2.45	2.55	2.65	2.75	2.85	2.95
频数	3	4	5	11	12	17	19	26
直径	3.05	3.15	3.25	3.35	3.45	3.55	3.65	3.75
频数	24	22	19	13	13	7	3	2

试问这批数据是否与正态分布相符 (用 D 检验, 取 $\alpha = 0.05$).

第7章 Bayes 方法和统计决策理论

7.1 引言和若干基本概念

7.1.1 引言

贝叶斯 (Bayes) 方法的基本观点是由 Bayes 公式引申而来的。学过初等概率论的人都知道 Bayes 公式, 此公式包含在英国学者 Bayes 于 1763 年 (在他去世后两年) 发表的文章《论有关机遇问题的求解》中。从形式上看这一公式不过是条件概率定义的一个简单推论, 但它却包含了归纳推理的一种思想, 后来学者把它发展成一种关于统计推断的系统理论和方法, 称为 Bayes 方法。由这种方法获得的统计推断的全部结果, 构成了 Bayes 统计学的内容。信奉 Bayes 统计, 乃至鼓吹 Bayes 观点是关于统计推断的唯一正确方法的那些学者形成了数理统计学中的 Bayes 学派。这一学派始于 20 世纪二三十年代, 到五六十年代引起人们广泛的关注。时至今日其影响日益扩大。Bayes 方法已渗透到了数理统计的几乎所有领域, 每个学习数理统计学的人, 都应当对这个学派的观点和方法有所了解。

Bayes 学派的观点在统计学界引起了广泛的争论。频率 (经典) 学派与 Bayes 学派近几十年来的争论推动了数理统计学的发展。这两大学派之间有共同点, 也有不同点, 为了弄清他们的主要差别, 首先介绍统计推断中所使用的三种信息。

数理统计的任务是要通过样本推断总体。样本有两重性, 当把样本视为随机变量, 它有概率分布, 称为总体分布。如果已经知道总体的分布形式, 这就给了一种信息, 称为总体信息。例如, 样本来自于正态总体, 它暗示很多信息, 如它的密度函数是倒立的钟形曲线, 它的所有阶矩存在, 任何事件的概率都可以通过查表求出。由正态总体还可导出与之相关的 χ^2 分布、 t 分布、 F 分布等。因此总体的信息是很重要的, 但是获得总体的信息是要付出代价的。例如, 在工业可靠性问题中, 要想获得电子元器件的寿命分布, 要利用成千上万的元器件, 做大量的实验, 再进行统计分析从而导出其分布, 要费钱、费力、费时。

另外一种信息是样本信息, 即从总体中抽取的样本提供的信息。这是最“鲜活”的信息。样本越多, 提供的信息越多。我们希望通过把样本加工、整理, 对总体的分布或总体的某些数字特征作出统计推断。没有样本就没有统计推断, 这是任何一种统计推断都需要有的信息。总体信息和样本信息统称为抽样信息。

基于上述两种信息进行统计推断的方法和理论称为经典 (古典) 统计学。它的

基本观点是：把样本看成来自有一定概率分布的总体，所研究的对象是总体。从 19 世纪至 20 世纪中叶，R.A. Fisher, K. Pearson, J. Neyman 等的杰出工作创立了经典统计学。直到 20 世纪 50 年代这个学派占据了数理统计学的主导地位。经典统计学在自然科学和社会科学中的各个领域得到发展，但也暴露出它的一些缺点，导致了新的统计学的产生。

第三种信息称为先验信息，就是在抽样之前，有关统计推断问题中未知参数的一些信息。一般先验信息来自经验和历史资料。下面两例说明先验信息存在且被人们利用。

例 7.1.1 英国统计学家 Savage 在 1961 年提出一个令人信服的例子说明先验信息有时是很重要的，且看下面两个试验：

(1) 一位常饮牛奶和茶的女士说她能辨别先倒进杯子里的是茶还是牛奶。对此做了 10 次试验她都说对了。

(2) 一位音乐家说他能够从一页乐谱辨别是海顿 (Haydn) 还是莫扎特 (Mozart) 的作品，在 10 次试验中他都说对了。

在上面两个试验中，如果认为试验者是猜对的，每次成功概率为 0.5，则 10 次都猜中的概率为 $(1/2)^{10} = 2^{-10} = 0.0009766$ ，这是一个很小的概率，几乎不可能发生。故每次猜对的概率为 0.5 的假设被否定。他们每次说对的概率比 0.5 要大得多，这不能认为是猜测，而是经验帮了忙。可见经验（先验信息）在推断中不可忽视，应当加以利用。

例 7.1.2 某工厂生产一种产品，每日抽查一部分 (n 件) 产品以检查废品率 θ ，经过一段时间后获得大量数据，对 θ 作出估计。就当日被抽查的那批产品的废品率 θ 而言，它只是一个固定的数，并无随机性可言；但逐日的废品率 θ 受随机因素的影响多少会有些波动。从长期看，将“一日废品率 θ ”视为随机变量，而要估计的某日的废品率是这个随机变量的一个观察值。根据历史资料可构造一个分布

$$P\left(\theta = \frac{i}{n}\right) = \pi_i, \quad i = 0, 1, 2, \dots, n.$$

这个对先验信息进行整理加工而得到的分布称为先验分布。先验分布总结了该厂过去产品质量情况。如果这个分布的概率大多集中在 $\theta = 0$ 附近，那该产品可以认为是“信得过产品”。假如以后多次抽样与历史资料提供的先验分布一致，使用单位就可以作出“免检产品”的决定，或者每月抽一、两次就足够了，省去大量人力和物力。

基于上述三种信息进行统计推断的方法和理论称为 Bayes 统计学。它与经典统计学的主要区别在于是否利用先验信息。在使用样本上也是存在差别的。Bayes 方法重视已出现的样本，对尚未发生的样本值不予考虑。Bayes 学派重视先验信息的

收集、挖掘和加工, 使之形成先验分布, 参加到统计推断中来, 以提高统计推断的效果. 忽视先验分布的利用, 有时是一种浪费.

Bayes 方法的一个主要问题是如何确定先验分布, 先验分布的确定有很大的主观随意性. 当先验分布完全未知或部分未知时, 如果人为给出的先验分布与实际情形偏离较大时, Bayes 解的性质就较差. 经验 Bayes 方法 (empirical Bayes procedure, EB 方法) 就是针对这一问题提出来的. 它的实质是利用历史样本对先验分布或先验分布的某些数字特征作出直接或间接的估计, 因此 EB 方法是对 Bayes 方法的改进和推广. 它是介于经典统计学和 Bayes 统计学之间的一种统计推断方法.

7.1.2 先验分布与后验分布

定义 7.1.1 (先验分布) 参数空间 Θ 上的任一概率分布都称为先验分布 (prior distribution).

先验分布有不同的类型, 比较重要的两个概念是无信息先验分布和共轭先验分布. 另外一个问题是如何确定和选择先验分布. 将在 7.2 节介绍.

先验分布 $F^\pi(\theta)$ (如有密度, 以 $\pi(\theta)$ 记其密度函数) 是在抽取样本 X 之前对参数 θ 可能取值的认识. 后验分布 $F^\pi(\theta|x)$ (如有密度, 以 $\pi(\theta|x)$ 记其后验密度函数) 是反映人们在获取样本后对 θ 的新认识. 这是由于样本 X 也包含 θ 的信息, 因此一旦获得样本 X 后, 人们对 θ 的认识就发生了变化和调整. 所以, 后验分布 $F^\pi(\theta|x)$ 可以看作是人们用总体信息和样本信息 (综合称为抽样信息) 对先验分布 $F^\pi(\theta)$ 作调整的结果.

为叙述方便, 本章中常假定有关的先验密度函数 $\pi(\theta)$ 和后验密度函数 $\pi(\theta|x)$ 皆存在.

定义 7.1.2 (后验分布) 在获得样本 X 后, θ 的后验分布 (posterior distribution) 就是给定 $X = x$ 条件下 θ 的条件分布, 后验密度 $\pi(\theta|x)$ 有下列计算公式:

$$\pi(\theta|x) = \frac{f(x, \theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}. \quad (7.1.1)$$

(注: 在离散情形下它就转变成 Bayes 公式), 其中

$$f(x, \theta) = f(x|\theta)\pi(\theta)$$

为 X 和 θ 的联合密度, 而

$$m(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta$$

为 X 的边缘密度.

从 Bayes 学派的观点看获取后验分布后, 对参数 θ 的任何统计推断(估计、检验等)必须基于且只能基于 θ 的后验分布.

7.1.3 古典学派和 Bayes 学派的争论

频率学派和 Bayes 学派是当今数理统计学的两大学派. 凡是坚持概率的频率解释, 因而对数理统计学中的概念、结果和方法性能的评价等, 必须在大量重复的意义上去理解的, 都属于频率学派. 20 世纪初数理统计大发展以来, 一些起领导作用的学者, 如 R.A. Fisher, K. Pearson, J. Neyman, E.S. Pearson 等都属于这一学派. 直到 20 世纪 50 年代为止, 这个学派占据主导地位. 20 世纪五六十年代以来 Bayes 学派迅速崛起, 达到可以与频率学派分庭抗礼的程度. 由于其发展较新, 因此 Bayes 学派常常把频率学派称为古典学派. 两个学派常常发生激烈的争论, 发表了许多文章和言论, 是当代数理统计学发展中的特有现象, 双方的主要论点至今并无定论.

频率学派对 Bayes 学派的批评, 主要集中在主观概率以及相关的先验分布的确定上. 频率学派认为一个事件的概率可以用大量重复试验下的频率来解释. 这种解释不应该因人而异, 即不同的人都应给以同样的解释. 主观概率则理解为认识主体对事情发生机会的个人信念, 即不同的人对同一事件的概率可以得到不同的结果. 坚持频率解释的人认为这难以捉摸, 是主观随意性的产物, 没有客观性, 因而也就没有科学性, 当然也不能接受建立在这个基础上的统计方法.

频率学派还认为在许多情况下将 θ 视为随机变量是不合理的. 例如, 估计某矿体内一种金属的含量, 很难把这一问题纳入 Bayes 观点的模式中. 此外, 就算在某些问题中将 θ 看作随机变量有一定的合理性, 但关于 θ 的先验知识, 往往不是确切到可以提出一定的先验分布来. 在这种情况下, 指定一种先验分布不免带有人为性, 而这未必有助于问题的合理解决.

Bayes 学派对频率学派的批评, 主要在以下方面: 首先涉及“频率解释”本身. Bayes 学派认为许多应用问题是一次性的, 在严格或大致相同条件下重复事实上是不可能的. 因此, 在许多情况下, 统计概念和方法的频率解释完全没有现实意义. 在诸多自然现象、社会现象和经济决策问题中, “事件”常常是不能大量重复的. 例如, 灾害预报中, 像地震、水灾等都不可在相同条件下重复.

Bayes 学派认为, 虽然频率学派没有明白使用先验分布, 但事实上, 经典统计中的一些重要统计推断方法之所以能站住脚, 只是它暗合于某个合理的 Bayes 解. 例如, $N(\theta, 1)$ 中的 θ 的无偏估计 \bar{X} , 恰好是当 θ 有广义先验密度 $\pi(\theta) \equiv 1$, 即所谓无信息先验分布之下的 Bayes 解.

频率学派承认 Bayes 方法在一些情况下可用, 但限于先验分布可给予某种频率解释的时候. 至于两派涉及的基本哲学观点, 看来是无法调和的. 关于这方面的详细介绍见文献 [2] 第 212 页.

频率学派与 Bayes 学派有不少共同点, 如都承认样本有概率分布, 概率的计算遵循共同的准则。分歧在于把未知参数 θ 看成一个未知的固定量, 还是看成一个随机变量, 其余的分歧多少都是由此派生而来。对上述争论有一个至高无上的裁判者, 即应用的结果如何。统计方法无论在理论上如何精细高明, 总要用实践来检验。迄今为止, 实践显示这两派的得分都不低, 也正是因为它们在应用上的表现不错, 才能各自聚合了一些追随者形成各自的学派。作为一个统计学者, 可以不必执著于任何一派的观点, 而是取其所长为我所用。

7.1.4 历史

如前所述, Bayes 统计起源于英国学者 T. Bayes (1701-1761) 在去世两年后 (1763 年) 发表的一篇论文。在此论文中他提出著名的 Bayes 公式和一种归纳推理的方法。Bayes 在 18 世纪上半叶的欧洲学术界并不是很知名的人物, 他在生前没有发表科学论著。那时学者之间的私人通信是传播和交流科学成果的一种重要方式。许多这类信件得以保存下来并发表传世, 成为科学史上的重要文献。对 Bayes 来说这方面的材料有一些, 但不多。Bayes 在 1742 年当选为英国皇家统计学会会员 (相当于科学院院士), 因而可以想见, 他必定以某种方式表现出其学术造诣而为当时的学术界所承认。他是一个生性孤僻、哲学气味重于数学气味的怪杰。他的上述遗作发表后, 很长一段时间在学术界没有引起什么反响, 但到 20 世纪中叶突然受到人们的重视, 成为 Bayes 学派的奠基石。为了纪念他, 1958 年国际权威性的统计杂志 *Biometrika* 全文重新刊登了这篇文章。

据记载, Bayes 在他逝世之前 4 个月, 在一封遗书中将此文及 100 英镑托付给一个叫普莱斯的学者, 而当时 Bayes 对此人在何处也不了解。所幸的是后来普莱斯在 Bayes 的文件中发现了此论文, 他于 1763 年 12 月 23 日在英国皇家学会上宣读了此文, 并在同年得到发表。Bayes 的遗著在当时学术界没有引起重视, 其主要原因可能是当时起领导作用的统计学家, 如 Fisher, Neyman 等对 Bayes 方法持否定态度。另外 20 世纪上半叶正是经典统计学得到大发展的一个时期, 发现了一些具有普遍应用意义的有力的统计方法, 如创建假设检验理论 (Neyman-Pearson 引理)、似然比检验、拟合优度检验、列联表检验及估计的最忧性理论等。在这种情况下, 人们不会感到有要“另寻出路”的想法。自 20 世纪中叶以来, 经典统计学的发展遇到一些问题, 如数学化程度越来越高, 但有用的方法产生相对减少; 小样本方法缺乏进展从而越来越转向大样本理论的研究, 在应用工作中产生了不满。在这种背景下 Bayes 统计以其操作简单, 加之在解释上的某些合理性吸引了不少应用统计学者, 甚至一些频率学派学者后来也成为 Bayes 学派的成员就可以理解了。

Bayes 学派自 20 世纪下半叶进入全盛时期, 在这中间起过重要作用的有 H. Jeffreys, 他在 1939 年出版的《概率论》一书, 成了 Bayes 学派的经典著作。

L.J. Savage 在 1954 年出版了《统计推断》一书，也是 Bayes 学派的力作。还有 Lindley, Box & Tiao 和 Berger 等，他们也写了不少鼓吹 Bayes 统计的书。关于 Bayes 统计学的起源和发展史，详见文献 [5] 第 50 页。

7.2 先验分布的确定

7.2.1 主观概率

1. 引言和定义

在经典的统计学中概率是用公理化定义的，即用非负性、正则性和可加性三条公理定义的。概率的确定有两种方法：一是古典方法（包括几何方法）；另一种是频率方法。实际中大量使用的是频率方法，如掷均匀硬币正面出现的概率是 $1/2$ ，是通过大量重复抛掷硬币试验，发现出现正面次数的频率在 $1/2$ 附近。故经典统计研究的对象是能大量重复的随机现象，不是这类随机现象就不能用频率方法去确定有关事件的概率。

在诸多社会现象、经济领域和决策问题中，“事件”常常是不能大量重复的，如气象预报“明天是晴天的概率为 0.8”，就不能用频率去解释，因为天气随时间变化而变化，不可重复。又如国家统计局预测“2003 年失业率 θ ”是 3%-5% (θ 可以认为是随机变量)，这一事件也是不可重复的，因为不同年份的经济形势是不一样的。一位投资者认为明天“某种指定的股票行情上涨的可能性为 80%”这一事件也是不可重复的。因此主观概率的创立，使得在频率解释不能适用的情形也能讨论概率。从这个意义上说，主观概率至少是对概率的频率方法和古典方法的补充。

定义 7.2.1 主观概率 (subjective probability) 是人们根据经验对事件发生机会的个人信念。

例如，对一场足球赛胜负的打赌，对明天是否下雨的估计，对股票市场行情明天是升还是降的预测等都是采用主观概率方法。主观概率提出以来，使用的人越来越多，特别在社会、经济领域和决策分析中被较为广泛地使用。因为这些领域遇到的随机现象大多是不可重复的，无法用频率方法去确定概率。

2. 确定主观概率的方法

确定主观概率有下列一些方法：

(1) 对一些事件进行对比，获得相对似然性是确定主观概率最简单的办法。例如，某工厂已设计好一种新玩具，决策者要决定是否投产，需要评估新玩具畅销的概率。根据新玩具的特点和多年经验认为畅销 (A) 是不畅销 (\bar{A}) 的可能性的 2 倍，即 $P(A) = 2/3$, $P(\bar{A}) = 1/3$ ，由此决定投产。

(2) 利用专家意见来确定主观概率的方法是常用的. 例如, 有一项带有风险的生意, 欲估计成功的概率. 为此决策者采访这方面的专家, 向专家请教这生意成功的可能性有多大. 专家回答大约 0.6 (可以请教几位专家用其预测概率的平均值代替). 如果决策者对专家比较了解, 认为他的估计往往偏保守, 决策者可以修正这一估计, 将成功的概率修改为 0.7, 即 $P(A) = 0.7$, 这就是主观概率.

(3) 利用历史资料, 作一些对比修正, 确定主观概率. 例如, 某公司经营玩具, 现设计一种新式玩具将投放市场, 要估计未来市场销售情况. 经理查阅了本公司生产的 37 种玩具的销售记录, 得到销售状态如下: A_1 : 畅销; A_2 : 一般; A_3 : 滞销, 分别有 29, 6, 2 种, 于是得到销售状态的概率为

$$P(A_1) = \frac{29}{37} = 0.784, \quad P(A_2) = \frac{6}{37} = 0.162, \quad P(A_3) = \frac{2}{37} = 0.054.$$

考虑到新玩具不仅外形新颖而且能开发儿童智力, 认为它更畅销一些, 故对上述概率作了修正, 得到 $P(A_1) = 0.85, P(A_2) = 0.14, P(A_3) = 0.01$ 作为该产品畅销、一般和滞销的概率.

7.2.2 利用先验信息确定先验分布

在 Bayes 方法中关键的一步是如何确定先验分布, 当参数 θ 属于离散型随机变量, 即参数空间 Θ 是由有限或可列个点构成时, 可对 Θ 中每个点确定一个主观概率, 这就是前面所介绍的.

当参数 θ 是连续随机变量时, 即 θ 为实轴或其上的某个区间时, 构造一个先验密度就有些困难了. 当 θ 的先验信息足够多时, 下面的一些方法可以使用.

1. 直方图法

这个方法与一般直方图法类似, 可以通过下述步骤实现:

- (1) 当参数空间 θ 为实轴上的区间时, 先把 θ 分成一些小区间, 通常为等长的子区间;
- (2) 在每个小区间上决定主观概率或按历史数据算出频率;
- (3) 绘制直方图, 纵坐标为主观概率或 [频率/区间长];
- (4) 在直方图上画一条光滑曲线, 使下方与直方图面积相等. 此曲线即为先验密度 $\pi(\theta)$ (使曲线与横轴形成的曲边梯形的面积为 1).

例 7.2.1 云南某药店销售云南三七, 记录了 100 天的销售额, 每天销售最多 35kg, 最少 5kg, 数据见表 7.2.1. 要寻求每天销售量 θ 的概率分布.

解 利用直方图来确定 θ 的概率分布, 按下述步骤:

- (1) 把 θ 的区间 $(0, 35)$ 分成 7 个小区间, 每个小区间长为 5 个单位 (kg);

表 7.2.1 每天销售量统计表

销售量/kg	[0,5]	(5,10]	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]
天数	5	26	33	22	10	3	1
频率	0.05	0.26	0.33	0.22	0.10	0.03	0.01

- (2) 在每个小区间上依据历史数据确定频率 (表 7.2.1 已给出);
(3) 绘制频率直方图, 纵坐标为“频率/5”;
(4) 在直方图上画一条光滑曲线, 使下方与直方图面积相等. 此曲线即为先验密度 $\pi(\theta)$, 见图 7.2.1.

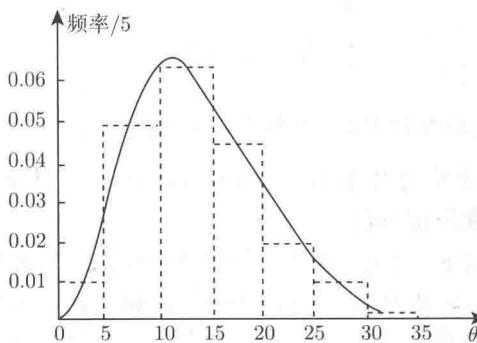


图 7.2.1 日销售量的直方图

利用此直方图可以计算有关的概率. 例如, $P(20 \leq \theta \leq 21) = 1 \times \pi(20.5) = 0.03$.

2. 相对似然法

此法大多用于 θ 为 $(-\infty, \infty)$ 的有限子区间的情形. 方法如下: 对 θ 中的各种点的直观“似然”进行比较, 再按确定了的值画图, 即可得到先验密度草图, 用下例来作说明.

例如, 设 $\theta = (0, 1)$, 从确定“最大可能”和“最小可能”的参数点的似然性入手. 设 $\theta = 3/4$ 为最大可能的点, $\theta = 0$ 为最小可能的点, 且 $\theta = 3/4$ 为 $\theta = 0$ 的似然性的 3 倍. 再确定 $\theta = 1/4$ 和 $\theta = 1/2$ 及 $\theta = 1$ 的相对似然性. 为简单计, 与 $\theta = 0$ 的可能性比较, $\theta = 1/2$ 和 $\theta = 1$ 的可能性 2 倍于 $\theta = 0$, 而 $\theta = 1/4$ 的可能性为 $\theta = 0$ 的可能性的 1.5 倍. 令基本点 $\theta = 0$ 的先验密度为 1, 由此画出 $\tilde{\pi}(\theta)$, 见图 7.2.2. 但 $\int_0^1 \tilde{\pi}(\theta) d\theta \neq 1$. 记 $\pi(\theta) = c\tilde{\pi}(\theta)$, 使 $\int_0^1 \pi(\theta) d\theta = c \int_0^1 \tilde{\pi}(\theta) d\theta = 1$, 则 $\pi(\theta)$ 即为 θ 的先验密度.

注 7.2.1 当 $\theta = (-\infty, \infty)$ 时此法会遇到较大困难. 上述两种确定先验密度的方法要求 θ 局限于 $(-\infty, \infty)$ 上的有限区间, 当 $\theta = (-\infty, \infty)$ 或其上无限区间

时便失效。下面介绍的方法更适合 θ 为无限区间的情形。

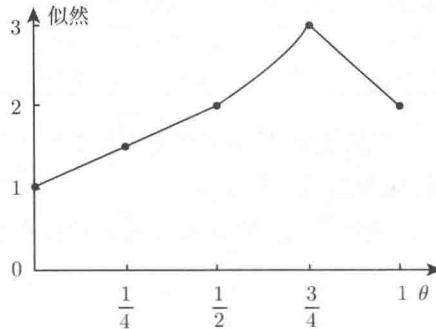


图 7.2.2 似然

3. 选定先验密度函数的形式，再估计超参数

先验分布中的参数称为超参数 (hyperparameter)。例如，先验分布 $\pi(\theta)$ 为 $N(\mu, \tau^2)$ ，则 μ 和 τ^2 称为超参数。

此法的主要思想如下：设先验分布的形式为 $\pi(\theta; \alpha, \beta)$ ，其中 α, β 为超参数。对超参数为 α, β 作出估计，得到 $\hat{\alpha}, \hat{\beta}$ ，使 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 和 $\pi(\theta; \alpha, \beta)$ 很接近，则 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 即为选定的先验密度函数。

这个方法的关键是 $\pi(\theta)$ 的形式的选定。若选择的不合适将导致失误。

例 7.2.2 在例 7.2.1 中设参数 θ 为销售量，选用正态分布 $N(\mu, \tau^2)$ 作为 θ 的先验分布 $\pi(\theta)$ ，试确定这一先验分布。

解 确定先验分布的问题就转化为估计超参数 μ 和 τ^2 的问题。这可用“每日销售量统计表”作出估计。若对表中 θ 的每个小区间用其中点代替，算得 μ 和 τ^2 的估计如下：

$$\hat{\mu} = 2.5 \times 0.05 + 7.5 \times 0.26 + \cdots + 32.5 \times 0.01 = 13.45,$$

$$\hat{\tau}^2 = (2.5 - \hat{\mu})^2 \times 0.05 + (7.5 - \hat{\mu})^2 \times 0.26 + \cdots$$

$$+ (32.5 - \hat{\mu})^2 \times 0.01 = 36.85,$$

故 $\theta \sim N(\hat{\mu}, \hat{\tau}^2) = N(13.45, 36.85)$ ，用先验分布可求下列概率，如

$$\begin{aligned} P(20 \leq \theta \leq 21) &= P\left(\frac{20 - 13.45}{\sqrt{36.85}} \leq \frac{\theta - 13.45}{\sqrt{36.85}} \leq \frac{21 - 13.45}{\sqrt{36.85}}\right) \\ &= \Phi(1.24) - \Phi(1.08) = 0.8925 - 0.8508 = 0.0417. \end{aligned}$$

在给定先验分布形式时，决定超参数的另一方法是从先验信息中获得几个分位数的估计值，再通过这些分位数的值确定超参数（即超参数的估计值），请看下例。

例 7.2.3 设参数 θ 的取值范围为 $(-\infty, \infty)$, 先验分布为正态分布, 若从先验信息得知: ① 先验的中位数为 0; ② 0.25 和 0.75 的分位数分别为 -1 和 +1, 试求此先验分布.

解 由于 $\theta \sim N(\mu, \tau^2)$, 所以确定先验分布的问题就转化为估计 μ 和 τ 的问题. 正态分布的中位数就是 μ , 故 $\mu = 0$. 由 0.75 的分位数为 1, 即 $0.75 = P(\theta < 1) = P(\theta/\tau < 1/\tau) = P(Z < 1/\tau)$, 其中 $Z = \theta/\tau \sim N(0, 1)$, 查标准正态分布表 $\Rightarrow 1/\tau = 0.675$, 即 $\tau = 1.481$, 故知 $\theta \sim N(\mu, \tau^2) = N(0, 1.481^2)$ 为 θ 的先验分布.

又若在本例中假定 θ 的先验不是正态分布, 而是 Cauchy 分布, 其余条件不变, 即 $\theta \sim \pi(\theta; \alpha, \beta) = \beta / \{\pi[\beta^2 + (\theta - \alpha)^2]\}$, $-\infty < \theta < \infty$, 确定先验分布的问题就转化为求 α, β 的估计.

由于 Cauchy 分布 $C(\alpha, \beta)$ 的均值和方差皆不存在, 但它关于 α 对称, 故有 $\alpha = 0$. 又由条件 0.25 分位数为 -1, 即

$$\int_{-\infty}^{-1} \frac{\beta}{\pi(\beta^2 + \theta^2)} d\theta = \frac{1}{\pi} \arctan\left(\frac{-1}{\beta}\right) + \frac{1}{2} = 0.25,$$

解出 $\beta = 1$, 故 $\theta \sim C(\alpha, \beta) = C(0, 1)$.

因此, 同样的先验信息有 2 个先验分布可供选择. 若 2 个先验分布差别不大, 可任选其一. 在本例中 $N(0, 1.481^2)$ 和 $C(0, 1)$ 密度函数形状上相似 (都关于 0 对称, 中间高两边低), 但 Cauchy 分布的尾部概率较大. 因此若 θ 的先验信息集中在中间, 则选择正态好些, 若先验信息较分散, 选择 Cauchy 分布更合适些.

7.2.3 无信息先验

Bayes 分析的一个重要特点就是在统计推断时要利用先验信息. 但常常会出现这样的情况: 没有先验信息或者只有极少的先验信息可利用, 但仍想用 Bayes 方法. 此时所需要的是一种无信息先验 (noninformative prior), 即对参数空间 Θ 中的任何一点 θ 没有偏爱、都是同等无知的先验信息. 这就引出了无信息先验分布的概念.

1. 均匀分布与广义先验分布

(1) 若 Θ 为有限集, 即 θ 只可能取有限个值, 如 $\theta = \theta_i$, $i = 1, 2, \dots, n$, 无信息先验给 Θ 中的每个元素以概率 $1/n$, 即 $P(\theta = \theta_i) = 1/n$, $i = 1, 2, \dots, n$.

(2) 若 Θ 为 R_1 上的有限区间 $[a, b]$, 则取无信息先验为区间 $[a, b]$ 上的均匀分布 $U(a, b)$ (有时也记为 $R(a, b)$).

(3) 问题是若参数空间 Θ 无界, 无信息先验如何选取? . 例如, 样本分布为 $N(\theta, \sigma^2)$, σ^2 已知, 此时 $\Theta = (-\infty, \infty)$. 若无信息先验取为 $\pi(\theta) \equiv 1$, 则 $\pi(\theta)$ 不

是通常的密度, 因为 $\int_{-\infty}^{\infty} \pi(\theta) d\theta = \infty$. 这就引出广义先验分布的概念.

定义 7.2.2 设随机变量 $X \sim f(x|\theta)$, $\theta \in \Theta$, 若 θ 的先验密度 $\pi(\theta)$ 满足下列条件:

$$(1) \pi(\theta) \geq 0 \text{ 且 } \int_{\Theta} \pi(\theta) d\theta = \infty;$$

(2) 后验密度 $\pi(\theta|x)$ 是正常的密度函数,

则称 $\pi(\theta)$ 为 θ 的广义先验密度 (improper prior density).

注 7.2.2 由定义可见广义先验密度 $\pi(\theta)$ 乘以任一给定的常数仍是一个广义先验密度.

例 7.2.4 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 $N(\theta, 1)$ 总体中抽取的随机样本, 设 θ 的先验密度 $\pi(\theta) \equiv 1$, 求 θ 的后验密度.

解 由式 (7.1.1) 可知

$$\begin{aligned}\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(\mathbf{x}|\theta)\pi(\theta)d\theta} = \frac{\exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2\right\}}{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2\right\} d\theta} \\ &= \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\theta - \bar{x})^2\right\}.\end{aligned}$$

这是正态分布 $N(\bar{x}, 1/n)$ 的密度函数, 后验分布 $\pi(\theta|\mathbf{x})$ 仍为正常的密度函数, 故 $\pi(\theta) \equiv 1$ 为广义先验密度, 它也是一种无信息先验.

对一般常见的概率分布, 如何求其参数的无信息先验分布? 下面将对位置参数和刻度参数分别加以介绍.

2. 位置参数的无信息先验

设总体分布的密度函数有形式 $f(x-\theta)$, $-\infty < \theta < \infty$, θ 为位置参数 (location parameter). 对 X 作平移变换得到 $Y = X + c$, 同时对 θ 也作平移变换得到 $\eta = \theta + c$. 显然 Y 的密度函数有形式 $f(y-\eta)$, η 仍为位置参数. 所以 (X, θ) 与 (Y, η) 的统计问题结构相同. 因此主张它们有相同的无信息先验是合理的. 理解这一点的另一方法: X 和 Y 的测量原点不同, 由于测量原点的选择是非常任意的, 所以无信息先验应当与这种选择无关. 如果无信息先验不依赖于原点的选择, 则它在等长区间内的先验概率应当一样. 换言之, 先验密度应当恒等于 1, 即取 θ 的无信息先验密度

$$\pi(\theta) \equiv 1.$$

它是一个广义先验密度.

这表明当 θ 为位置参数时, 其无信息先验密度取为常数或者 1.

例 7.2.5 (续例 7.2.4) 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, 其中 σ^2 已知. θ 无任何先验信息可用, 求 θ 的后验分布.

解 显见, 此时 $\bar{X} = \sum_{i=1}^n X_i/n$ 为充分统计量且 $\bar{X} \sim N(\theta, \sigma^2/n)$, 即

$$p(\bar{x}|\theta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right\}.$$

由于 θ 无任何先验信息可用, 此时如取无信息先验 $\pi(\theta) \equiv 1$, 则由例 7.2.4 可知后验密度是正态分布 $N(\bar{x}, \sigma^2/n)$ 的密度. 例如, 取 θ 的 Bayes 估计为后验均值, 则 Bayes 估计为 $\hat{\theta}_B = E(\theta|x) = \bar{x}$. 这个结果与经典统计中常用估计量一致.

这种现象被 Bayes 学派解释为, 经典统计学中一些成功的估计量可以看作使用合理的无信息先验的结果. 无信息先验的开发和使用是 Bayes 统计中最成功的结果之一.

3. 刻度参数的无信息先验

设总体分布的密度函数有形式 $\sigma^{-1}\varphi(x/\sigma)$, $\sigma > 0$ 为刻度参数 (scale parameter). 对 X 作变换 $Y = cX$, 同时对 σ 也作相应的变换 $\eta = c\sigma$. 不难算出 Y 的密度仍为 $\eta^{-1}\varphi(y/\eta)$. 可见 (X, σ) 和 (Y, η) 统计问题的结构相同, 故主张 σ 的无信息先验与 η 的无信息先验相同是合理的. 理解这一点的另一方法: X 和 Y 的度量单位不同, 先验分布应当不依赖于度量单位的选择, 则对任何 $a, b, 0 < a < b, c > 0$, σ 落在 $[a, b]$ 内的先验概率, 应当等于 η 落在 $[ca, cb]$ 内的先验概率. 不难看出, 这个事实只有在 σ 的先验密度为

$$\pi(\sigma) = \frac{1}{\sigma}, \quad \text{当 } \sigma > 0$$

时才能成立.

例 7.2.6 设总体 X 为指数分布, 其密度为

$$f(x|\lambda) = \lambda^{-1} \exp\{-x/\lambda\}, \quad x > 0,$$

其中 $\lambda > 0$ 为刻度参数. 令 $\mathbf{X} = (X_1, \dots, X_n)$ 是从上述分布中抽取的简单样本, λ 的先验密度为无信息先验, 求其后验密度.

解 记 $T = \sum_{i=1}^n x_i$, 由式 (7.1.1) 可知 λ 的后验密度为

$$\pi(\lambda|\mathbf{x}) = \frac{\prod_{i=1}^n f(x_i|\lambda)\pi(\lambda)}{\int_0^\infty \prod_{i=1}^n f(x_i|\lambda)\pi(\lambda)d\lambda} = \frac{\lambda^{-(n+1)} \exp\left\{-\frac{1}{\lambda}T\right\}}{\int_0^\infty \lambda^{-(n+1)} \exp\left\{-\frac{1}{\lambda}T\right\}d\lambda}$$

$$= \frac{T^n}{\Gamma(n)} \lambda^{-(n+1)} \exp \left\{ -\frac{1}{\lambda} T \right\}.$$

若取其 Bayes 估计为后验均值, 则 Bayes 估计为

$$\hat{\lambda}_B = E(\lambda|\mathbf{x}) = \frac{1}{n-1} \sum_{i=1}^n x_i,$$

其方差为 $\left(\sum_{i=1}^n x_i \right)^2 / [(n-1)^2(n-2)]$.

*4. 一般情形下的无信息先验

对非位置参数族和刻度参数族的无信息先验如何求, 广泛采用的是 Jeffreys 无信息先验, 由于推导涉及变换群和 Harr 测度, 这里只给出结果, 不推导结果是如何得来的.

假定样本分布族 $\{f(x|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ 满足 Cramer-Rao (C-R) 正则条件, 这里 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ 为 p 维参数向量. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从自总体 $f(x|\boldsymbol{\theta})$ 中抽取的随机样本. 在对 $\boldsymbol{\theta}$ 无先验信息可用时, Jeffreys 主张用 Fisher 信息阵行列式的平方根作为 $\boldsymbol{\theta}$ 的无信息先验, 这样的无信息先验称为 Jeffreys 无信息先验. 其求解步骤如下.

(1) 写出样本的对数似然函数

$$l(\boldsymbol{\theta}|\mathbf{x}) = \ln \left[\prod_{i=1}^n f(x_i|\boldsymbol{\theta}) \right] = \sum_{i=1}^n \ln f(x_i|\boldsymbol{\theta}).$$

(2) 求样本的信息阵

$$\mathbf{I}(\boldsymbol{\theta}) = (I_{ij}(\boldsymbol{\theta}))_{p \times p}, \quad I_{ij}(\boldsymbol{\theta}) = E_{\mathbf{X}|\boldsymbol{\theta}} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\}, \quad i, j = 1, \dots, p.$$

特别对 $p = 1$, 即 θ 为单参数的情形

$$I(\theta) = E_{\mathbf{X}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta^2} \right\},$$

(3) $\boldsymbol{\theta}$ 的无信息先验的密度为 $\pi(\boldsymbol{\theta}) = [\det \mathbf{I}(\boldsymbol{\theta})]^{1/2}$, 其中 $\det \mathbf{I}(\boldsymbol{\theta})$ 表示 p 阶方阵 $\mathbf{I}(\boldsymbol{\theta})$ 的行列式. 特别 $p = 1$, 即单参数场合 $\pi(\theta) = [I(\theta)]^{1/2}$.

例 7.2.7 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的简单样本, 记 $\boldsymbol{\theta} = (\mu, \sigma)$, 求 (μ, σ) 的联合无信息先验.

解 给定 \mathbf{X} 时, θ 的对数似然函数是

$$l(\boldsymbol{\theta}|\mathbf{x}) = -\frac{n}{2} \log 2\pi - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

记 $\mathbf{I}(\theta) = (I_{ij}(\theta))_{p \times p}$, 则有

$$\begin{aligned} I_{11}(\theta) &= E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \mu^2} \right\} = \frac{n}{\sigma^2}, \\ I_{22}(\theta) &= E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \sigma^2} \right\} = -\frac{n}{\sigma^2} + \frac{3}{\sigma^4} E \left\{ \sum_{i=1}^n (X_i - \mu)^2 \right\} = \frac{2n}{\sigma^2}, \\ I_{12}(\theta) = I_{21}(\theta) &= E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \mu \partial \sigma} \right\} = E \left\{ \frac{2}{\sigma^3} \sum_{i=1}^n (X_i - \mu) \right\} = 0. \end{aligned}$$

故有

$$\mathbf{I}(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{pmatrix}, \quad [\det \mathbf{I}(\theta)]^{1/2} = \frac{\sqrt{2}n}{\sigma^2}.$$

所以, (μ, σ) 的 Jeffreys 先验 (由于它是广义先验, 可以丢弃常数因子) 为

$$\pi(\mu, \sigma) = \frac{1}{\sigma^2},$$

即 (μ, σ) 的联合无信息先验为 $1/\sigma^2$. 它的三个特例为

- (1) 当 σ 已知时, $I(\mu) = E \left\{ -\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \mu^2} \right\} = n/\sigma^2$, 故 $\pi_1(\mu) \equiv 1$;
- (2) 当 μ 已知时, $I(\sigma) = E \left\{ -\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \sigma^2} \right\} = 2n/\sigma^2$, 故 $\pi_2(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$;
- (3) 当 μ 和 σ 独立时, $\pi(\mu, \sigma) = \pi_1(\mu)\pi_2(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.

由此可见, 当 μ 和 σ 的无信息先验不独立时, 它们的联合无信息先验为 $1/\sigma^2$; 而当 μ 和 σ 的无信息先验独立时, 它们的联合无信息先验为 $1/\sigma$. Jeffreys 最终推荐用 $\pi(\mu, \sigma) = 1/\sigma$ 为 μ 和 σ 的联合无信息先验.

例 7.2.8 设 θ 为 Bernoulli 试验中成功概率, 则在 n 次独立的 Bernoulli 试验中, 成功次数 $X \sim b(n, \theta)$, 即

$$P(X = x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \dots, n,$$

其对数似然函数为 $l(\theta|x) = \ln \binom{n}{x} + x \ln \theta + (n-x) \ln(1-\theta)$, 故有

$$I(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \theta^2} \right\} = E_{\mathbf{x}|\theta} \left\{ \frac{X}{\theta^2} + \frac{n-X}{(1-\theta)^2} \right\} = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}.$$

故取 $\pi(\theta) \propto I(\theta)^{1/2} = \theta^{-1/2}(1-\theta)^{-1/2}$, $\theta \in (0, 1)$, 添加正则化因子得到先验密度 $\pi(\theta)$, 它是一个 Beta 密度 $\text{Be}(1/2, 1/2)$.

注 7.2.3 一般说来, 无信息先验不唯一, 它们对 Bayes 推断影响都很小, 很少对结果产生较大的影响, 所以任何无信息先验都可以接受. 当今无论在统计理论

还是应用研究中无信息先验采用越来越多, 就连经典统计学者也认为无信息先验是客观的, 是可以接受的. 这是近几十年中 Bayes 学派研究中最成功的部分.

7.2.4 共轭先验分布

1. 共轭先验分布的概念

另外一种选择先验的方法是从理论角度出发的, 在已知样本分布的情形下, 为了理论上的需要常常选参数的先验分布为共轭先验分布, 其定义如下.

定义 7.2.3 设 θ 是总体分布中的参数, 样本 X 的分布为 $f(x|\theta)$, 如果对任取的先验分布 $\pi(\theta) \in \mathcal{F}$ 及样本 x , 后验分布 $\pi(\theta|x)$ 仍属于 \mathcal{F} , 则称 \mathcal{F} 是一个共轭先验分布族 (conjugate prior distribution family).

下面给出计算共轭先验分布的一个例子.

例 7.2.9 设 $X \sim b(n, \theta)$.

(1) 设 $\theta \sim U(0, 1)$, 即 $(0, 1)$ 上的均匀分布, 证明 θ 的后验分布为 Beta 分布;

(2) 若取 θ 的先验分布为 Beta 分布 $Be(a, b)$, 证明 θ 的后验分布仍为 Beta 分布, 即样本分布如果为二项分布, 则共轭先验分布为 Beta 分布.

证 (1) $U(0,1)$ 是 $Be(1, 1)$ 分布, $X \sim b(n, \theta)$, 其概率分布为

$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x},$$

而 θ 的先验分布为 $\pi(\theta) = 1$, 当 $\theta \in (0, 1)$, 故有

$$\pi(\theta|x) = \frac{\theta^x (1-\theta)^{n-x}}{\int_0^1 \theta^x (1-\theta)^{n-x} d\theta}. \quad (7.2.1)$$

计算积分得到

$$\int_0^1 \theta^x (1-\theta)^{n-x} d\theta = \frac{\Gamma(n-x+1)\Gamma(x+1)}{\Gamma(n+2)},$$

将上式结果代入式 (7.2.1), 得到后验密度

$$\pi(\theta|x) = \frac{\Gamma(n+2)}{\Gamma(n-x+1)\Gamma(x+1)} \theta^{(x+1)-1} (1-\theta)^{(n-x+1)-1},$$

即 θ 的后验分布是 Beta 分布 $Be(x+1, n-x+1)$.

(2) 又若 $\theta \sim Be(a, b)$, 则

$$\pi(\theta|x) = \frac{\theta^{x+a-1} (1-\theta)^{n-x+b-1}}{\int_0^1 \theta^x (1-\theta)^{n-x} \theta^{a-1} (1-\theta)^{b-1} d\theta}. \quad (7.2.2)$$

计算积分得到

$$\int_0^1 \theta^{x+a-1} (1-\theta)^{n-x+b-1} d\theta = \frac{\Gamma(x+a)\Gamma(n-x+b)}{\Gamma(n+a+b)}.$$

将上式结果代入式 (7.2.2), 得到后验密度

$$\pi(\theta|x) = \frac{\Gamma(n+a+b)}{\Gamma(x+a)\Gamma(n-x+b)} \theta^{(x+a)-1} (1-\theta)^{(n-x+b)-1},$$

即 θ 的后验分布是 Beta 分布 $\text{Be}(x+a, n-x+b)$. 因此, 样本分布若为二项分布, 其参数 θ 的共轭先验分布族为 Beta 分布族.

由此例可见计算后验分布时, 计算边缘分布时需要算积分. 下列的方法说明可以简化后验分布的计算, 省略计算边缘分布这一步骤.

2. 后验分布的计算

后验密度的计算公式由式 (7.1.1) 给出, 即

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}.$$

此处 $f(x|\theta)$ 是样本的密度函数 (也称为似然函数, 可以用 $l(\theta|x)$ 代替 $f(x|\theta)$), $\pi(\theta)$ 是 θ 的先验密度, $m(x)$ 是 X 的边缘密度. 由于 $m(x)$ 与 θ 无关, 故可将 $1/m(x)$ 看成与 θ 无关的常数, 因此有

$$\pi(\theta|x) \propto \frac{f(x|\theta)\pi(\theta)}{m(x)} \propto f(x|\theta)\pi(\theta), \quad (7.2.3)$$

其中符号 “ \propto ” 表示“正比于”, 即上式的左边和右边只相差一个正的常数, 此常数与 θ 无关, 但可以与 x 有关. 式 (7.2.3) 右端不是正常的密度函数, 但它正比于后验密度 $\pi(\theta|x)$ 的“核”(密度函数中仅与参数 θ 有关的因子称为它的核). 在共轭先验分布的场合, 很容易看出: 只要对后验密度的核, 添加一个正则化常数因子就可以得到后验密度.

因此, 对共轭先验分布情形, 求后验密度可按下列步骤.

(1) 写出似然函数 (即样本密度函数) $f(\theta|x)$ 的核, 即 $f(\theta|x)$ 中仅与参数 θ 有关的因子. 再写出先验密度 $\pi(\theta)$ 的核, 即 $\pi(\theta)$ 中仅与参数 θ 有关的因子.

(2) 类似式 (7.2.3), 写出后验密度的核, 即

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \{f(\theta|x) \text{ 的核}\} \cdot \{\pi(\theta) \text{ 的核}\}, \quad (7.2.4)$$

即“后验密度的核”是“似然函数的核与先验密度的核的乘积”.

(3) 将式(7.2.4)的右边添加一个正则化常数因子(可以与 x 有关),即可得到后验密度.

注 7.2.4 上述计算后验分布的简化方法,一般说来只对先验分布为共轭先验和无信息先验的情形有效.当先验分布为其他先验的情形,获得后验分布的核之后,如果不能判断出后验分布的类型,就不知道如何添加正则化常数因子,将“后验密度的核”变成“后验密度”.此时只有老老实实按式(7.1.1)去计算后验密度.

现在用上面介绍的方法来解例 7.2.9. 设 $X \sim b(n, \theta)$,若取 θ 的先验分布为 $Be(a, b)$,求 θ 的后验分布.

解 似然函数(即样本密度)的核是 $\theta^x(1-\theta)^{n-x}$,而先验密度的核是 $\theta^{a-1}(1-\theta)^{b-1}$.因此由式(7.2.4),有

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \theta^{x+a-1}(1-\theta)^{n-x+b-1}.$$

易见上式的右边是“ $Be(x+a, n-x+b)$ 密度的核”.因此,添加正则化因子得到后验密度

$$\pi(\theta|x) = \frac{\Gamma(n+a+b)}{\Gamma(x+a)\Gamma(n-x+b)}\theta^{(x+a)-1}(1-\theta)^{(n-x+b)-1}.$$

由此例可见,上面介绍的方法简化了后验分布的计算.下面再看计算后验分布的几个例子.

例 7.2.10 设 $X \sim N(\theta, \sigma^2)$, σ^2 已知而 θ 未知.令 θ 的先验分布 $\pi(\theta)$ 是 $N(\mu, \tau^2)$,其中 μ 和 τ^2 已知,求 θ 的后验分布 $\pi(\theta|x)$.

解 给定 θ 时 X 的条件分布记为 $f(x|\theta)$,令

$$A = \frac{1}{\sigma^2} + \frac{1}{\tau^2}, \quad B = \frac{x}{\sigma^2} + \frac{\mu}{\tau^2}, \quad C = \frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2}, \quad (7.2.5)$$

则有

$$\begin{aligned} \pi(\theta|x) &\propto f(x|\theta)\pi(\theta) \propto \exp\left\{-\frac{1}{2}\left[\frac{(x-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[A\theta^2 - 2\theta B + C\right]\right\} \\ &= \exp\left\{-\frac{A}{2}\left(\theta - \frac{B}{A}\right)^2 - \frac{1}{2}\left(C - \frac{B^2}{A}\right)\right\} \\ &= \exp\left\{-\frac{1}{2\eta^2}(\theta - \mu(x))^2\right\} \cdot \exp\left\{-\frac{(x-\mu)^2}{2(\sigma^2 + \tau^2)}\right\} \\ &\propto \exp\left\{-\frac{1}{2\eta^2}(\theta - \mu(x))^2\right\} \end{aligned} \quad (7.2.6)$$

其中

$$\mu(x) = \frac{B}{A} = \frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \quad \eta^2 = \frac{1}{A} = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}, \quad C - \frac{B^2}{A} = \frac{(x-\mu)^2}{\sigma^2 + \tau^2}.$$

易见, 式 (7.2.6) 右边是 $N(\mu(x), \eta^2)$ 密度的核, 添加正则化因子, 得到后验密度

$$\pi(\theta|x) = \frac{1}{\sqrt{2\pi}\eta} \exp \left\{ -\frac{1}{2\eta^2} [\theta - \mu(x)]^2 \right\}.$$

将式 (7.2.6) 倒数第二式对 θ 积分, 添加正则化常数得到 X 的边缘密度

$$f(x) = \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)}} \exp \left\{ -\frac{(\bar{x} - \mu)^2}{2(\sigma^2 + \tau^2)} \right\}$$

若进一步假定 X_1, \dots, X_n i.i.d. $\sim N(\theta, \sigma^2)$, σ^2 已知, $\theta \sim N(\mu, \tau^2)$, 求 θ 的后验密度.

由于样本均值 \bar{X} 为 θ 的充分统计量, 且给定 θ 时 \bar{X} 的分布是 $N(\theta, \sigma^2/n)$, 故将上述结果中的 x 用 \bar{x} 代替, σ^2 用 σ^2/n 代替得到 θ 的后验分布

$$\pi(\theta|\bar{x}) \sim N(\mu_n(\bar{x}), \eta_n^2),$$

其中

$$\mu_n(\bar{x}) = \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \mu + \frac{\tau^2}{\sigma^2/n + \tau^2} \bar{x}, \quad \eta_n^2 = \frac{\tau^2 \cdot \sigma^2/n}{\sigma^2/n + \tau^2} = \frac{\sigma^2 \tau^2}{n \tau^2 + \sigma^2}.$$

此例表明当样本分布为方差已知的正态分布时, 则均值参数 θ 的共轭先验分布族是正态分布族.

例 7.2.11 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 Poisson 分布 $P(\lambda)$ 中抽取的简单样本, λ 的先验分布为 Gamma 分布 $\Gamma(b, a)$. 证明给定 $\mathbf{X} = \mathbf{x}$ 时, λ 的后验分布仍为 Gamma 分布.

证 λ 的似然函数为

$$l(\lambda|\mathbf{x}) = \frac{e^{-n\lambda} \cdot \lambda^{n\bar{x}}}{x_1! \cdots x_n!} \propto \lambda^{n\bar{x}} e^{-n\lambda},$$

其中 $\bar{x} = \sum_{i=1}^n x_i/n$. 易见 λ 的先验分布是

$$\pi(\lambda) = \frac{a^b}{\Gamma(b)} \lambda^{b-1} e^{-a\lambda} \propto \lambda^{b-1} e^{-a\lambda},$$

则由式 (7.2.4) 可知

$$\pi(\lambda|\mathbf{x}) \propto l(\lambda|\mathbf{x}) \pi(\lambda) \propto \lambda^{n\bar{x}+b-1} e^{-(n+a)\lambda},$$

上式右边是 $\Gamma(n\bar{x} + b, n + a)$ 密度的核. 添加正则化常数因子, 得到后验密度

$$\pi(\lambda|\mathbf{x}) = \frac{(n+a)^{n\bar{x}+b}}{\Gamma(n\bar{x}+b)} \lambda^{n\bar{x}+b-1} e^{-(n+a)\lambda}.$$

此例表明当样本分布为参数 λ 的 Poisson 分布时, 则 λ 的共轭先验是 Gamma 分布族.

例 7.2.12 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 Gamma 分布 $\Gamma(r, \lambda)$ 中抽取的简单样本, 其中 r 已知. 假定 λ 的先验分布为 Gamma 分布 $\Gamma(b, a)$, 证明给定 \mathbf{X} 时, λ 的后验分布仍为 Gamma 分布.

证 λ 的似然函数为

$$l(\lambda|\mathbf{x}) = \frac{\lambda^{nr}}{(\Gamma(r))^n} \prod_{i=1}^n x_i^{r-1} \cdot \exp\{-\lambda n\bar{x}\} \propto \lambda^{nr} \exp\{-\lambda n\bar{x}\},$$

其中 $\bar{x} = \sum_{i=1}^n x_i/n$. 而 λ 的先验分布是

$$\pi(\lambda) = \frac{a^b}{\Gamma(b)} \lambda^{b-1} \exp\{-a\lambda\} \propto \lambda^{b-1} \exp\{-a\lambda\},$$

则由式 (7.2.4) 可知

$$\pi(\lambda|\mathbf{x}) \propto l(\lambda|\mathbf{x}) \pi(\lambda) \propto \lambda^{nr+b-1} \exp\{-(a+n\bar{x})\lambda\}.$$

上式右边是 $\Gamma(nr+b, a+n\bar{x})$ 密度的核. 添加正则化常数因子, 得到后验密度

$$\pi(\lambda|\mathbf{x}) = \frac{(a+n\bar{x})^{nr+b}}{\Gamma(nr+b)} \lambda^{nr+b-1} \exp\{-(a+n\bar{x})\lambda\}.$$

此例表明当样本分布为 Gamma 分布 $\Gamma(r, \lambda)$, 其中 r 已知时, 则 λ 的共轭先验是 Gamma 分布族. 由于指数分布是 Gamma 分布的特例, 所以当样本分布为指数分布 $\text{Exp}(\lambda)$ 时, λ 的共轭先验也是 Gamma 分布族.

定义 7.2.4 (逆 Gamma 分布) 若样本 X 的密度函数为

$$f(x|\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\lambda}{x}}, \quad x > 0,$$

则称随机变量 X 的分布为逆 Gamma 分布, 记为 $X \sim \Gamma^{-1}(\alpha, \lambda)$.

例 7.2.13 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态分布 $N(\mu, \sigma^2)$ 中抽取的简单样本, 其中 μ 已知, 求 σ^2 的共轭先验分布.

解 σ^2 的似然函数为

$$\begin{aligned} l(\sigma^2 | \mathbf{x}) &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \\ &\propto (\sigma^2)^{-n/2} \exp \left\{ -A/\sigma^2 \right\}, \end{aligned}$$

其中 $A = \sum_{i=1}^n (x_i - \mu)^2 / 2$, 可见似然函数 $l(\sigma^2 | \mathbf{x})$ 作为 $\theta = \sigma^2$ 的函数为逆 Gamma 分布, 故取先验分布为逆 Gamma 分布 $\Gamma^{-1}(\alpha, \lambda)$, 即

$$\pi(\sigma^2) = \frac{\lambda^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} e^{-\frac{\lambda}{\sigma^2}} \propto (\sigma^2)^{-(\alpha+1)} e^{-\frac{\lambda}{\sigma^2}}, \quad \sigma^2 > 0,$$

则由式 (7.2.4) 可知

$$\pi(\sigma^2 | \mathbf{x}) \propto l(\sigma^2 | \mathbf{x}) \pi(\sigma^2) \propto (\sigma^2)^{-(n/2+\alpha+1)} e^{-\frac{\lambda+A}{\sigma^2}}.$$

易见上式右边是逆 Gamma 分布 $\Gamma^{-1}(n/2 + \alpha, \lambda + A)$ 密度的核. 添加正则化常数因子, 得到后验密度

$$\pi(\sigma^2 | \mathbf{x}) = \frac{(\lambda + A)^{n/2+\alpha}}{\Gamma(n/2 + \alpha)} (\sigma^2)^{-(n/2+\alpha+1)} e^{-\frac{\lambda+A}{\sigma^2}}.$$

因此当样本分布为正态分布, 均值参数已知, 则 σ^2 的共轭先验是逆 Gamma 分布族.

注 7.2.5 共轭先验分布具有下列优点: ① 计算方便; ② 后验分布的某些参数常可以得到很好的解释. 例如, 例 7.2.10 中, 后验分布为 $N(\mu_n(\bar{x}), \eta_n^2)$, 其中

$$\mu_n(\bar{x}) = \frac{\tau^2}{\sigma^2/n + \tau^2} \bar{x} + \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \mu = r\bar{x} + (1-r)\mu$$

是样本均值 \bar{x} 和先验均值 μ 的加权平均. 若 σ^2/n 很小, 即样本信息量很大, 相对的先验信息很少, 则后验均值主要由 \bar{x} 决定, 反之若 σ^2/n 很大 (即样本信息很少), 则后验均值主要由先验分布的均值 μ 来决定. 由此可见后验均值是样本均值和先验均值的一个折中.

7.3 Bayes 统计推断

统计模型中参数的 Bayes 分析有两种方法: 一种方法是从后验分布出发考虑模型中参数的统计推断问题, 此时不考虑损失函数; 另一种方法是用统计决策的方法来考虑, 需要考虑损失函数. 本节和 7.4 节将分别讨论这两种方法, 本节讨论前者.

7.3.1 点估计

1. Bayes 点估计方法

有了后验分布后, 可从后验分布出发, 按经典方法(用后验分布代替通常的样本分布)求未知参数 θ 的点估计, 如极大似然估计(MLE)、后验中位数估计和后验期望估计等. 下面首先给出三种估计的定义.

定义 7.3.1 使后验密度 $\pi(\theta|x)$ 达到最大时 θ 的值称为后验分布的众数(mode), 用后验分布的众数作为 θ 的估计称为后验众数估计, 也称为广义极大似然(广义MLE)估计, 记为 $\hat{\theta}_{\text{MD}}$; 用后验分布的中位数(median)作为 θ 的估计, 称为 θ 的后验中位数估计, 记为 $\hat{\theta}_{\text{ME}}$; 用后验分布的期望(expectation)作为 θ 的估计称为后验期望估计(或后验均值估计), 记为 $\hat{\theta}_E$. 三个估计都称为 Bayes 估计, 在不会引起混淆的情况下上述三个估计皆用 $\hat{\theta}_B$ 来记.

注 7.3.1 一般场合下这三种估计是不同的, 但当后验密度为对称时, θ 的上述三种 Bayes 估计重合. 使用时可根据需要选用其中的一种. 一般来说, 当先验分布为共轭先验时, 求上述三种估计比较容易.

例 7.3.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, 其中 σ^2 已知. 取 θ 的先验为共轭先验分布 $N(\mu, \tau^2)$, 求 θ 的 Bayes 估计.

解 由例 7.2.10 可知 $\pi(\theta|\mathbf{x})$ 是正态分布 $N(\mu_n(\mathbf{x}), \eta_n^2)$, 其中

$$\mu_n(\mathbf{x}) = \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \mu + \frac{\tau^2}{\sigma^2/n + \tau^2} \bar{x}, \quad \eta_n^2 = \frac{\sigma^2 \tau^2}{n \tau^2 + \sigma^2}.$$

由于后验分布 $N(\mu_n(\mathbf{x}), \eta_n^2)$ 为对称分布, 故后验众数估计, 后验中位数估计和后验均值皆相同, 故 θ 的 Bayes 估计为

$$\hat{\theta}_B = \mu_n(\mathbf{x}) = \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \mu + \frac{\tau^2}{\sigma^2/n + \tau^2} \bar{x}.$$

如例 7.2.10 所述, $\hat{\theta}_B$ 是 \bar{x} 和 μ 加权平均, 是对样本信息和先验信息的综合.

例 7.3.2 为估计产品不合格品率 θ , 今从一批产品中随机抽取 n 件检查, 检查结果为 X_1, \dots, X_n , 其中 $X_i = 1$ 表示抽出的第 i 件不合格, $X_i = 0$ 表示抽出的第 i 件合格, 不合格品数 $X = \sum_{i=1}^n X_i$. 若取 θ 的先验分布为共轭先验 $\text{Be}(a, b)$, 求 θ 的后验众数估计和后验均值估计, 并比较这两个估计.

解 易见 $X \sim b(n, \theta)$. 则由例 7.2.9 可知 θ 的后验密度为

$$\pi(\theta|\mathbf{x}) = \frac{\Gamma(n+a+b)}{\Gamma(x+a)\Gamma(b-x+n)} \theta^{x+a-1} (1-\theta)^{n-x+b-1},$$

即 $\pi(\theta|\mathbf{x}) \sim \text{Be}(x+a, n-x+b)$. 易知 θ 的后验众数估计为

$$\hat{\theta}_{\text{MD}} = \frac{(x+a)-1}{(X+a)+(n-X+b)-2} = \frac{x+a-1}{n+a+b-2};$$

后验均值估计为

$$\hat{\theta}_E = \frac{x + a}{n + a + b}.$$

特别当先验分布中取 $a = 1, b = 1$, 则 $\pi(\theta)$ 是均匀分布 $U(0, 1)$, 此时

$$\hat{\theta}_{MD} = \frac{x}{n}, \quad \hat{\theta}_E = \frac{x + 1}{n + 2}.$$

对这两个估计作如下说明: ① θ 的后验众数估计 (即广义 MLE) 就是经典统计学中的 MLE, 即不合格率 θ 的 MLE 就是取先验分布为无信息先验 $U(0, 1)$ 下的 Bayes 估计, 这种现象以后还会看到. Bayes 学派对这种现象的看法是: 任何使用经典统计方法的人都自觉或不自觉地使用 Bayes 方法, 即经典统计方法下的许多估计量是特殊先验分布下的 Bayes 估计. ② θ 的后验期望估计要比后验众数更合适一些, 表 7.3.1 列出了 4 个特殊试验结果.

表 7.3.1 不合格品率 θ 的两种 Bayes 估计的比较表

试验号	样本量 n	不合格品数 x	$\hat{\theta}_{MD} = x/n$	$\hat{\theta}_E = (x + 1)/(n + 2)$
1	3	0	0	0.200
2	10	0	0	0.083
3	3	3	1	0.800
4	10	10	1	0.917

1 号试验与 2 号试验各抽 3 个与 10 个, 其中没有一个不合格, 这两件事在人们心目中留下的印象是不同的, 后者的质量要比前者更信得过. 但 $\hat{\theta}_{MD}$ 皆为 0, 显示不出二者的差别, 而 $\hat{\theta}_E$ 可显示出二者的差别. 对 3 号和 4 号试验, 也有类似问题. 由此例看到, 当 $x = 0$ 或 $x = n$ 时, $\hat{\theta}_{MD} = 0$ 或 1, 这种估计未免太极端一些, 而 $\hat{\theta}_E$ 可以看作是对它的修正, 更合理些, 所以在实际中人们更愿意选用后验均值估计作为 Bayes 估计. 由于 $\hat{\theta}_{MD}$ 与经典估计相同, 故 Bayes 估计 $\hat{\theta}_E$ 显示出相对于经典估计的优点.

由此可见, 作为 θ 的估计, 后验均值估计常优于后验众数估计.

例 7.3.3 设随机变量 $X \sim f(x|\theta) = e^{-(x-\theta)} I_{[\theta, \infty)}(x)$, 此处 $-\infty < \theta < +\infty$ 为位置参数, 取 θ 的先验分布为 Cauchy 分布, 即先验密度 $\pi(\theta) = 1/[\pi(1+\theta^2)]$, $-\infty < \theta < +\infty$, 求 θ 的后验众数估计.

解 θ 的后验密度为

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} = \frac{e^{-(x-\theta)}}{m(x)\pi(1+\theta^2)}, \quad x \geq \theta.$$

要找使 $\pi(\theta|x)$ 最大化的 $\hat{\theta}$, 对后验密度关于 θ 求导数, 可得

$$\frac{d\pi(\theta|x)}{d\theta} = \frac{e^{-x}}{\pi m(x)} \left[\frac{e^\theta}{1+\theta^2} - \frac{2\theta e^\theta}{(1+\theta^2)^2} \right] = \frac{e^{-x}}{\pi m(x)} \cdot \frac{e^\theta(\theta-1)^2}{(1+\theta^2)^2},$$

故 $\pi(\theta|x)$ 的导数在 $\theta \leq x$ 范围是内非负的, 因而 $\pi(\theta|x)$ 是单调增的, 故在 $\theta = x$ 时达到最大, 即 $\hat{\theta}_{\text{MD}} = x$ 是后验众数估计. 本例中的另外两个估计, 后验期望估计和后验中位数估计需要通过数值计算得到. 本例中的三种 Bayes 估计是互不相同的.

*2. Bayes 点估计的精度——估计的误差

设 θ 的后验分布为 $\pi(\theta|x)$, θ 的 Bayes 估计为 $\delta(x)$, 知道在经典方法中衡量一个估计量的优劣看其均方误差(在无偏估计情形看方差)大小, 一个估计量均方误差(MSE)越小越好. 在此处对 Bayes 估计 $\delta(x)$, 衡量优劣用如下定义的后验均方误差(posterior mean square error, PMSE):

$$\text{PMSE}(\delta(x)) = E^{\theta|x}[(\theta - \delta(x))^2]$$

来度量估计量的精度, PMSE 越小越好. 特别若 $\delta(x) = E(\theta|x)$ 时, 则 $\delta(x)$ 的 PMSE 即后验方差, 即

$$\text{PMSE}(\delta(x)) = E^{\theta|x}[(\theta - \delta(x))^2] = D(\theta|x).$$

记 θ 的后验均值 $E(\theta|x) = \mu^\pi(x)$, 后验方差 $V^\pi(x)$, 则 PMSE 与后验方差 $V^\pi(x)$ 的关系如下:

$$\begin{aligned}\text{PMSE}(\delta(x)) &= E^{\theta|x}[(\theta - \delta(x))^2] = E^{\theta|x}[(\theta - \mu^\pi(x)) + (\mu^\pi(x) - \delta(x))]^2 \\ &= V^\pi(x) + (\mu^\pi(x) - \delta(x))^2 \\ &\geq V^\pi(x),\end{aligned}$$

且等号成立的充要条件是 $\delta(x) = \mu^\pi(x)$, 即 θ 的后验期望估计使 PMSE 达到最小. 故后验期望估计是在 PMSE 准则下的最优估计. 这就是为什么习惯上在三种 Bayes 估计(后验众数估计, 后验中位数估计, 后验期望估计)中常取后验期望 $\mu^\pi(x) = E(\theta|x)$ 作为 θ 的 Bayes 估计.

例 7.3.4 (续例 7.3.1) 设 $X \sim N(\theta, \sigma^2)$, σ^2 已知, θ 未知, 设 θ 的先验分布为 $\pi(\theta) \sim N(\mu, \tau^2)$, 用 $\hat{\theta}_E = (\sigma^2\mu + \tau^2x)/(\sigma^2 + \tau^2)$ 作为 θ 的估计, 求估计量 $\hat{\theta}_E$ 的方差 $V^\pi(x)$, 并将其与 θ 的经典估计 $\delta(x) = x$ 的 PMSE 进行比较.

解 由例 7.3.1 中(取 $n = 1$) 可知后验分布 $\pi(\theta|x)$ 是 $N(\mu(x), \eta^2)$, 后验期望 $\mu(x) = \hat{\theta}_E$, 而后验方差为 η^2 , 故知

$$V^\pi(x) = \eta^2 = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}.$$

而 $\delta(x) = x$ 的 PMSE 为

$$\begin{aligned}\text{PMSE}(\delta(x)) &= \text{PMSE}(x) = V^\pi(x) + (\mu^\pi(x) - x)^2 \\ &= V^\pi(x) + \left(\frac{\sigma^2}{\sigma^2 + \tau^2}\right)^2 (\mu - x)^2 \geq V^\pi(x).\end{aligned}$$

由此可见, Bayes 估计 $\hat{\theta}_E$ 比经典估计 $\delta(x) = x$ 的估计误差要小.

例 7.3.5 设 $X \sim N(\theta, \sigma^2)$, 其中 σ^2 已知. 取 θ 的先验分布为无信息先验即 $\pi(\theta) \equiv 1$, 求 θ 的后验期望估计.

解 在例 7.2.4 中取 $n = 1$, 知 θ 的后验分布是 $N(x, \sigma^2)$, θ 的后验期望估计是

$$\mu^\pi(x) = E(\theta|x) = x, \quad V^\pi(x) = \sigma^2.$$

本例说明无信息先验得到的 Bayes 估计与经典方法所得 θ 的估计一致. 这又一次说明经典方法获得的估计是特殊先验分布下的 Bayes 估计.

例 7.3.6 (续例 7.3.2) 在例 7.3.2 中, $X \sim b(n, \theta)$, θ 的先验分布是 $Be(a, b)$, 求 θ 后验期望估计的后验方差. 又当取 θ 的先验分布是均匀分布 $U(0, 1)$ 时, 求后验期望估计的后验方差.

解 (1) 由例 7.3.2 已求出了 θ 的后验分布是 $Be(x+a, n-x+b)$, 故知 θ 的后验均值估计为 $\hat{\theta}_E = (x+a)/(n+a+b)$, 其后验方差为

$$V^\pi(x) = \frac{(x+a)(n-x+b)}{(a+b+n)^2(a+b+n+1)}.$$

(2) 若 θ 的先验分布是均匀分布 $U(0, 1)$, 即 $Be(1, 1)$, 故 θ 的后验分布是 $Be(x+1, n-x+1)$, 则后验均值估计和后验众数估计分别为

$$\hat{\theta}_E = \frac{x+1}{n+2}, \quad \hat{\theta}_{MD} = \frac{x}{n}.$$

易见 $\hat{\theta}_E$ 的后验方差是

$$V^\pi(x) = \frac{(x+1)(n-x+1)}{(n+2)^2(n+3)}.$$

而 $\hat{\theta}_{MD}$ 的后验均方误差是

$$\begin{aligned} PMSE(\hat{\theta}_{MD}) &= V^\pi(x) + (\hat{\theta}_{MD} - \hat{\theta}_E)^2 \\ &= \frac{(x+1)(n-x+1)}{(n+2)^2(n+3)} + \left(\frac{x+1}{n+2} - \frac{x}{n} \right)^2 \\ &\geq V^\pi(x). \end{aligned}$$

可见后验均值估计的精度比后验众数估计的精度高.

*3. 多参数情形

若 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$ 是向量, $\boldsymbol{\theta}$ 的后验分布为 $\pi(\boldsymbol{\theta}|x)$, 估计 $\boldsymbol{\theta}$ 的方法如下.

(1) 后验众数估计, 即从后验分布出发用广义极大似然估计 (后验众数估计);

(2) 后验期望估计, $\mu^\pi(x) = E(\theta|x) = (\mu_1^\pi(x), \dots, \mu_p^\pi(x))'$, 估计量的精度用后验协方差阵(记为 $\text{Cov}^\pi(x)$)来衡量

$$\text{Cov}^\pi(x) = E^{\theta|x}[(\theta - \mu^\pi(x))(\theta - \mu^\pi(x))'].$$

对 θ 的任一估计 $\delta(x)$, 其后验协方差阵可分解为

$$\text{Cov}_\delta^\pi(x) = E^{\theta|x}[(\theta - \delta)(\theta - \delta)'] = \text{Cov}^\pi(x) + (\mu^\pi(x) - \delta)(\mu^\pi(x) - \delta)'.$$

可见后验均值估计仍使后验协方差矩阵达到最小.

7.3.2 区间估计

对于区间估计问题, Bayes 方法具有处理方便和含义清晰的优点. 而经典方法寻求的置信区间常受到批评.

当 θ 的后验分布 $\pi(\theta|x)$ 获得后, 若存在区间 $[a, b]$, 使 θ 落在 $[a, b]$ 内的后验概率为 $1 - \alpha$, 即

$$P(a \leq \theta \leq b|x) = 1 - \alpha,$$

则称 $[a, b]$ 为 θ 的 Bayes 区间估计, 又称为可信区间. 这是对 θ 的后验分布为连续情形, 若 θ 为离散型随机变量, 对给定概率 $1 - \alpha$, 上述的区间 $[a, b]$ 不一定存在, 而要将左边概率适当放大一点使 $P(a \leq \theta \leq b|x) \geq 1 - \alpha$, 这样的区间也是 θ 的 Bayes 可信区间. 定义如下:

定义 7.3.2 (可信区间) 设参数 θ 的后验分布为 $\pi(\theta|x)$, 对给定的样本 x 和 $0 < \alpha < 1$ (通常 α 取较小的数), 若存在两个统计量 $\hat{\theta}_1 = \hat{\theta}_1(x)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(x)$, 使得

$$P(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2|x) \geq 1 - \alpha,$$

则称 $[\hat{\theta}_1, \hat{\theta}_2]$ 为 θ 的可信水平为 $1 - \alpha$ 的 Bayes 可信区间 (Bayesian credible interval). 而满足

$$P(\theta \geq \hat{\theta}_L|x) \geq 1 - \alpha$$

的 $\hat{\theta}_L = \hat{\theta}_L(x)$ 称为 θ 的可信水平为 $1 - \alpha$ 的可信下限. 而满足

$$P(\theta \leq \hat{\theta}_U|x) \geq 1 - \alpha$$

的 $\hat{\theta}_U = \hat{\theta}_U(x)$ 称为 θ 的可信水平为 $1 - \alpha$ 的可信上限.

可信水平和可信区间与经典统计方法中的置信水平和置信区间虽是同类概念, 但二者存在本质差别, 主要表现如下.

(1) 基于后验分布 $\pi(\theta|x)$, 在给定 x 和 $1-\alpha$ 后求得的可信区间, 如 $1-\alpha = 0.90$ 的可信区间为 $[1.2, 2.0]$, 这时可以写为

$$P(1.2 \leq \theta \leq 2.0|x) = 0.90.$$

既可以说 “ θ 属于这个区间的概率为 0.90”, 也可以说 “ θ 落入这个区间的概率为 0.90”. 可对置信区间就不能这样说. 因为经典统计方法认为 θ 为常数, 它要么 θ 在 $[1.2, 2.0]$ 之内, 要么在其外, 不能说 “ θ 落在 $[1.2, 2.0]$ 中的概率为 0.90”, 而只能说 “100 次重复使用这个置信区间, 大约有 90 次能覆盖 θ ”. 这种频率解释对仅使用这个置信区间一次或两次的人来说是毫无意义的. 可见 Bayes 可信区间简单, 自然易被人们接受和理解. 事实上很多实际工作者把求得的置信区间当可信区间去用.

(2) 经典统计方法寻求置信区间有时是困难的, 要设法构造一个枢轴变量, 要求它的表达式与 θ 有关, 而它的分布与 θ 无关. 这是一项技术性很强的工作, 有时找抽样分布相当困难, 而寻求可信区间只需要利用后验分布, 不需要寻求另外的分布, 与经典统计方法相比要简单得多.

例 7.3.7 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, 其中 σ^2 已知. θ 的先验分布是 $N(\mu, \tau^2)$, 其中 μ, τ^2 已知. 求 θ 的可信水平 $1-\alpha$ 的可信区间.

解 由例 7.2.10 已知 θ 的后验分布是 $N(\mu_n(\mathbf{x}), \eta_n^2)$, 故有

$$P(\mu_n(\mathbf{x}) - \eta_n u_{\alpha/2} \leq \theta \leq \mu_n(\mathbf{x}) + \eta_n u_{\alpha/2} | \mathbf{x}) = 1 - \alpha,$$

其中 $u_{\alpha/2}$ 为 $N(0, 1)$ 的上侧 $\alpha/2$ 分位数. 因此 $[\mu_n(\mathbf{x}) - \eta_n u_{\alpha/2}, \mu_n(\mathbf{x}) + \eta_n u_{\alpha/2}]$ 就是 θ 的可信水平 $1-\alpha$ 的可信区间.

看一个数值例子 (其中 $n = 1$). 在儿童智商测验的例子中, 设测验结果 $X \sim N(\theta, 100)$, 其中 θ 为被测验儿童智商 IQ 真值, θ 的先验分布是 $N(100, 225)$, 由前面结果可知后验分布 $\pi(\theta|x)$ 是 $N(\mu(x), \eta^2)$, 当该儿童测验得分 $x = 115$ 时, 有

$$\begin{aligned}\mu(x) &= \frac{\sigma^2}{\sigma^2 + \tau^2} \mu + \frac{\tau^2}{\sigma^2 + \tau^2} x = \frac{4}{13} \times 100 + \frac{9}{13} \times 115 = 110.38, \\ \eta^2 &= \frac{100 \times 225}{100 + 225} = 8.32^2.\end{aligned}$$

故 θ 的后验分布是 $N(110.38, 8.32^2)$, 因此求得 θ 的可信系数为 0.95 的可信区间为

$$[\mu(x) - 1.96\eta, \mu(x) + 1.96\eta] = [94.07, 126.69].$$

在这个例子中, 若不用先验信息, 用经典方法 $X \sim N(\theta, 100)$, $x = 115$ 可求 θ 的置信系数 0.95 的置信区间为

$$[115 - 10 \times 1.96, 115 + 10 \times 1.96] = [95.4, 134.6].$$

这两个区间是不同的, 区间长度也不同. 可信区间长度短了一些, 这是由于利用了先验信息. 另一个不同是经典方法不能说“ θ 落入区间 $[95.4, 134.6]$ 中的概率为 0.95”, 也不能说“ θ 属于此区间的概率为 0.95”. 在这一束缚下这个区间还能有什么用? 这就是经典置信区间常受到批评的原因. 但仍有不少人在使用这个区间, 把它当可信区间去用.

7.3.3 假设检验

1. 假设检验的一般方法

设检验问题仍如式 (5.1.1) 所述, 即

$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1.$$

经典的假设检验方法中, 对上述检验问题要通过对犯 I, II 型错误的概率的大小来评价检验优劣.

Bayes 方法处理假设检验问题是直截了当的. 在求得 θ 的后验分布 $\pi(\theta|x)$ 后, 计算 Θ_0 和 Θ_1 的后验概率

$$\alpha_0 = P(\theta \in \Theta_0|x), \quad \alpha_1 = P(\theta \in \Theta_1|x).$$

比较 α_0 和 α_1 的大小决定接受 H_0 还是 H_1 . α_0 和 α_1 是综合样本信息和先验信息得出的两个假定实际发生的概率. 检验法则如下:

$$\text{当 } \frac{\alpha_0}{\alpha_1} > 1 \text{ 时接受 } H_0, \text{ 否则拒绝 } H_0. \quad (7.3.1)$$

由此可见, Bayes 假设检验方法是简单的. 与频率方法相比, 它无需选择检验统计量、确定抽样分布; 无需给出显著性水平, 确定否定域; 而且容易推广到多重假设检验情形.

例 7.3.8 设随机变量 X 是从二项分布 $b(n, \theta)$ 中抽取的一个样本, 取 θ 的先验分布为均匀分布 $U(0, 1)$, 求下列检验问题:

$$H_0 : \theta \leq 1/2 \leftrightarrow H_1 : \theta > 1/2.$$

解 由例 7.2.9 已知后验分布为 $\text{Be}(x+1, n-x+1)$, 故有

$$\alpha_0 = P(\Theta_0|x) = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} \int_0^{1/2} \theta^x (1-\theta)^{n-x} d\theta.$$

当取 $n=5$ 时可算得各种 x 下的后验概率及后验机会比 α_0/α_1 如下:

表 7.3.2 θ 的后验机会比

x	0	1	2	3	4	5
α_0	$63/64$	$57/64$	$42/64$	$22/64$	$7/64$	$1/64$
α_1	$1/64$	$7/64$	$22/64$	$42/64$	$57/64$	$63/64$
α_0/α_1	63.0	8.14	1.91	0.52	0.12	0.016

可见当 $x = 0, 1, 2$ 时接受 H_0 , 当 $x = 3, 4, 5$ 时拒绝 H_0 .

例 7.3.9 考虑对一个儿童作智力测验的情形. 设测验结果 $X \sim N(\theta, 100)$, 其中 θ 为这个孩子测验中的智商 IQ 真值, θ 的先验分布为 $N(100, 225)$. 该儿童测验得分 $x = 115$, 求下列检验问题:

$$H_0 : \theta \leq 100 \leftrightarrow H_1 : \theta > 100.$$

解 由例 7.3.7 可知 θ 的后验分布为 $N(110.38, 8.32^2)$, 查标准正态分布表求得

$$\alpha_0 = P(\theta \leq 100|x) = 0.106, \quad \alpha_1 = P(\theta > 100|x) = 0.894.$$

后验机会比 $\alpha_0/\alpha_1 = 1/8.44$. 这一比值远小于 1, 故否定 H_0 .

2. 多重假设检验

按 Bayes 分析的观点, 多重假设检验并不比两个假设检验更困难, 即直接计算每一个假设的后验概率, 并比较其大小. 设检验问题是

$$H_i : \theta \in \Theta_i, \quad i = 1, \dots, k, \quad \text{其中 } \Theta_1 \cup \Theta_2 \cup \dots \cup \Theta_k = \Theta,$$

其中 Θ 是参数空间, Θ_i ($i = 1, \dots, k$) 是 Θ 的非空真子集, 且两两不相交. 在求得 θ 的后验分布 $\pi(\theta|x)$ 后, 计算 Θ_i 的后验概率

$$\alpha_i = P(\Theta_i|x), \quad i = 1, \dots, k,$$

取其最大者, 则认为相应的假设成立.

例 7.3.10 (续例 7.3.9) 考虑对儿童作智力测验的情形, 求下列多重检验问题:

$$H_1 : \theta \leq 90, \quad H_2 : 90 < \theta < 110, \quad H_3 : \theta \geq 110.$$

解 由于后验分布 $\pi(\theta|x)$ 是 $N(110.38, 8.32^2)$, 故查标准正态分布表得出

$$P(\theta \leq 90|x = 115) = 0.007, \quad P(90 < \theta < 110|x = 115) = 0.473,$$

$$P(\theta \geq 110|x = 115) = 0.520.$$

由于 $\Theta_3 = \{\theta : \theta \geq 110\}$ 的后验概率最大, 故接受 H_3 .

*7.3.4 预测推断

一般预测问题的典型情况是：若 $X \sim f(x|\theta)$, $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 X 中抽取的历史样本，对具有密度为 $g(z|\theta)$ 的随机变量 Z 未来观察值作出预测。通常假定 Z 和 X 不相关， f 和 g 分别为密度函数。Bayes 预测的想法是：由 $\pi(\theta|x)$ 为 θ 的后验分布，于是 $g(z|\theta)\pi(\theta|x)$ 为给定 $X = x$ 的条件下 (Z, θ) 的联合分布，把它对 θ 积分得到给定 $X = x$ 时 Z 的边缘分布密度作为预测密度。

定义 7.3.3 设 θ 的后验密度是 $\pi(\theta|x)$ 。给定 \mathbf{x} 后，随机变量 Z 的后验预测密度 (posterior predictive density) 定义为

$$p(z|\mathbf{x}) = \int_{\Theta} g(z|\theta)\pi(\theta|\mathbf{x})d\theta. \quad (7.3.2)$$

例 7.3.11 一赌徒在过去 10 次赌博中赢了 3 次，现要对未来 5 次赌博中他赢的次数 Z 作出预测。

解 这个问题的一般提法是：在 n 次独立的 Bernoulli 试验中成功了 X 次，现要对未来的 k 次相互独立的 Bernoulli 试验中成功次数作预测。

现设成功概率为 θ ，则 n 次独立的 Bernoulli 试验中成功次数 X 的概率函数 (亦称为似然函数) 为

$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \dots, n.$$

若取 θ 的先验分布为共轭先验 $\text{Be}(a, b)$ ，则由例 7.2.9 可知，后验密度为

$$\pi(\theta|x) = \frac{\Gamma(n+a+b)}{\Gamma(x+a)\Gamma(n-x+b)} \theta^{x+a-1} (1-\theta)^{n-x+b-1}.$$

设 Z 表示“未来的 k 次相互独立的 Bernoulli 试验中成功的次数”，其概率函数为

$$g(z|\theta) = \binom{k}{z} \theta^z (1-\theta)^{k-z}, \quad z = 0, 1, \dots, k.$$

于是在给定 $X = x$ 时，随机变量 Z 的后验预测密度为

$$\begin{aligned} p(z|x) &= \int_0^1 \binom{k}{z} \theta^z (1-\theta)^{k-z} \pi(\theta|x)d\theta. \\ &= c \int_0^1 \theta^{z+x+a-1} (1-\theta)^{k-z+n-x+b-1} d\theta \\ &= c \frac{\Gamma(z+x+a)\Gamma(k-z+n-x+b)}{\Gamma(k+n+a+b)}, \end{aligned}$$

其中

$$c = \binom{k}{z} \frac{\Gamma(a+b+n)}{\Gamma(x+a)\Gamma(n-x+b)}.$$

在此问题中, $n = 10$, $x = 3$, $k = 5$. 取 $a = b = 1$, 即先验分布为 $\text{Be}(1, 1) = U(0, 1)$, 则

$$p(z|x=3) = \binom{5}{z} \frac{\Gamma(12)\Gamma(4+z)\Gamma(13-z)}{\Gamma(4)\Gamma(8)\Gamma(17)}, \quad z = 0, 1, 2, 3, 4, 5.$$

经计算可得 $z = 0, 1, 2, 3, 4, 5$ 时后验预测概率

$$p(0|3) = 0.1813, \quad p(1|3) = 0.3022, \quad p(2|3) = 0.2747,$$

$$p(3|3) = 0.1649, \quad p(4|3) = 0.0641, \quad p(5|3) = 0.02128.$$

由此后验预测分布可见, 它的概率集中在 $Z = 0, 1, 2, 3$ 之间, 即

$$P_{Z|x}(0 \leq Z \leq 3) = 0.9231.$$

这表明 $[0, 3]$ 是 Z 的 92% 预测区间. 另外分布的众数在 $z = 1$ 处, 第二大的概率在 $z = 2$ 处出现, 可见未来 5 次赌博中胜 1 次或 2 次的可能性最大.

7.4 Bayes 统计决策理论

7.4.1 统计决策理论的若干基本概念

1. 决策问题

为了说明什么是决策 (判决) 问题, 请看下例.

例 7.4.1 一位投资者有一笔资金要进行投资, 有如下几个投资方案供选择:

a_1 : 购买股票, 根据市场情况可净赚 5000 元, 但也可能亏损 10000 元.

a_2 : 购买基金, 根据市场情况可净赚 3000 元, 但也可能亏损 8000 元.

a_3 : 存入银行, 不管市场情况如何, 总可净赚 1000 元.

他应如何决策?

这位投资者在与金融市场博弈. 未来的金融市场也有两种情况: 看涨 (θ_1) 与看跌 (θ_2). 根据上述情况, 可写出投资者的收益矩阵如下:

表 7.4.1

行动	a_1	a_2	a_3
θ_1	5000	3000	1000
θ_2	-10000	-8000	1000

投资者将依据收益矩阵决定资金投向何方. 这种人与自然界 (或社会) 的博弈问题称为决策问题. 在决策问题中, 主要寻求人对自然界 (或社会) 的最优策略. 决策问题也不一定要涉及统计方法. 如果它满足以下条件, 那就必然与统计方法有关, 因而就可以称为统计决策问题. 这条件是: 在作出决策时所依据的事实中, 至少有一个部分是受到随机性影响的观察值 (或试验数据).

决策实际上是一个过程, 它可分为两部分: 第一部分是把决策问题描述清楚; 第二部分是如何作决策使得收益最大 (或损失最小). 显然第二部分是研究的重点, 但首先要把第一部分搞清楚, 这就需要下面一些基本概念.

2. 统计决策三要素和贝叶斯统计决策四要素

(1) 样本空间和样本分布族.

取值于样本空间 \mathcal{X} 内的随机变量 X 及其分布族是 $\{f(x, \theta), \theta \in \Theta\}$ 构成统计决策问题的第一个要素, 其中 $f(x, \theta)$ 是 X 的概率函数, θ 是未知参数, Θ 为参数空间, $\mathbf{X} = (X_1, \dots, X_n)$ 是总体 X 的简单样本.

(2) 行动空间.

决策者或统计工作者对某个统计决策问题可能采取的行动所构成的非空集合, 被称为行动空间 (action space), 记为 \mathcal{A} . 在估计问题中, \mathcal{A} 由一切估计量 $\delta(x)$ 构成, 常取 $\mathcal{A} = \Theta$. 在检验问题中 \mathcal{A} 只有两个行动组成, 即 $\mathcal{A} = \{a_0, a_1\}$, 其中 a_0 表示接受原假设 H_0 ; a_1 表示拒绝 H_0 , 接受 H_1 .

(3) 损失函数.

损失函数 (loss function) 是定义于 $\Theta \times \mathcal{A}$ 上的非负函数, 记为 $L(a, \theta)$. 它表示参数为 $\theta \in \Theta$ 时, 采取行动 $a \in \mathcal{A}$ 所蒙受的损失. 损失函数的类型很多, 常用的有“平方损失”、“绝对值损失”和“线性损失”等.

对 Bayes 统计决策问题, 除上式三要素外还应增加第四个要素:

(4) 先验分布.

定义在参数空间 Θ 上的先验分布函数 $F^\pi(\theta)$ 或概率函数 $\pi(\theta)$.

统计决策 (贝叶斯统计决策) 问题就是研究如何根据样本 x 的值, 恰当地选取行动 a 使得按样本分布 (或后验分布) 计算的平均损失最小.

3. 风险函数和一致最优决策函数

定义 7.4.1 定义于样本空间 \mathcal{X} 内而取值于行动空间 \mathcal{A} 内的函数 $\delta = \delta(x)$ 称为决策函数或判决函数 (decision rules).

设 δ 是采取的决策行动, 若参数为 θ , 则由此造成的损失是 $L(\delta, \theta)$, 这个量与样本 \mathbf{X} 有关, 因而是随机的. 故采取行动 δ 的效果用平均损失去度量是相对合理的. 这就引入如下风险函数的概念.

定义 7.4.2 设 δ 是一个决策函数, 称平均损失

$$R(\delta, \theta) = E[L(\delta(\mathbf{X}), \theta)] = \int_{\mathcal{X}} L(\delta(\mathbf{x}), \theta) f(x, \theta) dx \quad (7.4.1)$$

为 δ 的风险函数 (risk function).

按照 Wald 的统计决策理论, 评价一个决策函数的唯一依据, 就是其风险函数. 风险函数越小越好. 有了风险函数后可比较不同决策函数的优劣.

定义 7.4.3 (一致最优决策函数) 记 $R(\delta, \theta)$ 为决策函数 δ 的风险函数. 若存在一个决策函数 δ^* , 使得对任一决策函数 δ 有

$$R(\delta^*, \theta) \leq R(\delta, \theta), \quad \text{一切 } \theta \in \Theta, \quad (7.4.2)$$

则称 δ^* 为一致最优解 或一致最优决策函数.

注 7.4.1 若一致最优解 (或一致最优决策函数) 存在, 则毫无疑问应当采用它. 但是除了某些例外情形, 一致最优解通常不存在. 因此必须把标准放宽些, 引进一些比一致最优准则更弱的优良性准则. 大致有两种途径, 一是用某种方法制定优良性的综合指标, 以之作为比较的标准. 例如, 本节和 7.5 节将要讨论的 Bayes 准则和 Minimax 准则就属于这一类. 另一类制定优良性准则的方法是对决策函数提出合理的要求, 缩小所考虑的决策函数的范围, 从中找一致最优者. 例如, 无偏性准则就是这样一种要求, UMVUE 就是属于由这类最优化准则获得的估计. 7.5 节将要介绍的同变性也属于这一类.

4. Bayes 准则

前一段已说过, 一致最优的决策函数通常不存在, 故有必要把标准放宽些, 引进一些条件更弱的优良性准则. 下列定义的 Bayes 风险, 就是制定一种优良性的综合指标, 作为比较决策函数的标准.

定义 7.4.4 设 $R(\delta, \theta)$ 为 δ 的风险函数, $H(\theta)$ 为 θ 的先验分布 (若存在密度, 用 $\pi(\theta)$ 表示), 则称

$$R_H(\delta) = E^\theta[R(\delta, \theta)] = \int_{\Theta} R(\delta, \theta) dH(\theta) \quad (7.4.3)$$

为 δ 的 Bayes 风险 (Bayes risk).

由定义可见 Bayes 风险是将风险函数以 θ 的先验分布 $H(\theta)$ 为权的一种加权平均. 使 Bayes 风险达到最小的决策函数称为决策问题的 Bayes 解, 其定义如下.

定义 7.4.5 设 δ_1 和 δ_2 为 θ 的两个决策函数, 若 $R_H(\delta_1) \leq R_H(\delta_2)$, 则称 δ_1 在 Bayes 风险下优于 δ_2 . 若存在 δ^* , 使得对任一决策函数 δ 有

$$R_H(\delta^*) \leq R_H(\delta), \quad (7.4.4)$$

则称 δ^* 为所考虑的统计决策问题的 Bayes 解.

7.4.2 后验风险最小的原则

1. 后验风险

设给定 θ 时, 随机变量 $X \sim f(x|\theta)$, θ 的先验分布是 $H(\theta)$ (如有密度, 用 $\pi(\theta)$)

记之), θ 的后验分布函数为 $H(\theta|x)$ (如有后验密度, 以 $\pi(\theta|x)$ 记之). 令 $L(\theta, \delta)$ 为损失函数, 将其按后验分布 $H(\theta|x)$ 求平均, 得到后验风险. 其定义如下.

定义 7.4.6 设 $H(\theta|x)$ 为 θ 的后验分布函数, $L(\delta, \theta)$ 为损失函数, 则称

$$R(\delta|x) = E^{\theta|x}[L(\delta, \theta)] = \int_{\Theta} L(\delta, \theta) dH(\theta|x) \quad (7.4.5)$$

为决策函数 δ 的后验风险 (posterior risk).

若存在决策函数 δ^* , 使得对任一决策函数 δ 有

$$R(\delta^*|x) = \min_{\delta} R(\delta|x), \quad (7.4.6)$$

则称 δ^* 为后验风险最小准则下的最优 Bayes 决策函数.

2. 后验风险与 Bayes 风险的关系

由 Bayes 风险 $R_H(\delta)$ 的定义可知

$$\begin{aligned} R_H(\delta) &= E^{\theta}[R(\delta, \theta)] = \int_{\Theta} R(\delta, \theta) dH(\theta) = \int_{\Theta} \left[\int_{\mathcal{X}} L(\delta, \theta) dF(x|\theta) \right] dH(\theta) \\ &= \int_{\mathcal{X}} \left[\int_{\Theta} L(\delta, \theta) dH(\theta|x) \right] dF_m(x) = \int_{\mathcal{X}} R(\delta|x) dF_m(x) \\ &= E^X[R(\delta|x)], \end{aligned} \quad (7.4.7)$$

其中 $F_m(x)$ 是 X 的边缘分布 (如有密度, 以 $f_m(x)$ 记之), E^{θ} 表示“关于 θ 的先验分布”求期望, E^X 表示“关于 X 的绝对 (边缘) 分布”求期望. 由上式可见 Bayes 风险有两种表达式 $R_H(\delta) = E^{\theta}[R(\delta, \theta)] = E^X[R(\delta|x)]$, 即 Bayes 风险既可以看成将风险函数关于 θ 的先验分布求平均, 也可以看成是对后验风险关于 X 的绝对 (边缘) 分布求平均.

3. 后验风险最小的原则

可以证明: 后验风险最小准则下的决策函数 (7.4.6) 就是 Bayes 解. Bayes 解的定义由式 (7.4.4) 给出, 它是使 Bayes 风险达到最小的决策函数.

定理 7.4.1 对任何样本值 x , 若存在决策函数 $\delta_H(x)$, 满足条件

$$R(\delta_H|x) = \inf_{\delta \in \mathcal{A}} R(\delta|x),$$

则 δ_H 为先验分布 $H(\theta)$ 之下的 Bayes 解.

证 设 δ 为任一非随机化决策函数, 由已知条件可知

$$R(\delta|x) = \int_{\Theta} L(\delta, \theta) H(d\theta|x) \geq \int_{\Theta} L(\delta_H, \theta) H(d\theta|x) = R(\delta_H|x)$$

对一切 $x \in \mathcal{X}$ 成立. 将上式两边关于 X 的绝对分布求平均. 故由式 (7.4.7) 可得

$$R_H(\delta) = \int_{\mathcal{X}} R(\delta|x)dF_m(x) \geq \int_{\mathcal{X}} R(\delta_H|x)dF_m(x) = R_H(\delta_H),$$

即使 Bayes 风险达到最小的决策函数为 δ_H , 因此它是 Bayes 解.

注 7.4.2 当 θ 的先验分布为广义先验分布, 定理的结果仍然是对的. 此时后验风险最小准则下的决策函数, 称为广义 Bayes 解.

7.4.3 一般损失函数下的 Bayes 估计

常见损失函数有平方损失、加权平方损失、绝对值损失等. 下面将分别讨论在这几种损失函数下的 Bayes 估计.

1. 平方损失下的 Bayes 估计

定理 7.4.2 设 $a = a(x)$ 为一决策函数, 则在平方损失 $L(a, \theta) = (a - \theta)^2$ 下, θ 的 Bayes 估计为其后验期望. 即

$$\hat{\theta}_B(x) = E(\theta|x). \quad (7.4.8)$$

证 由定理 7.4.1 可知 Bayes 解就是后验风险达到最小的决策函数, 因此有

$$R(a|x) = E[(\theta - a)^2|x] = \int_{\Theta} [\theta^2 - 2a\theta + a^2] H(d\theta|x).$$

令

$$\frac{dR(a|x)}{da} = -2 \int_{\Theta} \theta H(d\theta|x) + 2a = 0,$$

解方程得 $a = \hat{\theta}_B(x) = \int_{\Theta} \theta H(d\theta|x) = E(\theta|x)$, 且 $\frac{d^2}{da^2}[R(a|x)] = 2 > 0$. 定理得证.

例 7.4.2 设 $X \sim N(\theta, \sigma^2)$, 其中 σ^2 已知, θ 的先验分布是 $N(\mu, \tau^2)$. 求平方损失下 θ 的 Bayes 估计.

解 由例 7.2.10 可知 θ 的后验分布是 $N(\mu(x), \eta^2)$, 其中

$$\mu(x) = \frac{\sigma^2}{\sigma^2 + \tau^2} \mu + \frac{\tau^2}{\sigma^2 + \tau^2} x, \quad \eta^2 = \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2}.$$

由定理 7.4.2 可知, 平方损失下 θ 的 Bayes 估计是

$$\hat{\theta}_B(x) = E(\theta|x) = \mu(x) = \frac{\sigma^2}{\sigma^2 + \tau^2} \mu + \frac{\tau^2}{\sigma^2 + \tau^2} x.$$

例 7.4.3 设 $X = (X_1, \dots, X_n)$ 是从 Poisson 分布 $P(\theta)$ 中抽取的简单样本. 取 θ 的先验分布为 $\Gamma(\alpha, \lambda)$, 即 $\pi(\theta) = (\lambda^\alpha / \Gamma(\alpha)) \theta^{\alpha-1} e^{-\lambda\theta} I_{[\theta>0]}$, 其中 $\lambda > 0$, $\alpha > 0$ 已知, 求平方损失下 θ 的 Bayes 估计.

解 由例 7.2.11 可知, θ 的后验分布为 $\Gamma(n\bar{x} + \alpha, n + \lambda)$, 故知 Bayes 估计

$$\hat{\theta}_B(x) = \frac{n\bar{x} + \alpha}{n + \lambda} = \frac{n}{n + \lambda}\bar{x} + \frac{\lambda}{n + \lambda} \cdot \left(\frac{\alpha}{\lambda}\right),$$

其中 \bar{x} 为样本均值, α/λ 为先验分布的均值, 可见 θ 的 Bayes 估计为样本均值和先验均值的加权平均. 当 $n \gg \lambda$ 时, 样本均值 \bar{x} 在 Bayes 估计中起主导作用, 当 $\lambda \gg n$ 时, 先验均值 α/λ 在 Bayes 估计中起主导作用, 所以 Bayes 估计是合理的.

2. 加权平方损失下的 Bayes 估计

定理 7.4.3 设 $a = a(x)$ 为一决策函数, 则在加权平方损失 $L(a, \theta) = w(\theta)(a - \theta)^2$ 下, θ 的 Bayes 估计为

$$\hat{\theta}_B = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)}, \quad (7.4.9)$$

其中 $w(\theta)$ 为参数空间上的正值实函数.

证 由定理 7.4.1 可知, Bayes 解就是后验风险达到最小的决策函数, 故有

$$R(a|x) = E[w(\theta)(\theta - a)^2|x] = \int_{\Theta} [w(\theta)\theta^2 - 2\theta w(\theta)a + w(\theta)a^2] H(d\theta|x).$$

令

$$\frac{d}{da}[R(a|x)] = -2 \int_{\Theta} \theta w(\theta)H(d\theta|x) + 2a \int_{\Theta} w(\theta)H(d\theta|x) = 0,$$

解方程得

$$a = \hat{\theta}_B = \frac{\int_{\Theta} \theta w(\theta)H(d\theta|x)}{\int_{\Theta} w(\theta)H(d\theta|x)} = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)},$$

且 $\frac{d^2}{da^2}[R(a|x)] > 0$. 定理得证.

例 7.4.4 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从下列指数分布中抽取的简单样本:

$$f(x|\theta) = \begin{cases} \theta^{-1} e^{-x/\theta}, & x > 0, \\ 0, & \text{其他,} \end{cases}$$

其中 $\theta > 0$. 设 θ 的先验分布服从逆 Gamma 分布 $\Gamma^{-1}(\alpha, \lambda)$, 即先验密度是

$$\pi(\theta) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{-(\alpha+1)} e^{-\lambda/\theta}, & \theta > 0, \\ 0, & \text{其他.} \end{cases}$$

求 θ 在加权平方损失 $L(\delta, \theta) = (\theta - \delta)^2/\theta^2$ 下的 Bayes 估计.

解 X 的似然函数为

$$L(\theta|x) = f(x|\theta) \propto \theta^{-n} e^{-n\bar{x}/\theta}.$$

类似于例 7.2.13, 可知后验分布

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \theta^{-(n+\alpha+1)} e^{-(n\bar{x}+\lambda)/\theta},$$

添加正则化常数

$$\pi(\theta|x) = \frac{(n\bar{x} + \lambda)^{n+\alpha}}{\Gamma(n+\alpha)} \cdot \theta^{-(n+\alpha+1)} e^{-(n\bar{x}+\lambda)/\theta} I_{(0,\infty)}(\theta).$$

θ 的 Bayes 估计由式 (7.4.9) 给出, 其中

$$\begin{aligned} E(\theta w(\theta)|x) &= E(\theta^{-1}|x) = \int_0^\infty \theta^{-1} \cdot \pi(\theta|x)d\theta = \frac{n+\alpha}{n\bar{x}+\lambda}, \\ E(w(\theta)|x) &= E(\theta^{-2}|x) = \int_0^\infty \theta^{-2} \cdot \pi(\theta|x)d\theta = \frac{(n+\alpha+1)(n+\alpha)}{(n\bar{x}+\lambda)^2}, \end{aligned}$$

因此得到

$$\hat{\theta}_B = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)} = \frac{(n+\alpha)/(n\bar{x}+\lambda)}{(n+\alpha+1)(n+\alpha)/(n\bar{x}+\lambda)^2} = \frac{n\bar{x}+\lambda}{n+\alpha+1}.$$

3. 绝对值损失下的 Bayes 解

定理 7.4.4 设 $\delta = \delta(x)$ 为一决策函数, 则在绝对损失 $L(\delta, \theta) = |\delta - \theta|$ 下, θ 的 Bayes 估计为后验中位数.

证 见文献 [16] (中译本) 第 117-118 页.

例 7.4.5 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从均匀分布 $U(0, \theta)$ 中抽取的简单样本, θ 的先验分布是 Pareto 分布, 其分布函数和密度函数为

$$H(\theta) = 1 - \left(\frac{\theta_0}{\theta}\right)^\alpha, \quad \theta > \theta_0; \quad \pi(\theta) = \frac{\alpha \theta_0^\alpha}{\theta^{\alpha+1}}, \quad \theta > \theta_0.$$

求 θ 在绝对值损失下的 Bayes 估计.

解 X_1, \dots, X_n 的联合密度 (或称为似然函数) 为

$$f(\mathbf{x}|\theta) = \frac{1}{\theta^n} I_{(0,\theta)}(x_{(n)}),$$

其中 $x_{(n)} = \max\{x_1, \dots, x_n\}$. 由式 (7.2.3) 可知

$$\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta) \propto \frac{1}{\theta^{n+\alpha+1}}, \quad \theta > \max\{x_{(n)}, \theta_0\}.$$

记 $\theta_1 = \max\{x_{(n)}, \theta_0\}$, 添加正则化常数得后验密度和后验分布函数如下:

$$\pi(\theta|x) = \frac{(\alpha+n)\theta_1^{\alpha+n}}{\theta^{n+\alpha+1}}, \quad \theta > \theta_1; \quad H(\theta|x) = 1 - \left(\frac{\theta_1}{\theta}\right)^{\alpha+n}, \quad \theta > \theta_1.$$

这仍为 Pareto 分布, 故上述先验分布是共轭先验分布. 在绝对值损失下 Bayes 估计 $\hat{\theta}_B$ 是后验中位数, 解方程 $1 - (\theta_1/\theta)^{\alpha+n} = 1/2$ 得

$$\hat{\theta}_B = 2^{\frac{1}{\alpha+n}} \cdot \theta_1,$$

若取平方损失, 则 θ 的 Bayes 估计为后验均值, 即

$$\hat{\theta}_{B_1} = \frac{\alpha+n}{\alpha+n-1} \theta_1.$$

这个估计要比经典方法中极大似然估计 $\hat{\theta} = \max(x_1, \dots, x_n)$ 大些, 因为

$$\frac{\alpha+n}{\alpha+n-1} > 1.$$

注 7.4.3 当后验分布是对称时, 后验均值也是后验中位数, 二者相同. 例如, 在例 7.3.7 儿童智商测验的例子中, 后验分布仍为正态, 后验均值也是后验中位数, 故绝对值损失下的 Bayes 解也为 $E(\theta|x) = \mu(x)$.

7.4.4 假设检验和有限行动 (分类) 问题

在估计问题中, 一般有无穷多个行动可供选择. 然而有很多重要的统计问题, 只能在有限个行动中选择. 最常见的有限行动问题是假设检验问题. 这类问题用 Bayes 决策方法是容易解决的. 设行动空间有 r 个行动组成, 即 $\mathcal{A} = \{a_1, \dots, a_r\}$. 令在参数为 θ , 行动为 a_i 下的损失函数是 $L(a_i, \theta)$, $i = 1, \dots, r$. 按后验风险最小准则, Bayes 决策就是使后验风险 $R(a_i|x) = E^{\theta|x}[L(a_i, \theta)]$ 达到最小的那个行动. 下面首先讨论两个行动的假设检验问题.

1. 假设检验问题

设有如下的两个假设:

$$H_0: \theta \in \Theta_0 \leftrightarrow H_1: \theta \in \Theta_1 \quad \text{且} \quad \Theta_0 \cup \Theta_1 = \Theta.$$

决策行动有两个: a_0 和 a_1 , 其中 a_i 表示接受 H_i 的行动, $i = 0, 1$. 决策者要在这两个行动中选择一个. 以下就不同的损失函数来考虑:

(1) 若为 0-1 损失.

$$L(a_i, \theta) = \begin{cases} 0, & \theta \in \Theta_i, \\ 1, & \theta \notin \Theta_i, \end{cases} \quad i = 0, 1.$$

其后验风险为

$$R(a_0|x) = \int_{\Theta} L(a_0, \theta) \pi(\theta|x) d\theta = \int_{\Theta_1} \pi(\theta|x) d\theta = P(\Theta_1|x),$$

类似可求

$$R(a_1|x) = E^{\theta|x} [L(a_1, \theta)] = P(\Theta_0|x).$$

按后验风险最小准则, 若 $R(a_0|x) \geq R(a_1|x)$, 则否定 H_0 , 即接受 H_1 . 等价地, 若

$$P(\Theta_1|x) \geq P(\Theta_0|x) \text{ 时否定 } H_0, \text{ 接受 } H_1. \quad (7.4.10)$$

因此, Bayes 决策就是接受具有较大后验概率的假设. 这与 Bayes 统计推断中的结论是一致的.

(2) 若为 $0-k_i$ 损失.

$$L(a_i, \theta) = \begin{cases} 0, & \theta \in \Theta_i, \\ k_i, & \theta \notin \Theta_i, \end{cases} \quad i = 0, 1.$$

此时后验风险分别为

$$\begin{aligned} R(a_0|x) &= E^{\theta|x} [L(a_0, \theta)] = k_0 P(\Theta_1|x), \\ R(a_1|x) &= E^{\theta|x} [L(a_1, \theta)] = k_1 P(\Theta_0|x). \end{aligned}$$

按后验风险最小准则, 若

$$R(a_0|x) > R(a_1|x) \quad (7.4.11)$$

则否定 H_0 ; 若不等式反向, 则接受 H_0 . 式 (7.4.11) 的等价形式为

$$k_0 P(\Theta_1|x) > k_1 P(\Theta_0|x),$$

即当

$$\frac{k_0}{k_1} > \frac{P(\Theta_0|x)}{P(\Theta_1|x)} \text{ 时否定 } H_0.$$

由于 $\Theta_0 \cup \Theta_1 = \Theta$, 故有 $P(\Theta_0|x) = 1 - P(\Theta_1|x)$, 所以上式可改写为

$$P(\Theta_1|x) > \frac{k_1}{k_0 + k_1} \text{ 时否定 } H_0, \text{ 否则接受 } H_0. \quad (7.4.12)$$

例 7.4.6 (续儿童智力测验问题) 考虑对一个儿童作智力测验的情形. 设测验结果 $X \sim N(\theta, 100)$, 其中 θ 为这个孩子测验中的智商 IQ 真值, θ 的先验分布为 $N(100, 225)$. 该儿童测验得分 $x = 115$. 取损失函数为 0-1 损失, 求下列检验问题

$$H_0 : \theta \leq 105 \leftrightarrow H_1 : \theta > 105.$$

解 由例 7.3.7 已知后验分布 $\pi(\theta|x=115)$ 是 $N(110.38, 8.32^2)$. 利用标准正态分布表可得

$$P(\Theta_0|x=115) = P(\theta \leq 105|x=115) = \Phi(-5.38/8.32) = 0.2578,$$

$$P(\Theta_1|x=115) = 1 - P(\Theta_0|x=115) = 0.7422.$$

由于 $P(\Theta_1|x=115) > P(\Theta_0|x=115)$, 故由检验准则 (7.4.10) 可知, 否定 H_0 .

2. 多行动 (分类) 问题

很多决策问题可能采取的行动多于两个. 例如, 在假设检验问题中常常存在两者皆可的区域, 即除 $\theta \in \Theta_0$ 及 $\theta \in \Theta_1$ 分别采取行动 a_0 和 a_1 之外, 还存在第三个行动 a_2 , 它表示当 $\theta \in \Theta_2$ 时采取两者皆可的行动. 例如, 若要求检验两种药物的治愈率, 合理方法是检验下列三个假设:

$$H_0 : \theta_1 - \theta_2 < -\varepsilon, \quad H_1 : \theta_1 - \theta_2 > \varepsilon, \quad H_2 : |\theta_1 - \theta_2| \leq \varepsilon,$$

其中 $\varepsilon > 0$ 的选择使得当 $|\theta_1 - \theta_2| \leq \varepsilon$ 时两种药物被认为等效.

即使在经典的假设检验中, 也有三个行动可供选择: a_0 表示接受 H_0 , a_1 表示拒绝 H_0 , 接受 H_1 , 而 a_2 表示接受 H_0 或 H_1 都无足够的证据. 经典方法是通过犯错误概率来作选择. 下面用 Bayes 统计决策方法来研究, 采用后验风险最小的原则, 即使后验风险达到最小的决策行动作为最优决策行动.

常见的有限行动问题的另一个类型是分类问题. 获得观测值后, 将未知参数分到几个可能的区域中去, 这与前面的多行动检验类似, 采用的准则仍是后验风险最小的原则, 即采用使后验风险达到最小的决策行动.

例 7.4.7 (续例 7.4.6) 在例 7.4.6 儿童作智力测验的例子中, 对那个孩子的智商作出如下三个假设:

$$H_1 : \theta < 90, \quad H_2 : 90 \leq \theta \leq 110, \quad H_3 : \theta > 110. \quad (7.4.13)$$

又设有三个行动 a_1, a_2, a_3 , 其中 a_i 表示接受 H_i , $i = 1, 2, 3$. 选择下列损失函数是合适的:

$$L(a_1, \theta) = \begin{cases} 0, & \theta < 90, \\ \theta - 90, & 90 \leq \theta \leq 110, \\ 2(\theta - 90), & \theta > 110, \end{cases}$$

$$L(a_2, \theta) = \begin{cases} 90 - \theta, & \theta < 90, \\ 0, & 90 \leq \theta \leq 110, \\ \theta - 110, & \theta > 110, \end{cases}$$

$$L(a_3, \theta) = \begin{cases} 2(110 - \theta), & \theta < 90, \\ 110 - \theta, & 90 \leq \theta \leq 110, \\ 0, & \theta > 110. \end{cases}$$

已知后验分布 $\pi(\theta|x)$ 是 $N(110.38, 8.32^2)$, 求多重检验问题 (7.4.13).

解 计算 a_i , $i = 1, 2, 3$ 的后验风险. 计算时将后验分布 $\pi(\theta|x)$ 由 $N(110.39, 8.32^2)$ 变换到 $N(0, 1)$, 查标准正态分布表得到

$$R(a_1|x) = E^{\theta|x}[L(a_1, \theta)] = \int_{90}^{110} (\theta - 90)\pi(\theta|x)d\theta + 2 \int_{110}^{\infty} (\theta - 90)\pi(\theta|x)d\theta \\ = 6.49 + 27.83 = 34.23,$$

$$R(a_2|x) = E^{\theta|x}[L(a_2, \theta)] = \int_{-\infty}^{90} (90 - \theta)\pi(\theta|x)d\theta + \int_{110}^{\infty} (\theta - 110)\pi(\theta|x)d\theta \\ = 0.02 + 3.53 = 3.55,$$

$$R(a_3|x) = E^{\theta|x}[L(a_3, \theta)] = \int_{-\infty}^{90} 2(110 - \theta)\pi(\theta|x)d\theta + \int_{90}^{110} (110 - \theta)\pi(\theta|x)d\theta \\ = 0.32 + 2.95 = 3.27.$$

按后验风险最小准则, 采取行动 a_3 , 即接受 H_3 , 即认为该儿童智商属于 $\theta \geq 110$.

*7.4.5 统计决策中的区间估计问题

区间估计问题, 也可以用统计决策的方法去考虑. 为简单计, 设 $C(x) = (d_1(x), d_2(x))$ 为 θ 的一个区间估计, 损失函数的一种取法为

$$L(C(x), \theta) = m_1[d_2(x) - d_1(x)] + m_2[1 - I_{C(x)}(\theta)],$$

其中 $m_1 > 0$, $m_2 > 0$ 为给定常数. 显见第一部分表示区间长度引起的损失, 第二部分表示当 θ 不属于 $C(x)$ 引起的损失. 按后验风险最小的原则, 应使

$$R(C(x)|x) = E^{\theta|x}[L(C(x), \theta)] = m_1[d_2(x) - d_1(x)] + m_2 P_{\theta|x}(\theta \notin C(x))$$

越小越好. 但是要找出最优解, 并非易事. 优化问题能够得以解决的不多.

*7.5 Minimax 准则

7.5.1 引言及定义

7.4节已经说过,一致最优的决策函数通常不存在,因此必须把标准放宽些,引进一些比一致最优准则更弱的优良性准则。途径之一是用某种方法制定优良性的综合指标,以之作为比较的标准。Bayes准则属于这一类。和 Bayes 风险 $R_\pi(\delta)$ 一样,下列定义的最大风险 $M(\delta)$ 也是一种优良性的综合指标,用它作为比较决策函数的标准,称为 Minimax 准则。因此 Minimax 准则是从综合指标考虑的另一种优良性准则。

设 δ 为一决策函数, $R(\delta, \theta)$ 为其风险函数,令

$$M(\delta) = \sup_{\theta \in \Theta} R(\delta, \theta).$$

易见 $M(\delta)$ 表示采用决策函数 δ 时所遭受的最大风险。如果在某项应用中,使这个最大风险尽可能小是很重要的,就可以制定如下准则,通常称为 Minimax 准则或极小极大准则。

定义 7.5.1 设 δ_1 和 δ_2 为同一个统计决策问题的两个决策函数,若 $M(\delta_1) < M(\delta_2)$, 则称决策函数 δ_1 优于 δ_2 。如果存在某个决策函数 δ^* , 对任何决策函数 δ 都有

$$M(\delta^*) \leq M(\delta),$$

则称 δ^* 为该统计决策问题的 Minimax 解,也称 δ^* 为 Minimax 决策函数。当统计决策问题为估计或检验时,也称 δ^* 为 Minimax 估计或 Minimax 检验。

注 7.5.1 以最大风险的大小作为评判决策函数的标准,是考虑最不利的情形,使最不利情形尽可能地好。因此 Minimax 准则是一种偏保守的准则。常有如图 7.5.1 的情形发生,其中 $M(\delta_1) < M(\delta_2)$, 故按 Minimax 准则而言, δ_1 优于 δ_2 。但是仔细观看二者的风险函数,发现对大多数 θ 而言, δ_2 优于 δ_1 , 仅当 $a < \theta < b$ 时, δ_1 优于 δ_2 。如果没有足够的先验信息说明 θ 处在 a, b 之间,就很难说 δ_1 优于 δ_2 了。因此 Bayes 学派认为,只是在人们对 θ 的先验分布很没把握的时候,作为一种替代,才使用 Minimax 解。只要对先验分布有一定把握,则宁肯采用 Bayes 准则。

在实际中常使用 Minimax 准则这种策略思想作决策。形象地说,这一准则“不求得到很多,但求不失去很多”,如在地震多发地区,重要建筑物的建筑设计按 Minimax 准则,要求在可抗八级地震的条件下,尽量减少建造费用。重要的防洪堤坝的设计也要按 Minimax 准则,在保证能抵御百年一遇洪水的条件下,尽可能地减少建设费用。参加人寿保险或财产保险,也是出于此种策略。

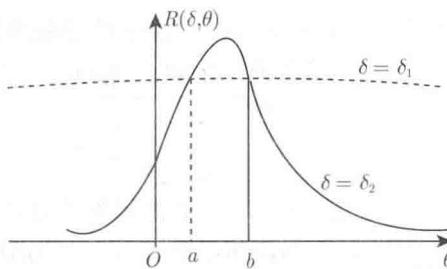


图 7.5.1 风险与 Minimax 准则

7.5.2 Minimax 解的求法

求 Minimax 解通常比较困难, 下列两个定理与其说是求 Minimax 解的方法, 不如说是验证某一特定的解为 Minimax 解的方法.

定理 7.5.1 设 δ^* 为在先验分布 $H(\theta)$ 之下的 Bayes 解, 且 δ^* 的风险函数为常数 c , 即对任何 $\theta \in \Theta$ 有 $R(\delta^*, \theta) = c$, 则 δ^* 为一个 Minimax 解.

证 采用反证法. 若不然, δ^* 不是 Minimax 解, 则存在估计量 δ 使得

$$\sup_{\theta \in \Theta} R(\delta, \theta) < \sup_{\theta \in \Theta} R(\delta^*, \theta) \equiv c. \quad (7.5.1)$$

故 $R(\delta, \theta) < c$ 对一切 $\theta \in \Theta$ 成立, 这时将两边关于 θ 的先验分布 $H(\theta)$ 求平均可得

$$R_H(\delta) = \int_{\Theta} R(\delta, \theta) dH(\theta) < c \int_{\Theta} dH(\theta) = \int_{\Theta} R(\delta^*, \theta) dH(\theta) = R_H(\delta^*),$$

即 $R_H(\delta) < R_H(\delta^*)$ (其中 $R_H(\delta)$ 为 δ 的 Bayes 风险, $R_H(\delta^*)$ 为 δ^* 的 Bayes 风险), 这与 δ^* 为 Bayes 解矛盾.

例 7.5.1 设 $X \sim b(n, \theta)$, θ 的先验分布是 $Be(a, b)$. 损失函数为 $L(d, \theta) = (\theta - d)^2$, 求 θ 的 Minimax 估计.

解 由例 7.2.9 可知 θ 的后验分布 $\pi(\theta|x)$ 是 $Be(x+a, n-x+b)$, 故 θ 的 Bayes 估计为 $\delta_{a,b}(x) = (x+a)/(n+a+b)$, 其风险函数为

$$\begin{aligned} R(\delta_{a,b}, p) &= E \left(\frac{X+a}{n+a+b} - \theta \right)^2 = E \left[\left(\frac{X - E(X)}{n+a+b} \right) + \left(\frac{E(X)+a}{n+a+b} - \theta \right) \right]^2 \\ &= D \left(\frac{X}{n+a+b} \right) + \left(\frac{a+n\theta}{n+a+b} - \theta \right)^2 \\ &= \frac{n\theta(1-\theta)}{(n+a+b)^2} + \left(\frac{a-(a+b)\theta}{n+a+b} \right)^2. \end{aligned}$$

若取 $a = b = \sqrt{n}/2$, 则上式右边等于一个常数, 即

$$R(\delta_{\sqrt{n}/2, \sqrt{n}/2}(x), \theta) = \frac{n}{4(n+\sqrt{n})^2}.$$

由于风险为常数, 由定理 7.5.1 可知

$$\delta_{\sqrt{n}/2, \sqrt{n}/2}(x) = \frac{x + \sqrt{n}/2}{n + \sqrt{n}}$$

是 θ 的 Minimax 估计.

二项分布中参数 θ 的传统的估计 $\delta_2 = \bar{X}$ 的风险函数是 $R(\delta_2, \theta) = \theta(1 - \theta)/n$. Minimax 估计 $\delta_1 = \delta_{\sqrt{n}/2, \sqrt{n}/2}(x)$ 与决策函数 δ_2 的风险函数的图形与图 7.5.1 类似, 所不同的是 δ_1 的风险函数是一条严格的水平直线. 因此, 如注 7.5.1 所述, 尽管 $M(\delta_1) < M(\delta_2)$, 但对大多数 θ 而言, δ_2 优于 δ_1 , 仅当 $a < \theta < b$ 时, δ_1 优于 δ_2 . 如果没有足够的先验信息说明 θ 以较大的概率落在 a, b 之间, 就很难说 δ_1 优于 δ_2 了. 因此, 除非有某种特殊的原因, 人们宁可用 δ_2 而不用 δ_1 .

定理 7.5.1 的使用面较窄, 因为一般很难找到一个其风险函数为常数的 Bayes 解. 下面定理的应用要广泛得多.

定理 7.5.2 设 δ_k 为一个统计决策问题在先验分布 H_k 之下的 Bayes 解, 假定 δ_k 的 Bayes 风险为 r_k , $k = 1, 2, \dots$, 且有

$$\lim_{k \rightarrow \infty} r_k = r < \infty. \quad (7.5.2)$$

又设 δ^* 为同一问题的一个决策函数, 满足条件

$$M(\delta^*) \leq r, \quad (7.5.3)$$

则 δ^* 为此统计决策问题的 Minimax 解.

证 反证法. 若不然, δ^* 不是 Minimax 解, 则必存在决策函数 δ 使得

$$M(\delta) < M(\delta^*) \leq r = \lim_{k \rightarrow \infty} r_k.$$

由式 (7.5.2) 和式 (7.5.3) 可知, 当 k 充分大时有

$$M(\delta) < r_k.$$

于是有

$$\begin{aligned} R_{H_k}(\delta) &= \int_{\Theta} R(\delta, \theta) dH_k(\theta) \leq \int_{\Theta} M(\delta) dH_k(\theta) = M(\delta) \\ &< r_k = \int_{\Theta} R(\delta_k, \theta) dH_k(\theta) = R_{H_k}(\delta_k). \end{aligned}$$

这与 δ_k 为先验分布 $H_k(\theta)$ 之下的 Bayes 解矛盾.

例 7.5.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态总体 $N(\theta, 1)$ 中抽取的简单样本, 取损失函数为 $L(d, \theta) = (d - \theta)^2$, 求 θ 的 Minimax 估计.

解 找一串 θ 的先验分布 $\{H_k\}$, H_k 为 $N(0, k^2)$, $k = 1, 2, \dots$. 由例 7.2.10 可知 θ 的后验分布是 $N(\mu_k(\mathbf{x}), \eta^2)$. 再由定理 7.4.2 可知, 平方损失下的 Bayes 解为后验分布的均值, 即

$$\delta_k(\mathbf{x}) = \mu_k(\mathbf{x}) = \frac{nk^2}{nk^2 + 1} \bar{x},$$

其风险函数

$$\begin{aligned} R(\delta_k, \theta) &= E \left(\frac{nk^2 \bar{X}}{nk^2 + 1} - \theta \right)^2 \\ &= D \left(\frac{nk^2 \bar{X}}{nk^2 + 1} \right) + \left(\frac{nk^2}{nk^2 + 1} E(\bar{X}) - \theta \right)^2 \\ &= \frac{nk^4}{(nk^2 + 1)^2} + \frac{\theta^2}{(nk^2 + 1)^2} = \frac{nk^4 + \theta^2}{(nk^2 + 1)^2}, \end{aligned}$$

其中均值 E 和方差 D 是关于“给定 θ 时 X 的分布”计算的. 故 δ_k 的 Bayes 风险是

$$r_k = R_{H_k}(\delta_k) = E^\theta \left[\frac{nk^4 + \theta^2}{(nk^2 + 1)^2} \right] = \frac{nk^4 + k^2}{(nk^2 + 1)^2} = \frac{k^2}{nk^2 + 1},$$

显然有

$$\lim_{k \rightarrow \infty} r_k = \lim_{k \rightarrow \infty} \frac{k^2}{nk^2 + 1} = \frac{1}{n} = r.$$

若取 $\delta^* = \bar{X}$, 其风险函数 $R(\delta^*, \theta) = E(\bar{X} - \theta)^2 = 1/n \leq r$, 故由定理 7.5.2 可知, $\delta^* = \bar{X}$ 为 θ 的 Minimax 估计.

*7.6 同变估计及可容许性

7.6.1 同变估计及例子

前两节讨论的 Bayes 准则和 Minimax 准则都是用某种方法制定的优良性的综合指标 $R_H(\delta)$ 和 $M(\delta)$ 等, 以其作为比较的准则. 另一类制定优良性准则的方法是对决策函数提出合理的要求, 缩小所考虑的决策函数的范围, 从中找一致最优者, 如无偏性准则就是这样一种要求, UMVUE 就是属于由这类准则获得的最优估计. 本节介绍另一个这类准则, 称为同变原理. 请看下例.

例如, 设要估计某物体的重量 a , 将它放在某一架天平上称量 n 次, 得样本 X_1, \dots, X_n , 假定它们相互独立, 同服从正态分布 $N(a, \sigma^2)$. 用 $\delta(X_1, \dots, X_n)$ 去估计 a ; 如果用另一架不精确的天平去称同样的物体, 称出来的数字系统地偏大 c , 即得到一批数据 X'_1, \dots, X'_n , $X'_i = X_i + c$, $c > 0$, $i = 1, 2, \dots, n$, 用这批数据去估计

物重 a , 即用 $\delta(x'_1, \dots, x'_n) = \delta(x_1 + c, \dots, x_n + c)$ 估计 a 是系统地偏高了, 它的平均值应当是 $a + c$, 因此应当用 $\delta(x_1 + c, \dots, x_n + c) - c$ 估计 a 的值. 提出合理要求: 估计值不应与天平的系统偏差有关, 应消除这种偏差, 即要求所用的估计量 δ 满足条件: 对一切实数 c , 有

$$\delta(x'_1, \dots, x'_n) = \delta(x_1 + c, \dots, x_n + c) = \delta(x_1, \dots, x_n) + c, \quad (7.6.1)$$

满足式 (7.6.1) 要求的估计量称为在平移变换下的位置同变估计 (location equivariant estimation).

若考虑另一种情形, 如测量两地间的距离 n 次, 得样本 X_1, \dots, X_n , 假定它们相互独立, 同服从正态分布 $N(b, \sigma^2)$. 用原来长度单位 (如 m) 得到距离 b 的估计值为 $\delta(x_1, \dots, x_n)$, 若把测量单位改变 (如将 m 改为 cm), 则得到数据为 X'_1, \dots, X'_n , $X'_i = cX_i, i = 1, 2, \dots, n, 1 < c < \infty$. 用 $\delta(x'_1, \dots, x'_n) = \delta(cx_1, \dots, cx_n)$ 去估计 b 显然偏高, 它的平均值应当是 cb . 因此, 应当用 $\delta(x'_1, \dots, x'_n)/c$ 作为 b 的估计才与单位无关. 提出合理要求: 估计 b 应当与单位无关, 即要求所用的估计量 δ 满足条件

$$\delta(x'_1, \dots, x'_n) = \delta(cx_1, \dots, cx_n) = c\delta(x_1, \dots, x_n), \quad c > 0. \quad (7.6.2)$$

满足式 (7.6.2) 要求的估计量称为在刻度变换下的刻度同变估计 (scale equivariant estimation).

综上所述, 可以给出同变估计的一个描述性定义, 其严格定义见文献 [13] 第 3 章或文献 [1] 第 163 页.

定义 7.6.1 设 $\delta(x_1, \dots, x_n)$ 是参数 θ 的一个估计量. 假如对样本作了某种变换, 估计量 $\delta(x_1, \dots, x_n)$ 仍保持此变换的统计特性, 则称 $\delta(x_1, \dots, x_n)$ 是在该变换下 θ 的同变估计. 在变换是明确无误的情形下, 可简称 $\delta(x_1, \dots, x_n)$ 是参数 θ 的同变估计.

例如, 若对样本作平移变换, 则位置参数 θ 的同变估计满足性质 (7.6.1); 若对样本作刻度变换, 则刻度参数 θ 的同变估计满足性质 (7.6.2).

引入损失函数后可以在同变估计类中找最优同变估计. 设 $R(\delta, \theta)$ 为同变估计 δ 的风险函数, 则有如下定义.

定义 7.6.2 设 $\delta^*(x_1, \dots, x_n)$ 是某特定变换下 θ 的一个同变估计, 如果对此种变换下的任一同变估计 $\delta(x_1, \dots, x_n)$, 有

$$R(\delta^*, \theta) \leq R(\delta, \theta), \quad \text{一切 } \theta \in \Theta,$$

则称 $\delta^*(x_1, \dots, x_n)$ 是此种变换下 θ 的最优同变估计.

下面来看一个最优同变估计的例子. 为此需要下述引理.

引理 7.6.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从正态总体 $N(a, \sigma^2)$ 中抽取的简单样本, 而 $f(X_1, \dots, X_n)$ 满足条件: $f(X_1 + c, \dots, X_n + c) = f(X_1, \dots, X_n)$, 对任何实数 c , 则 \bar{X} 与 $f(X_1, \dots, X_n)$ 独立.

证 这是第 2 章中的最后一个习题. 读者也可查看文献 [2] 第 235 页引理 5.1 的证明.

例 7.6.1 设 X_1, \dots, X_n 为从正态总体 $N(a, \sigma^2)$ 中抽取的简单样本, 则样本均值 \bar{X} 是 a 的平移变换下最优同变估计.

证 设 $\delta(\mathbf{X}) = \delta(X_1, \dots, X_n)$ 为 a 的任一位置同变估计 (即满足式 (7.6.1)), 显然, \bar{X} 也是 a 的位置同变估计. 记

$$\delta_0(\mathbf{X}) = \delta(\mathbf{X}) - \bar{X},$$

则 δ_0 满足: $\delta_0(X_1 + c, \dots, X_n + c) = \delta_0(X_1, \dots, X_n)$. 由引理 7.6.1, 可知 \bar{X} 与 $\delta_0(\mathbf{X})$ 独立. 若取损失函数为平方损失, 则有

$$R(\delta(\mathbf{X}), a) = E_a[\delta(\mathbf{X}) - a]^2 = E_a[(\delta(\mathbf{X}) - \bar{X}) + (\bar{X} - a)]^2.$$

由于 \bar{X} 与 $\delta_0(\mathbf{X})$ 独立, 故有

$$\begin{aligned} R(\delta(\mathbf{X}), a) &= E_a(\bar{X} - a)^2 + E_a[\delta_0^2(\mathbf{X})] + 2E_a(\bar{X} - a) \cdot E_a[\delta_0(\mathbf{X})] \\ &= E_a(\bar{X} - a)^2 + E_a[\delta_0^2(\mathbf{X})] \\ &\geq E_a(\bar{X} - a)^2 = R(\bar{X}, a), \quad \text{对一切 } a. \end{aligned}$$

这就证明了 \bar{X} 是 a 的一切位置同变估计中风险一致最小者, 因而 \bar{X} 是 a 的平移变换下最优同变估计.

7.6.2 决策函数的可容许性

1. 可容许性的概念及判别方法

在统计决策问题中按风险函数越小越好的原则找一致最优的决策函数常常难以实现, 于是降低要求, 寻找可容许决策函数. 容许性这个概念本身并不是一个优良性准则, 但在很大程度上可以说, 它是任何优良决策函数所应具备的条件. 下面给出它的定义.

定义 7.6.3 记 $R(\delta, \theta)$ 为决策函数 δ 的风险函数. 若存在决策函数 δ_1 使得

- (1) 对任何 $\theta \in \Theta$, 有 $R(\delta_1, \theta) \leq R(\delta, \theta)$;
- (2) 至少存在一个 $\theta_0 \in \Theta$, 使 $R(\delta_1, \theta_0) < R(\delta, \theta_0)$,

则称 δ_1 一致地优于 δ , 而称 δ 是不可容许的决策函数. 反之, 若不存在一致地优于 δ 的决策函数, 则称 δ 是可容许的决策函数 (admissible decision rule).

确定一个决策函数是可容许或不可容许的并不容易, 极难办到. 有时可限制在一定范围内讨论可容许性. 例如, 限制在线性估计类中讨论可容许性, 是最有兴趣也是文献中研究最多的情形. 这类决策函数一般都有某种优良性. 例如, 决策函数是某个统计决策问题的 Bayes 解或 Minimax 解, 或者是由直观方法产生的优良解. 这方面的研究工作难度很大, 因为缺乏一般的有效方法. 下面的几个定理分别讨论 Bayes 解和 Minimax 解的可容许性.

定理 7.6.1 在 Bayes 决策问题中, 设 δ_H 为先验分布 $H(\theta)$ 下的 Bayes 解, 若 $H(\theta)$ 和 δ_H 满足下列条件:

- (1) 先验分布 $H(\theta)$ 对参数空间 Θ 的任何非空开子集有正概率;
- (2) δ_H 的 Bayes 风险 $R_H(\delta_H) < \infty$;
- (3) 对任何决策函数 δ , $R(\delta, \theta)$ 为 θ 的连续函数,

则 δ_H 为可容的.

注 7.6.1 这个定理虽然简单, 但有很大的实用价值. 首先它说明一大批 Bayes 估计是可容许的, 如由共轭先验分布得到的 Bayes 估计都是可容许的. 其次, 这个定理还指出: 要证明“一个估计是可容许的”, 只要能找到满足定理 7.6.1 的先验分布, 使得在此先验分布下的 Bayes 估计就是所讨论的估计即可.

证 若不然, 必存在 $\delta^*(x)$, 使得

$$R(\delta^*, \theta) \leq R(\delta_H, \theta), \quad \text{一切 } \theta \in \Theta,$$

且至少对某个 $\theta_1 \in \Theta$ 有严格不等式, 即

$$R(\delta^*, \theta_1) < R(\delta_H, \theta_1).$$

由 $R(\delta_H, \theta)$ 的连续性, 知必存在函数 $\varepsilon > 0$ 及 θ_1 的 ε 邻域 $\mathcal{S}_\varepsilon(\theta_1)$, 使得

$$R(\delta^*, \theta) < R(\delta_H, \theta) - \varepsilon, \quad \text{一切 } \theta \in \mathcal{S}_\varepsilon(\theta_1).$$

于是对 δ^* 的 Bayes 风险有

$$\begin{aligned} R_H(\delta^*) &= \int_{\mathcal{S}_\varepsilon(\theta_1)} R(\delta^*, \theta) dH(\theta) + \int_{\Theta - \mathcal{S}_\varepsilon(\theta_1)} R(\delta^*, \theta) dH(\theta) \\ &\leq \int_{\mathcal{S}_\varepsilon(\theta_1)} [R(\delta_H, \theta) - \varepsilon] dH(\theta) + \int_{\bar{\mathcal{S}}_\varepsilon(\theta_1)} R(\delta_H, \theta) dH(\theta) \\ &= R_H(\delta_H) - \varepsilon P(\theta \in \mathcal{S}_\varepsilon(\theta_1)). \end{aligned}$$

由假设 (1) 可知 $P(\theta \in \mathcal{S}_\varepsilon(\theta_1)) > 0$, 从而有 $R_H(\delta^*) < R_H(\delta_H)$, 这与 δ_H 为 θ 的 Bayes 解矛盾.

定理 7.6.2 在给定的 Bayes 决策问题中, 若在给定的先验分布 $H(\theta)$ 下的 Bayes 估计 δ_H 是唯一的, 则它也是可容许的.

证 反证. 若 δ_H 不是可容许的, 则必存在另一个估计 $\delta^* \neq \delta_H$, 使得

$$R(\delta^*, \theta) \leq R(\delta_H, \theta), \quad \text{一切 } \theta \in \Theta,$$

且严格不等式至少对某个 $\theta \in \Theta$ 成立. 上式两边关于先验分布 $H(\theta)$ 求期望得到

$$R_H(\delta^*) = \int_{\Theta} R(\delta^*, \theta) dH(\theta) \leq \int_{\Theta} R(\delta_H, \theta) dH(\theta) = R_H(\delta_H).$$

故 $\delta^* = \delta_H$ 与 δ_H 的唯一性矛盾.

定理 7.6.3 在统计决策问题中, 假如 δ_0 为参数 θ 的唯一 Minimax 估计, 则 δ_0 也是 θ 的可容许估计.

证 反证法. 若不然, 则必存在另一个估计 $\delta^* \neq \delta_0$, 使得

$$R(\delta^*, \theta) \leq R(\delta_0, \theta), \quad \text{一切 } \theta \in \Theta,$$

且严格不等式至少对某个 $\theta \in \Theta$ 成立, 因此有

$$\sup_{\theta \in \Theta} R(\delta^*, \theta) \leq \sup_{\theta \in \Theta} R(\delta_0, \theta),$$

从而 δ^* 也是 θ 的 Minimax 估计, 与唯一性矛盾.

例 7.6.2 设随机变量 $X \sim b(n, \theta)$, $\theta \sim \text{Be}(a, b)$, 损失函数 $L(d, \theta) = (d - \theta)^2$.

证明参数 θ 的 Minimax 估计 $\delta_{\sqrt{n}/2, \sqrt{n}/2}(x) = (x + \sqrt{n}/2)/(n + \sqrt{n}/2)$ 是可容许的.

证 由例 7.2.9 可知给定 x , θ 的后验分布是 $\text{Be}(a+x, b+n-x)$. 易知平方损失下的 Bayes 估计是后验均值, 即

$$\delta_{a,b}(x) = \frac{x+a}{a+b+n},$$

当取 $a = b = \sqrt{n}/2$ 时, 风险函数为常数, 故

$$\delta_{\sqrt{n}/2, \sqrt{n}/2}(x) = \frac{x + \sqrt{n}/2}{n + \sqrt{n}/2} \tag{7.6.3}$$

是 θ 的 Minimax 估计.

易验证 θ 的先验分布 $\text{Be}(a, b)$ 在 $0 < \theta < 1$ 内有处处大于 0 的密度, 故定理 7.6.1 的条件 (1) 成立. 显然定理 7.6.1 的条件 (2) 也成立. 至于条件 (3), 只要注意到任一估计量 δ 的风险函数

$$R(\delta, \theta) = E(\delta - \theta)^2 = \sum_{i=0}^n (\delta(i) - \theta)^2 \binom{n}{i} \theta^i (1-\theta)^{n-i}$$

是 θ 的连续函数, 故定理 7.6.1 的条件 (3) 也成立. 因此, 由定理 7.6.1 可知 θ 的 Minimax 估计 (7.6.3) 是可容许的.

例 7.6.3 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从正态分布 $N(\theta, 1)$ 中抽取的简单样本, 其中 $-\infty < \theta < \infty$. 令损失函数是 $L(d, \theta) = (d - \theta)^2$, $0 < c < 1$ 为常数, 则 $c\bar{X}$ 是 θ 的可容许估计.

证 事实上, 若取 θ 的先验分布 $H(\theta)$ 为 $N(0, \tau^2)$, 其中 τ^2 为已知常数, 取 $c = n\tau^2/(1 + n\tau^2)$, 则由例 7.2.10 可知 θ 的 Bayes 估计是后验均值, 即

$$\delta_H(\mathbf{X}) = \frac{n\tau^2}{1 + n\tau^2} \bar{X} = c\bar{X}.$$

不难验证定理 7.6.1 的条件 (1)-(3) 皆成立, 故 $\delta_H(\mathbf{X})$ 是 θ 的可容许估计.

此处不能用定理 7.6.1 的方法直接证明 θ 的常见估计 $\bar{X} = \mathbf{X}/n$ 的可容许性. 但 \bar{X} 确是 θ 的可容许估计, 对此证明感兴趣的读者可查看文献 [2] 第 240-241 页中被称之为“C-R 不等式法”的证明方法.

2. James-Stein 估计

设 X_1, \dots, X_n 为从 p 维正态分布 $N(\boldsymbol{\theta}, \mathbf{I}_p)$ 中抽取的简单样本, 取损失函数 $L(\boldsymbol{\theta}, \mathbf{d}) = \|\mathbf{d} - \boldsymbol{\theta}\|^2$, 其中 $\boldsymbol{\theta}, \mathbf{d}$ 皆为 p 维向量. $\boldsymbol{\theta}$ 的通常估计 $\bar{\mathbf{X}} = \sum_{i=1}^n \mathbf{X}_i/n$ 是否为可容许的呢? 在 $p = 1, 2$ 时, 回答是肯定的, 自然猜想对 $p > 2$ 时 $\bar{\mathbf{X}}$ 也是可容许的. C. Stein 在 1955 年证明: 当 $p \geq 3$, 样本均值向量 $\bar{\mathbf{X}}$ 是正态均值向量 $\boldsymbol{\theta}$ 的不可容许估计. 这一结果曾在统计学界引起轰动. 1960 年, James 和 Stein 给出了比 $\bar{\mathbf{X}}$ 更优的估计

$$\boldsymbol{\theta}_{JS} = \left(1 - \frac{p-2}{\|\bar{\mathbf{X}}\|^2}\right) \bar{\mathbf{X}}.$$

这个估计称为 James-Stein 估计.

习 题 7

1. 设参数 θ 的先验分布为贝塔 (Beta) 分布 $Be(\alpha, \beta)$, 若从先验信息中获得其均值和方差分别为 $1/3$ 与 $1/45$, 试确定该先验分布.
2. 设 θ 的先验分布是伽马分布, 其均值为 10, 方差为 5, 试确定 θ 的先验分布.
3. 设 θ 是一批产品的不合格率, 已知它不是 0.1 就是 0.2, 且其先验分布为 $\pi(0.1) = 0.7$, $\pi(0.2) = 0.3$. 假如从这批产品中随机抽取 8 个进行检查, 发现有 2 个不合格, 求 θ 的后验分布.
4. 设一卷磁带上的缺陷数服从 poission 分布 $P(\lambda)$, 其中 λ 可取 1.0 和 1.5 中的一个, 又设 λ 的先验分布为 $\pi(1.0) = 0.4$, $\pi(1.5) = 0.6$. 假如检查一卷磁带发现 3 个缺陷, 求 λ 的后验分布.

5. 考虑一个试验, 对给定的 θ , 试验结果 X 有如下密度函数:

$$p(x|\theta) = 2x/\theta^2, \quad 0 < x < \theta < 1.$$

- (1) 假如 θ 的先验分布是 $(0, 1)$ 上的均匀分布, 试求 θ 的后验分布;
 (2) 假如 θ 的先验密度是 $\pi(\theta) = 3\theta^2$, $0 < \theta < 1$, 试求 θ 的后验分布.
 6. 对下列每个分布中的未知参数使用 Fisher 信息量决定 Jeffreys 无信息先验:
 (1) Poission 分布 $P(\lambda)$;
 (2) 二项分布 $b(n, \theta)$, 其中 n 已知;
 (3) Gamma 分布 $\Gamma(\alpha, \lambda)$, 其中 α 已知.

7. 设随机变量 X 服从负二项分布, 其概率分布为

$$f(x|p) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \quad x = k, k+1, \dots$$

- 证明其成功概率 p 的共轭先验分布族为 Beta 分布族.
 8. 设随机变量 X 服从参数为 λ 的指数分布 $\text{Exp}(\lambda)$, 证明指数分布中参数 λ 的共轭先验分布族为 Gamma 分布族.
 9. 设随机变量 X 服从指型分布, 其密度函数为 $f(x|\theta) = \exp\{a(\theta)b(x) + c(\theta) + d(x)\}$, 其中 $a(\theta), c(\theta)$ 是 θ 的函数, $b(x), d(x)$ 是 x 的函数. 证明先验分布 $h(\theta) = Ae^{k_1 a(\theta) + k_2 c(\theta)}$ 是参数 θ 的共轭先验分布, 其中 A 为常数, k_1, k_2 是与 θ 无关的常数.
 10. 设随机变量 X 的密度函数为

$$f(x|\lambda) = \begin{cases} \lambda^{-1} e^{-x/\lambda}, & 0 < x < \infty, \\ 0, & x \leq 0. \end{cases}$$

- 证明参数 λ 的共轭先验分布族为逆 Gamma 分布族.
 11. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从均匀分布 $U(0, \theta)$ 中抽取的简单样本, 又假设 θ 的先验分布是 Pareto 分布, 其密度函数为

$$\pi(\theta) = \begin{cases} \frac{\alpha \theta_0^\alpha}{\theta^{\alpha+1}}, & \theta > \theta_0, \\ 0, & \theta \leq \theta_0, \end{cases}$$

- 其中 $\theta_0 > 0$, $\alpha > 0$. 证明 Pareto 分布是 $U(0, \theta)$ 中端点 θ 的共轭先验分布.
 12. 从一批产品中抽检 100 个, 发现 3 个不合格品, 假定该产品不合格率 θ 的先验分布为 Beta 分布 $\text{Be}(2, 200)$, 求 θ 的后验分布.
 13. 设 X_1, \dots, X_n 为从正态总体 $N(\theta, 2^2)$ 中抽取随机样本, 又设 θ 的先验分布为正态分布.
 (1) 若样本容量 $n = 100$, 证明不管先验标准差为多少, 后验标准差一定小于 $1/5$.
 (2) 若 θ 的先验分布的标准差为 1, 要使后验方差不超过 0.1, 最少要抽取样本容量多大的样本?

14. 设随机变量 X 服从几何分布 $p(x|\theta) = P(X=x|\theta) = \theta(1-\theta)^{x-1}$, $x=1, 2, \dots$, 其中参数 θ 的先验分布为均匀分布 $U(0, 1)$.
- 若只对 X 作一次观察, 观察值为 3, 求 θ 的后验期望估计;
 - 若对 X 作三次观察, 观察值为 3, 2, 5, 求 θ 的后验期望估计.
15. 对正态总体 $N(\theta, 1)$ 作三次观察, 获得样本的具体观察值为 2, 4, 3. 若 θ 的先验分布为正态分布 $N(3, 1)$, 求 θ 的可信系数为 0.95 的可信区间.
16. 接到船运来的一大批零件, 从中抽检 5 件, 看有多少次品, 假设零件中的次品数 X 服从二项分布 $b(5, \theta)$. 从以往各批的情况中已知 θ 的先验分布为贝塔分布 $Be(1, 9)$, 若观测值 $X=0$, 求 θ 的可信系数为 95% 且长度最短的可信区间.
17. 设 X 为从总体中抽取的样本, 总体密度 $f(x|\theta)$ 和 θ 的先验密度 $\pi(\theta)$ 分别为

$$f(x|\theta) = \begin{cases} e^{-(x-\theta)}, & x > \theta, \\ 0, & x \leq \theta, \end{cases} \quad \pi(\theta) = \begin{cases} e^{-\theta}, & \theta > 0, \\ 0, & \theta \leq 0. \end{cases}$$

求检验问题 $H_0: \theta \leq 1 \leftrightarrow H_1: \theta > 1$.

18. 设 X_1, \dots, X_m i.i.d. $\sim N(a, 1)$, Y_1, \dots, Y_n i.i.d. $\sim N(b, 1)$, 其中 a, b 为参数, 又样本 $X_1, \dots, X_m, Y_1, \dots, Y_n$ 独立. 令 $\theta = (a, b)$ 的先验分布为: a, b 独立, $a \sim N(\mu_1, \tau_1^2)$, $b \sim N(\mu_2, \tau_2^2)$. 求检验问题 $H_0: a - b \leq 0 \leftrightarrow H_1: a - b > 0$.
19. 设 X_1, \dots, X_n 是来自正态总体 $N(\mu, 1)$ 的一个样本, 其中未知参数 μ 的先验分布是 $N(0, 1)$, 在平方损失函数即 $L(\mu, \hat{\mu}) = (\mu - \hat{\mu})^2$ 下求 μ 的 Bayes 估计.
20. 设 X_1, \dots, X_n 是来自参数为 θ 的几何分布 (参见题 14) 的一个样本, 假如未知参数 θ 的先验分布为 Beta 分布 $Be(\alpha, \beta)$. 在平方损失函数下求 θ 的 Bayes 估计.
21. 设 X_1, \dots, X_n 为抽自参数为 θ 的 Poisson 分布的一个样本, 假如未知参数 θ 的先验分布是伽马分布, 其密度为

$$h(\theta) = \begin{cases} \theta e^{-\theta}, & \theta > 0, \\ 0, & \theta \leq 0. \end{cases}$$

在平方损失函数下, 求 θ 的 Bayes 估计.

22. 某产品的寿命服从指数分布 $Exp(\theta)$. 对 n 个这种产品进行寿命试验, 获得了一个样本 X_1, \dots, X_n , 假如未知参数 θ 的先验分布为 Gamma 分布 $\Gamma(\beta, \alpha)$. 在平方损失下, 试求 θ 和 $1/\theta$ 的 Bayes 估计.
23. 设 X_1, \dots, X_n 是从均匀分布 $\{U(0, \theta), \theta > 0\}$ 中抽取的简单样本. 设 θ 的先验分布为 $U(0, a)$, $a > 0$ 已知. 求平方损失下 θ 的 Bayes 估计.
24. 设 X_1, \dots, X_n 是来自正态总体 $N(0, \tau)$ 的一个样本, 其中方差 τ 的先验分布为逆 Gamma 分布, 即

$$h(\tau) = \begin{cases} \frac{\alpha^\beta}{\Gamma(\beta)} \tau^{-(\beta+1)} e^{-\frac{\alpha}{\tau}}, & \tau > 0, \\ 0, & \tau \leq 0. \end{cases}$$

在加权平方损失函数 $L(\tau, \hat{\tau}) = (\tau - \hat{\tau})^2 / \tau^2$ 下, 求 τ 的 Bayes 估计.

25. 设 $X \sim b(n, \theta)$, $\theta \sim \text{Be}(\alpha, \beta)$, 在损失函数

$$L(a, \theta) = \frac{(\theta - a)^2}{\theta(1 - \theta)}$$

下, 求 θ 的 Bayes 估计.

26. 接到船运来的一大批零件, 从中抽检 5 件. 假设其中不合格品数 $X \sim b(5, \theta)$, 又从以往各批的先验信息中确定 θ 的先验分布为 $\text{Be}(1, 9)$. 若观察值为 $x = 0$, 在下列损失函数下求检验问题:

$$H_0 : 0 \leq \theta \leq 0.15 \leftrightarrow H_1 : \theta > 0.15 .$$

设行动 a_i 表示接受 H_i , $i = 0, 1$.

$$(1) L(a_0, \theta) = \begin{cases} 0, & \theta \leq 0.15, \\ 1, & \theta > 0.15, \end{cases} \quad L(a_1, \theta) = \begin{cases} 2, & \theta \leq 0.15, \\ 0, & \theta > 0.15. \end{cases}$$

$$(2) L(a_0, \theta) = \begin{cases} 0, & \theta \leq 0.15, \\ 1, & \theta > 0.15, \end{cases} \quad L(a_1, \theta) = \begin{cases} 0.15 - \theta, & \theta \leq 0.15, \\ 0, & \theta > 0.15. \end{cases}$$

27. 在上题中若取损失函数为 0-1 损失, 其他不变, 考虑检验问题

$$H_0 : 0 \leq \theta \leq 0.15 \leftrightarrow H_1 : \theta > 0.15 .$$

28. 设随机变量 X 服从二项分布 $b(n, \theta)$, $0 < \theta < 1$, 证明 $d(x) = x/n$ 在损失函数 $L(d, \theta) = (d - \theta)^2 / [\theta(1 - \theta)]$ 下是 θ 的极小极大估计.(提示: 先验分布取均匀分布)

29. 设 X_1, \dots, X_n 为自正态总体 $N(0, \sigma^2)$ 中抽取的 i.i.d. 样本, 令 $\tau = \sigma^2$ 的先验分布为无信息先验, 即 $\pi(\tau) = (1/\tau)I_{(0, \infty)}(\tau)$, 取损失函数为 $L(d, \tau) = (d - \tau)^2 / \tau^2$, 求 τ 的贝叶斯估计 $\hat{\tau}_B$, 并证明 $\hat{\tau}_B$ 是 τ 的 Minimax 估计.

30. 设 X_1, \dots, X_n i.i.d. $\sim N(\theta, 1)$ 要估计 θ . 令损失函数为平方损失, 取估计量为 $\hat{\theta}_n = c_1 X_1 + \dots + c_n X_n$. 证明若 $c_1 + \dots + c_n = 1$, 则除非 $c_1 = \dots = c_n = 1/n$, $\hat{\theta}_n$ 是不可容许的.

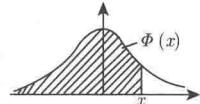
参 考 文 献

- [1] 陈希孺. 数理统计引论. 北京: 科学出版社, 1981, 1998.
- [2] 陈希孺, 倪国熙. 数理统计学教程. 上海: 上海科学技术出版社, 1988.
- [3] 克拉美 H. 统计学数学方法. 魏宗舒, 等译. 上海: 上海科学技术出版社, 1966.
- [4] 莫诗松, 王静龙. 数理统计. 上海: 华东师范大学出版社, 1990.
- [5] 陈希孺. 数理统计学简史. 长沙: 湖南教育出版社, 2002.
- [6] 陈家鼎, 孙山泽, 李东风. 数理统计学讲义. 北京: 高等教育出版社, 1993.
- [7] 傅权, 胡蓓华. 基本统计方法教程. 上海: 华东师范大学出版社, 1989.
- [8] 莫诗松, 王静龙, 潘晓龙. 高等数理统计. 北京: 高等教育出版社, 1998.
- [9] 中华人民共和国国家标准 GB4882—85. 数据的统计处理和解释, 正态性检验. 北京: 中国标准出版社, 1985.
- [10] Bickel P J, Doksum K A. Mathematical Statistics-Basic Ideas and Selected Topics. San Francisco: Holden-Day, 1977.
中译本, 数理统计——基本概念及专题. 李泽慧, 等译, 陈希孺校. 兰州: 兰州大学出版社, 1991.
- [11] Mood A M, Graybill F A. Introduction to the Theory of Statistics. 3rd ed. New York: McGraw-Hill, 1974.
中译本, 统计学导论. 史定华译, 汤旦林校. 北京: 科学出版社, 1982.
- [12] Lehmann E L. Testing Statistical Hypotheses. 2nd ed. New York: John Wiley & Sons, 1986.
- [13] Lehmann E L, George C. Theory of Point Estimation. 2nd ed. New York: Springer-Verlag, 1998.
中译本, 点估计理论. 郑忠国, 等译. 北京: 中国统计出版社, 2005.
- [14] Box G E P, Tiao G. Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, MA, 1973.
- [15] 莫诗松. 贝叶斯统计. 北京: 中国统计出版社, 1999.
- [16] Berger J O. Statistical Decision Theory and Bayesian Analysis. 2nd ed. New York: Springer-Verlag, 1985.
中译本, 统计决策论及贝叶斯分析. 贾乃光译, 吴喜之校. 北京: 中国统计出版社, 1998.
- [17] Shao Jun. Mathematical Statistics. 2nd ed. New York: Springer, 2003.
- [18] 陈希孺, 柴根象. 非参数统计教程. 上海, 华东师范大学出版社, 1993.
- [19] Parzen E. On the estimation of probability density function and mode. Ann. Math. Statist., 1962, 33: 1065-1076.
- [20] Prakasa Rao, B L S. Nonparametric Functional Estimation. New York: Academic Press, 1983.
- [21] Serfling R J. Approximation Theorems of Mathematical Statistics. New York: John Wiley & Sons, 1980.

附录

附表 1 标准正态分布表

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

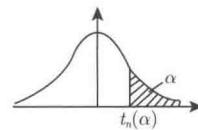


x	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.	0.9987	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

注: 表中末行为函数值 $\Phi(3.0), \Phi(3.1), \dots, \Phi(3.9)$.

附表 2 t 分布表

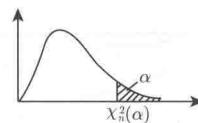
$$P(t_n > t_n(\alpha)) = \alpha$$



$n \backslash \alpha$	0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.654	1.960	2.326	2.576

附表 3 χ^2 分布表

$$P(\chi_n^2 > \chi_n^2(\alpha)) = \alpha$$



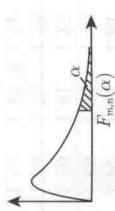
$n \backslash \alpha$	0.995	0.99	0.975	0.95	0.90	0.75
1	—	—	0.001	0.004	0.016	0.102
2	0.010	0.020	0.051	0.103	0.211	0.575
3	0.072	0.115	0.216	0.352	0.584	1.213
4	0.207	0.297	0.484	0.711	1.064	1.923
5	0.412	0.554	0.831	1.145	1.610	2.675
6	0.676	0.872	1.237	1.635	2.204	3.455
7	0.989	1.239	1.690	2.167	2.833	4.255
8	1.344	1.646	2.180	2.733	3.490	5.071
9	1.735	2.088	2.700	3.325	4.168	5.899
10	2.156	2.558	3.247	3.940	4.865	6.737
11	2.603	3.053	3.816	4.575	5.578	7.584
12	3.074	3.571	4.404	5.226	6.304	8.438
13	3.565	4.107	5.009	5.892	7.042	9.299
14	4.075	4.660	5.629	6.571	7.790	10.165
15	4.601	5.229	6.262	7.261	8.547	11.037
16	5.142	5.812	6.908	7.962	9.312	11.912
17	5.697	6.408	7.564	9.672	10.085	12.792
18	6.265	7.015	8.231	9.390	10.865	13.675
19	6.844	7.633	8.907	10.117	11.651	14.562
20	7.434	8.260	9.591	10.851	12.443	15.452
21	8.034	8.897	10.283	11.591	13.240	16.344
22	8.643	9.542	10.982	12.338	14.042	17.240
23	9.260	10.196	11.689	13.091	14.848	18.137
24	9.886	10.856	12.401	13.848	15.659	19.037
25	10.520	11.524	13.120	14.611	16.473	19.939
26	11.160	12.198	13.844	15.379	17.292	20.843
27	11.808	12.879	14.573	16.151	18.114	21.749
28	12.461	13.565	15.308	16.928	18.939	22.657
29	13.121	14.257	16.047	17.708	19.768	23.567
30	13.787	14.954	16.791	18.493	20.599	24.478
35	17.192	18.509	20.569	22.465	24.797	29.054
40	20.707	22.164	24.433	26.509	29.051	33.660
45	24.311	25.901	28.366	30.612	33.350	38.291

续表

$n \backslash \alpha$	0.50	0.25	0.10	0.05	0.025	0.01	0.005
1	0.455	1.323	2.706	3.841	5.024	6.635	7.879
2	1.386	2.773	4.605	5.991	7.378	9.210	10.597
3	2.366	4.108	6.251	7.815	9.348	11.345	12.838
4	3.357	5.385	7.779	9.488	11.143	13.277	14.860
5	4.351	6.626	9.236	11.071	12.833	15.086	16.750
6	5.348	7.841	10.645	12.592	14.449	16.812	18.548
7	6.346	9.037	12.017	14.067	16.013	18.475	20.278
8	7.344	10.219	13.362	15.507	17.535	20.090	21.955
9	8.343	11.389	14.684	16.919	19.023	21.666	23.589
10	9.342	12.549	15.987	18.307	20.483	23.209	25.188
11	10.341	13.701	17.275	19.675	21.920	24.725	26.757
12	11.340	14.845	18.549	21.026	23.337	26.217	28.299
13	12.340	15.984	19.812	22.362	24.736	27.688	29.819
14	13.339	17.117	21.064	23.685	26.119	29.141	31.319
15	14.339	18.245	22.307	24.996	27.488	30.578	32.801
16	15.338	19.369	23.542	26.296	28.845	32.000	34.267
17	16.338	20.489	24.769	27.587	30.191	33.409	35.718
18	17.338	21.605	25.989	28.869	31.526	34.805	37.156
19	18.338	22.718	27.204	30.144	32.852	36.191	38.582
20	19.337	23.828	28.412	31.410	34.170	37.566	39.997
21	20.337	24.935	29.615	32.671	35.479	38.932	41.401
22	21.337	26.039	30.813	33.924	36.781	40.289	42.796
23	22.337	27.141	32.007	35.172	38.076	41.638	44.181
24	23.337	28.241	33.196	36.415	39.364	42.980	45.559
25	24.337	29.339	34.382	37.652	40.646	44.314	46.928
26	25.336	30.435	35.563	38.885	41.923	45.642	48.290
27	26.336	31.528	36.741	40.113	43.194	46.963	49.645
28	27.336	32.620	37.916	41.337	44.461	48.278	50.993
29	28.336	33.711	39.087	42.557	45.722	49.588	52.336
30	29.336	34.800	40.256	43.773	46.979	50.892	53.672
35	34.336	40.223	46.059	49.802	53.203	57.342	60.275
40	39.335	45.616	51.805	55.758	59.342	63.691	66.766
45	44.335	50.985	57.505	61.656	65.410	69.957	73.166

附表 4 F 分布表

· 333 ·



$$P\{F_{m,n} > F_{m,n}(\alpha)\} = \alpha$$

 $\alpha = 0.10$

m	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.48	9.49	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.16	5.15	5.15	5.14	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.78	4.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48

续表

		$\alpha = 0.10$																	
$n \setminus m$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00
		$\alpha = 0.05$																	
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.47	19.48	19.49	19.50	
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71

续表

附表 4 F 分布表

		$\alpha = 0.05$																	
$m \setminus n$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00
		$\alpha = 0.025$																	
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.47	39.48	39.49	39.50	
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.20	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33		
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13	
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00

续表

$\alpha = 0.025$																				
n	m	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97		
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94	
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91	
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88	
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85	
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83	
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81	
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79	
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64	
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48	
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31	
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00	
$\alpha = 0.01$																				
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366	
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.49	99.49	99.50	
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13	
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46	
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02	
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88	
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65	
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86	
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31	
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91		
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60	
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36	
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17	
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00	
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87	
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75	
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65	
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57	

附表 4 F 分布表

表
綱

$\alpha = 0.01$														$\alpha = 0.005$					
$m \backslash n$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

表

$\alpha = 0.005$																				
m	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
n	15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11	
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98	
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87	
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78	
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69	
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61	
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.54	3.36	3.18	3.08	2.98	2.77	2.66	2.55	2.55	
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48	
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43	
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38	
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33	
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29	
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25	
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21	
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18	
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93	
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69	
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43	
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00	

续表

附表 4 F 分布表

 $\alpha = 0.001$

m	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
n	12	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.00	6.71	6.40	6.25	6.09	5.93	5.76	5.59	5.42
	13	17.81	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98	6.80	6.52	6.23	5.93	5.78	5.63	5.47	5.30	5.14	4.97
	14	17.14	11.78	9.73	8.62	7.92	7.43	7.08	6.80	6.58	6.40	6.13	5.85	5.56	5.41	5.25	5.10	4.94	4.77	4.60
	15	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.81	5.54	5.25	5.10	4.95	4.80	4.64	4.47	4.31
	16	16.12	10.97	9.00	7.94	7.27	6.81	6.46	6.19	5.98	5.81	5.55	5.27	4.99	4.85	4.70	4.54	4.39	4.23	4.06
	17	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75	5.58	5.32	5.05	4.78	4.63	4.48	4.33	4.18	4.02	3.85
	18	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56	5.39	5.13	4.87	4.59	4.45	4.30	4.15	4.00	3.84	3.67
	19	15.08	10.16	8.28	7.26	6.62	6.18	5.85	5.59	5.39	5.22	4.97	4.70	4.43	4.29	4.14	3.99	3.84	3.68	3.51
	20	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.82	4.56	4.29	4.15	4.00	3.86	3.70	3.54	3.38
	21	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11	4.95	4.70	4.44	4.17	4.03	3.88	3.74	3.58	3.42	3.26
	22	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99	4.83	4.58	4.33	4.06	3.92	3.78	3.63	3.48	3.32	3.15
	23	14.19	9.47	7.67	6.69	6.08	5.65	5.33	5.09	4.89	4.73	4.48	4.23	3.96	3.82	3.68	3.53	3.38	3.22	3.05
	24	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.39	4.14	3.87	3.74	3.59	3.45	3.29	3.14	2.97
	25	13.88	9.22	7.45	6.49	5.88	5.46	5.15	4.91	4.71	4.56	4.31	4.06	3.79	3.66	3.52	3.37	3.22	3.06	2.89
	26	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64	4.48	4.24	3.99	3.72	3.59	3.44	3.30	3.15	2.99	2.82
	27	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57	4.41	4.17	3.92	3.66	3.52	3.38	3.23	3.08	2.92	2.75
	28	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50	4.35	4.11	3.86	3.60	3.46	3.32	3.18	3.02	2.86	2.69
	29	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45	4.29	4.05	3.80	3.54	3.41	3.27	3.12	2.97	2.81	2.64
	30	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.00	3.75	3.49	3.36	3.22	3.07	2.92	2.76	2.59
	40	12.61	8.25	6.60	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.64	3.40	3.15	3.01	2.87	2.73	2.57	2.41	2.23
	60	11.97	7.76	6.17	5.31	4.76	4.37	4.09	3.87	3.69	3.54	3.31	3.08	2.83	2.69	2.55	2.41	2.25	2.08	1.89
	120	11.38	7.32	5.79	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.02	2.78	2.53	2.40	2.26	2.11	1.95	1.76	1.54
	∞	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.74	2.51	2.27	2.13	1.99	1.84	1.66	1.45	1.00

注: † 表示将此数乘以 100.

附表 5 泊松分布表

$$P(X \geq x) = \sum_{k=x}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}$$

x	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$
0	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	0.1812692	0.2591818	0.3296800	0.393469	0.451188	0.503415
2	0.0175231	0.0369363	0.0615519	0.090204	0.121901	0.155805
3	0.0011485	0.0035995	0.0079263	0.014388	0.023115	0.034142
4	0.0000568	0.0002658	0.0007763	0.001752	0.003358	0.005753
5	0.0000023	0.0000158	0.0000612	0.000172	0.000394	0.000786
6	0.0000001	0.0000008	0.0000040	0.000014	0.000039	0.000090
7			0.0000002	0.000001	0.000003	0.000009
8					0.000001	0.000001
x	$\lambda = 0.8$	$\lambda = 0.9$	$\lambda = 1.0$	$\lambda = 1.2$	$\lambda = 1.5$	$\lambda = 2.0$
0	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	0.550671	0.593430	0.632121	0.698806	0.776870	0.864665
2	0.191208	0.227518	0.264241	0.337373	0.442175	0.593994
3	0.047423	0.062857	0.080301	0.120513	0.191153	0.323324
4	0.009080	0.013459	0.018988	0.033769	0.065642	0.142877
5	0.001411	0.002344	0.003660	0.007746	0.018576	0.052653
6	0.000184	0.000343	0.000594	0.001500	0.004456	0.016564
7	0.000021	0.000043	0.000083	0.000251	0.000926	0.004534
8	0.000002	0.000005	0.000010	0.000037	0.000170	0.001097
9			0.000001	0.000005	0.000028	0.000237
10				0.000001	0.000004	0.000046
11					0.000001	0.000008
12						0.000001
x	$\lambda = 2.5$	$\lambda = 3.0$	$\lambda = 3.5$	$\lambda = 4.0$	$\lambda = 4.5$	$\lambda = 5.0$
0	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	0.917915	0.950213	0.969803	0.981684	0.988891	0.993262
2	0.712703	0.800852	0.864112	0.908422	0.938901	0.959572
3	0.456187	0.576810	0.679153	0.761897	0.826422	0.875348
4	0.242424	0.352768	0.463367	0.566530	0.657704	0.734974
5	0.108822	0.184737	0.274555	0.371163	0.467896	0.559507
6	0.042021	0.083918	0.142386	0.214870	0.297070	0.384039
7	0.014187	0.033509	0.065288	0.110674	0.168949	0.237817
8	0.004247	0.011905	0.026739	0.051134	0.086586	0.133372
9	0.001140	0.003803	0.009874	0.021363	0.040257	0.068094
10	0.000277	0.001102	0.003315	0.008132	0.017093	0.031828
11	0.000062	0.000292	0.001019	0.002840	0.006669	0.013695
12	0.000013	0.000071	0.000289	0.000915	0.002404	0.005453
13	0.000002	0.000016	0.000076	0.000274	0.000805	0.002019
14		0.000003	0.000019	0.000076	0.000252	0.000698
15		0.000001	0.000004	0.000020	0.000074	0.000226
16			0.000001	0.000005	0.000020	0.000069
17				0.000001	0.000005	0.000020
18					0.000001	0.000005
19						0.000001

附表 6 正态分布容许限 $\bar{X} + \lambda s$ 或 $\bar{X} - \lambda s$ 中系数 $\lambda(n, \beta, \gamma)$ 值表

$$\left(s^2 = \sum \frac{(x_i - \bar{x})^2}{n-1} \right)$$

n	1 - $\gamma = 0.95$			1 - $\gamma = 0.99$		
	1 - $\beta = 0.90$	1 - $\beta = 0.95$	1 - $\beta = 0.99$	1 - $\beta = 0.90$	1 - $\beta = 0.95$	1 - $\beta = 0.99$
5	3.41	4.21	5.75			
6	3.01	3.71	5.07	4.41	5.41	7.33
7	2.76	3.40	4.64	3.86	4.73	6.41
8	2.58	3.19	4.36	3.50	4.29	5.81
9	2.45	3.03	4.14	3.24	3.97	5.39
10	2.36	2.91	3.98	3.05	3.74	5.08
11	2.28	2.82	3.85	2.90	3.56	4.83
12	2.21	2.74	3.75	2.77	3.41	4.63
13	2.16	2.67	3.66	2.68	3.29	4.47
14	2.11	2.61	3.59	2.59	3.19	4.34
15	2.07	2.57	3.52	2.52	3.10	4.22
16	2.03	2.52	3.46	2.46	3.03	4.12
17	2.00	2.49	3.41	2.41	2.96	4.04
18	1.97	2.45	3.37	2.36	2.91	3.96
19	1.95	2.42	3.33	2.32	2.86	3.89
20	1.93	2.40	3.30	2.28	2.81	3.83
22	1.89	2.35	3.23	2.21	2.73	3.73
24	1.85	2.31	3.18	2.15	2.66	3.64
26	1.82	2.27	3.13	2.10	2.60	3.56
28	1.80	2.24	3.09	2.06	2.55	3.50
30	1.78	2.22	3.06	2.03	2.52	3.45
35	1.73	2.17	2.99	1.96	2.43	3.33
40	1.70	2.13	2.94	1.90	2.37	3.25
45	1.67	2.09	2.90	1.86	2.31	3.18
50	1.65	2.07	2.86	1.82	2.27	3.12
60	1.61	2.02	2.81	1.76	2.20	3.04
70	1.58	1.99	2.77	1.72	2.15	2.98
80	1.56	1.97	2.73	1.69	2.11	2.93
90	1.54	1.94	2.71	1.66	2.08	2.89
100	1.53	1.93	2.68	1.64	2.06	2.85
150	1.48	1.87	2.62	1.57	1.97	2.74
200	1.45	1.84	2.57	1.52	1.92	2.68
250	1.43	1.81	2.54	1.50	1.89	2.64
300	1.42	1.80	2.52	1.48	1.87	2.61
400	1.40	1.78	2.49	1.45	1.84	2.57
500	1.39	1.76	2.48	1.43	1.81	2.54
1000	1.35	1.73	2.43	1.38	1.76	2.47
∞	1.28	1.64	2.33	1.28	1.64	2.33

附表 7 正态分布容许区间 $\bar{X} \pm \lambda s$ 中系数 $\lambda(n, \beta, \gamma)$ 值表

$$\left(s^2 = \sum \frac{(x_i - \bar{x})^2}{n-1} \right)$$

n	1 - $\gamma = 0.95$			1 - $\gamma = 0.99$		
	1 - $\beta = 0.90$	1 - $\beta = 0.95$	1 - $\beta = 0.99$	1 - $\beta = 0.90$	1 - $\beta = 0.95$	1 - $\beta = 0.99$
5	4.28	5.08	6.63	6.61	7.86	10.26
6	3.71	4.41	5.78	5.34	6.35	8.30
7	3.37	4.01	5.25	4.61	5.49	7.19
8	3.14	3.73	4.89	4.15	4.94	6.47
9	2.97	3.53	4.63	3.82	4.55	5.97
10	2.84	3.38	4.43	3.58	4.27	5.59
11	2.74	3.26	4.28	3.40	4.05	5.31
12	2.66	3.16	4.15	3.25	3.87	5.08
13	2.59	3.08	4.04	3.13	3.73	4.89
14	2.53	3.01	3.96	3.03	3.61	4.74
15	2.48	2.95	3.88	2.95	3.51	4.61
16	2.44	2.90	3.81	2.87	3.41	4.49
17	2.40	2.86	3.75	2.81	3.35	4.39
18	2.37	2.82	3.70	2.75	3.28	4.31
19	2.34	2.78	3.66	2.70	3.22	4.23
20	2.31	2.75	3.62	2.66	3.17	4.16
22	2.26	2.70	3.54	2.58	3.08	4.04
24	2.23	2.65	3.48	2.52	3.00	3.95
26	2.19	2.61	3.43	2.47	2.94	3.87
28	2.16	2.58	3.39	2.43	2.89	3.79
30	2.14	2.55	3.35	2.39	2.84	3.73
35	2.09	2.49	3.27	2.31	2.75	3.61
40	2.05	2.45	3.21	2.25	2.68	3.52
45	2.02	2.41	3.17	2.20	2.62	3.44
50	2.00	2.38	3.13	2.16	2.58	3.39
60	1.96	2.33	3.07	2.10	2.51	3.29
70	1.93	2.30	3.02	2.06	2.45	3.23
80	1.91	2.27	2.99	2.03	2.41	3.17
90	1.89	2.25	2.96	2.00	2.38	3.13
100	1.87	2.23	2.93	1.98	2.36	3.10
150	1.83	2.18	2.86	1.91	2.27	2.98
200	1.80	2.14	2.82	1.87	2.22	2.92
250	1.78	2.12	2.79	1.84	2.19	2.88
300	1.77	2.11	2.77	1.82	2.17	2.85
400	1.75	2.08	2.74	1.79	2.14	2.81
500	1.74	2.07	2.72	1.78	2.12	2.78
1000	1.71	2.04	2.68	1.74	2.07	2.72
∞	1.64	1.96	2.58	1.64	1.96	2.58

附表 8 非参数容许限 —— 相应于总体比例 $1-\beta$ 和置信水平 $1-\gamma$ 的样本容量 n

$1 - \gamma$	$1 - \beta$	0.50	0.75	0.90	0.95	0.99	0.999
0.50		1	3	7	14	60	693
0.75		3	5	14	28	138	1386
0.90		4	9	22	45	230	2302
0.95		5	11	29	59	299	2995
0.99		7	17	44	90	459	4603
0.999		10	25	66	135	688	6905

附表 9 非参数容许区间 —— 相应于总体比例 $1-\beta$ 和置信水平 $1-\gamma$ 的样本容量 n

$1 - \gamma$	$1 - \beta$	0.50	0.75	0.90	0.95	0.99	0.999
0.50		3	7	17	34	169	1679
0.75		5	10	27	53	269	2692
0.90		7	15	38	77	388	3889
0.95		8	18	46	93	473	4742
0.99		11	24	64	130	662	6636
0.999		14	33	89	181	920	9230

附表 10 符号检验临界值表

本表列出了满足 $P(n_+ \geq c) \leq \alpha$ 的临界值 c

n	α	0.01	0.05	0.10	n	α	0.01	0.05	0.10
5					18		15	13	13
6				6	19		15	14	13
7		7	6		20		16	15	14
8	8	7	7		21		17	15	14
9	9	8	7		22		17	16	15
10	10	9	8		23		18	16	16
11	10	9	9		24		19	17	16
12	11	10	9		25		19	18	17
13	12	10	10		26		20	18	17
14	12	11	10		27		20	19	18
15	13	12	11		28		21	19	18
16	14	12	12		29		22	20	19
17	14	13	12		30		22	20	20

续表

$n \backslash \alpha$	0.01	0.05	0.10	$n \backslash \alpha$	0.01	0.05	0.10
31	23	21	20	41	29	27	26
32	24	22	21	42	29	27	26
33	24	22	21	43	30	28	27
34	25	23	22	44	31	28	27
35	25	23	22	45	31	29	28
36	26	24	23	46	32	30	28
37	26	24	23	47	32	30	29
38	27	25	24	48	33	31	29
39	28	26	24	49	34	31	30
40	28	26	25	50	34	32	31

附表 11 符号秩和检验临界值表

$$P(w^+ \geq c) \leq \alpha$$

n	α			
	0.01	0.025	0.05	0.10
4	11	11	11	10
5	16	16	15	13
6	22	21	19	18
7	28	26	25	23
8	35	33	31	28
9	42	40	37	35
10	50	47	45	41
11	59	56	53	49
12	68	65	61	57
13	79	74	70	65
14	90	85	80	74
15	101	95	90	84
16	113	107	101	94
17	126	119	112	105
18	139	131	124	116
19	153	144	137	128
20	167	158	150	141

附表 12 秩和检验临界值表

$$P(W \geq c) \leq \alpha$$

附表 13 柯尔莫哥洛夫检验临界值 $D_{n,\alpha}$

$$P(D_n \geq D_{n,\alpha}) = \alpha$$

$n \backslash \alpha$	0.20	0.10	0.05	0.02	0.01
1	0.900	0.950	0.975	0.990	0.995
2	0.684	0.776	0.842	0.900	0.929
3	0.565	0.636	0.708	0.785	0.829
4	0.493	0.565	0.624	0.689	0.734
5	0.447	0.509	0.563	0.627	0.669
6	0.410	0.468	0.519	0.577	0.617
7	0.381	0.436	0.483	0.538	0.576
8	0.358	0.410	0.454	0.507	0.542
9	0.339	0.387	0.430	0.480	0.513
10	0.323	0.369	0.409	0.457	0.489
11	0.308	0.352	0.391	0.437	0.468
12	0.296	0.338	0.375	0.419	0.449
13	0.285	0.325	0.361	0.404	0.432
14	0.275	0.314	0.349	0.390	0.418
15	0.266	0.304	0.338	0.377	0.404
16	0.258	0.295	0.327	0.366	0.392
17	0.250	0.286	0.318	0.355	0.381
18	0.244	0.279	0.309	0.346	0.371
19	0.237	0.271	0.301	0.337	0.361
20	0.232	0.265	0.294	0.329	0.352
21	0.226	0.259	0.287	0.321	0.344
22	0.221	0.253	0.281	0.314	0.337
23	0.216	0.247	0.275	0.307	0.330
24	0.212	0.242	0.269	0.301	0.323
25	0.208	0.238	0.264	0.295	0.317
26	0.204	0.233	0.259	0.290	0.311
27	0.200	0.229	0.254	0.284	0.305
28	0.197	0.225	0.250	0.279	0.300
29	0.193	0.221	0.246	0.275	0.295
30	0.190	0.218	0.242	0.270	0.290
31	0.187	0.214	0.238	0.266	0.285
32	0.184	0.211	0.234	0.262	0.281
33	0.182	0.208	0.231	0.258	0.277
34	0.179	0.205	0.227	0.254	0.273
35	0.177	0.202	0.224	0.251	0.269
36	0.174	0.199	0.221	0.247	0.265
37	0.172	0.196	0.218	0.244	0.262
38	0.170	0.194	0.215	0.241	0.258
39	0.168	0.191	0.213	0.238	0.255
40	0.165	0.189	0.210	0.235	0.252
对 $n > 40$ 近似	$1.07/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.52/\sqrt{n}$	$1.63/\sqrt{n}$

附表 14 柯尔莫哥洛夫检验统计量 D_n 的极限分布附表 14 柯尔莫哥洛夫检验统计量 D_n 的极限分布

$$K(\lambda) = \lim_{n \rightarrow \infty} P(D_n \leq \lambda / \sqrt{n}) = \sum_{j=-\infty}^{\infty} (-1)^j \cdot \exp\{-2j^2 \lambda^2\}$$

λ	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	λ
0.2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000004	0.2
0.3	0.000009	0.000021	0.000046	0.000091	0.000171	0.000303	0.000511	0.000826	0.001285	0.001929	0.3
0.4	0.002808	0.003972	0.005476	0.007377	0.009730	0.012590	0.016005	0.020022	0.024682	0.030017	0.4
0.5	0.036055	0.042814	0.050306	0.058534	0.067497	0.077183	0.087577	0.098656	0.110395	0.122760	0.5
0.6	0.135718	0.149229	0.163225	0.177153	0.192677	0.207987	0.223637	0.239582	0.255780	0.272189	0.6
0.7	0.288765	0.305471	0.322265	0.339113	0.355981	0.372833	0.389640	0.406372	0.423002	0.439505	0.7
0.8	0.455857	0.472041	0.488030	0.503080	0.519366	0.534682	0.549744	0.564546	0.579070	0.593316	0.8
0.9	0.607270	0.620928	0.634286	0.647338	0.660082	0.672516	0.684630	0.696444	0.707940	0.719126	0.9
1.0	0.730000	0.740566	0.750826	0.760780	0.770434	0.777974	0.788860	0.797636	0.806128	0.814342	1.0
1.1	0.822282	0.829950	0.837356	0.844502	0.851394	0.858038	0.864442	0.870612	0.876548	0.882558	1.1
1.2	0.887750	0.893030	0.898104	0.902972	0.907648	0.912132	0.916432	0.920556	0.924505	0.928288	1.2
1.3	0.931908	0.935370	0.938682	0.941848	0.944872	0.947756	0.950512	0.953142	0.955650	0.958040	1.3
1.4	0.960318	0.962486	0.964552	0.966516	0.968382	0.970158	0.971846	0.973448	0.974970	0.976412	1.4
1.5	0.977782	0.979080	0.980310	0.981476	0.982578	0.983622	0.984610	0.985544	0.986426	0.987260	1.5
1.6	0.988048	0.988791	0.989492	0.990154	0.990777	0.991364	0.991917	0.992438	0.992928	0.993389	1.6
1.7	0.993823	0.994230	0.994612	0.99472	0.995309	0.995625	0.995922	0.996200	0.996460	0.996704	1.7
1.8	0.996332	0.997146	0.997346	0.997533	0.997707	0.997870	0.998023	0.998145	0.998297	0.998421	1.8
1.9	0.998536	0.998644	0.998744	0.998837	0.998924	0.999004	0.999079	0.999149	0.999123	0.999273	1.9
2.0	0.999329	0.999380	0.999428	0.999474	0.999516	0.999552	0.999588	0.999620	0.999650	0.999680	2.0
2.1	0.999705	0.999728	0.999750	0.999770	0.999790	0.999806	0.999822	0.999838	0.999852	0.999864	2.1
2.2	0.999874	0.999886	0.999896	0.999904	0.999912	0.999920	0.999926	0.999934	0.999940	0.999944	2.2
2.3	0.999949	0.999954	0.999958	0.999962	0.999965	0.999968	0.999970	0.999973	0.999976	0.999978	2.3
2.4	0.999980	0.999982	0.999984	0.999986	0.999987	0.999988	0.999988	0.999990	0.999991	0.999992	2.4

附表 15 W 检验 统计量 W 的系数 $a_i(n)$ 的值

$i \backslash n$	3	4	5	6	7	8	9	10	
1	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739	
	—	0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291	
	—	—	—	0.0875	0.1401	0.1743	0.1976	0.2141	
	—	—	—	—	—	0.0561	0.0947	0.1224	
	—	—	—	—	—	—	—	0.0399	
	11	12	13	14	15	16	17	18	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808
2	0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253	0.3232
3	0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0.2540	0.2553	0.2561
4	0.1429	0.1586	0.1707	0.1802	0.1878	0.1939	0.1988	0.2027	0.2059
5	0.0695	0.0922	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587	0.1641
6	—	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271
7	—	—	—	0.0240	0.0433	0.0593	0.0725	0.0837	0.0932
8	—	—	—	—	—	0.0196	0.0359	0.0496	0.0612
9	—	—	—	—	—	—	0.0163	0.0303	0.0422
10	—	—	—	—	—	—	—	—	0.0140
	21	22	23	24	25	26	27	28	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291
2	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968
3	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2510	0.2499
4	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2153	0.2151	0.2150
5	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836	0.1848	0.1857	0.1864
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616
7	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316	0.1346	0.1372	0.1395
8	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089	0.1128	0.1162	0.1192
9	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876	0.0923	0.0965	0.1002
10	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672	0.0728	0.0778	0.0822
11	—	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650
12	—	—	—	0.0107	0.0200	0.0284	0.0358	0.0424	0.0483
13	—	—	—	—	—	0.0094	0.0178	0.0253	0.0320
14	—	—	—	—	—	—	0.0084	0.0159	0.0227
15	—	—	—	—	—	—	—	—	0.0076
	31	32	33	34	35	36	37	38	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989
2	0.2921	0.2898	0.2876	0.2854	0.2834	0.2813	0.2794	0.2774	0.2755
3	0.2475	0.2463	0.2451	0.2439	0.2427	0.2415	0.2403	0.2391	0.2380
4	0.2145	0.2141	0.2137	0.2132	0.2127	0.2121	0.2116	0.2110	0.2098
5	0.1874	0.1878	0.1880	0.1882	0.1883	0.1883	0.1883	0.1881	0.1878

附表 15 W 检验 统计量 W 的系数 $a_i(n)$ 的值

续表

附表 16 W 检验 统计量 W 的 α 分位数 W_α

$n \backslash \alpha$	0.01	0.05	0.10
3	0.753	0.767	0.789
4	0.687	0.748	0.792
5	0.686	0.762	0.806
6	0.713	0.788	0.826
7	0.730	0.803	0.838
8	0.749	0.818	0.851
9	0.764	0.829	0.859
10	0.781	0.842	0.869
11	0.792	0.850	0.876
12	0.805	0.859	0.883
13	0.814	0.866	0.889
14	0.825	0.874	0.895
15	0.835	0.881	0.901
16	0.844	0.887	0.906
17	0.851	0.892	0.910
18	0.858	0.897	0.914
19	0.863	0.901	0.917
20	0.868	0.905	0.920
21	0.873	0.908	0.923
22	0.878	0.911	0.926
23	0.881	0.914	0.928
24	0.884	0.916	0.930
25	0.888	0.918	0.931
26	0.891	0.920	0.933
27	0.894	0.923	0.935
28	0.896	0.924	0.936
29	0.898	0.926	0.937
30	0.900	0.927	0.939
31	0.902	0.929	0.940
32	0.904	0.930	0.941
33	0.906	0.931	0.942
34	0.908	0.933	0.943
35	0.910	0.934	0.944
36	0.912	0.935	0.945
37	0.914	0.936	0.946
38	0.916	0.938	0.947
39	0.917	0.939	0.948
40	0.919	0.940	0.949
41	0.920	0.941	0.950
42	0.922	0.942	0.951
43	0.923	0.943	0.951
44	0.924	0.944	0.952
45	0.926	0.945	0.953
46	0.927	0.945	0.953
47	0.928	0.946	0.954
48	0.929	0.947	0.954
49	0.929	0.947	0.955
50	0.930	0.947	0.955

附表 17 D 检验 统计量 Y 的 α 分位数 Y_α

$n \backslash \alpha$	0.005	0.025	0.05	0.95	0.975	0.995
50	-3.91	-2.74	-2.21	0.937	1.06	1.24
60	-3.81	-2.68	-2.17	0.997	1.13	1.34
70	-3.73	-2.64	-2.14	1.05	1.19	1.42
80	-3.67	-2.60	-2.11	1.08	1.24	1.48
90	-3.61	-2.57	-2.09	1.12	1.28	1.54
100	-3.57	-2.54	-2.07	1.14	1.31	1.59
150	-3.41	-2.45	-2.00	1.23	1.42	1.75
200	-3.30	-2.39	-1.96	1.29	1.50	1.85
250	-3.23	-2.35	-1.93	1.33	1.55	1.93
300	-3.17	-2.32	-1.91	1.36	1.58	1.98
350	-3.13	-2.29	-1.89	1.38	1.61	2.03
400	-3.09	-2.27	-1.87	1.40	1.63	2.06
450	-3.06	-2.25	-1.86	1.41	1.65	2.09
500	-3.04	-2.24	-1.85	1.42	1.67	2.11
550	-3.02	-2.23	-1.84	1.43	1.68	2.14
600	-3.00	-2.22	-1.83	1.44	1.69	2.15
650	-2.98	-2.21	-1.83	1.45	1.70	2.17
700	-2.97	-2.20	-1.82	1.46	1.71	2.18
750	-2.96	-2.19	-1.81	1.47	1.72	2.20
800	-2.94	-2.18	-1.81	1.47	1.73	2.21
850	-2.93	-2.18	-1.80	1.48	1.74	2.22
900	-2.92	-2.17	-1.80	1.48	1.74	2.23
950	-2.91	-2.16	-1.80	1.49	1.75	2.24
1000	-2.91	-2.16	-1.79	1.49	1.75	2.25

索引

- 2×2 列联表, 258
Bayes 风险, 307, 308
Bayes 解, 278, 307
Bayes 统计学, 277
Bayes 学派, 276
Behrens-Fisher 问题, 138, 152, 188
C-R 不等式, 110, 114
C-R 正则条件, 109
 D 检验, 271, 272
 F 分布, 39, 187
 F 检验, 187
Fisher 信息量, 109, 113
James-Stein 估计, 324
Jeffreys 无信息先验, 288
Lehmann-Scheffé 定理, 103, 104
Minimax 解, 316
Neyman-Pearson 引理, 203
Wilcoxon 两样本秩和统计量, 243
 χ^2 分布, 32, 180, 248
 χ^2 检验, 180
 $r \times s$ 列联表, 259
 r 阶矩相合估计, 74, 82
 t 分布, 36, 37, 189
备择假设, 168
不可容许的决策函数, 321
参数, 15
参数估计, 70
参数空间, 15, 85, 306
参数统计推断, 16
参数向量, 15
超参数, 284
充分统计量, 50, 51, 54, 56
充分完全统计量法, 98
抽样, 8
抽样信息, 276
次序统计量, 19, 28, 57, 159, 269
大样本分布, 43
大样本性质, 21, 44, 74, 79
单边检验, 175
第二类错误, 170
第一类错误, 170
点估计, 70
定时截尾试验, 91
定数截尾试验, 91
对立假设, 168
对数似然函数, 85
非参数统计推断, 16
非随机化检验, 169
非中心 χ^2 分布, 36
非中心 F 分布, 41
非中心 t 分布, 38, 158
风险函数, 306
否定域, 169
符号检验, 233
符号秩和, 239
符号秩和检验, 238, 240
复合假设, 168, 206
格里汶科定理, 21, 22
个体, 8
功效函数, 171
共轭先验分布, 278

- 共轭先验分布族, 290
广义 Bayes 解, 309
广义似然比检验, 194
广义先验密度, 286
- 核估计, 118
后验分布, 278
后验风险, 308, 309
后验风险最小准则, 308
后验均方误差, 298
后验密度, 278
后验期望估计, 296
后验预测密度, 304
后验中位数估计, 296
后验众数估计, 296
- 极差, 32
极大似然估计, 86
极小充分统计量, 58
极值, 19
检验的 p 值, 223
检验函数, 169
检验统计量, 170
简单假设, 168
简单随机样本, 10
渐近无偏估计, 72, 80, 81
渐近有效估计, 115
接受域, 169
经典(古典)统计学, 276
经验 Bayes 方法, 278
经验分布函数, 20, 21
精度, 127, 128
矩法, 76, 79
矩估计量, 75, 76
拒绝域, 169
均方误差, 97
均方相合估计, 74, 82
- 柯尔莫哥洛夫检验, 265
可估参数, 98
可估函数, 98
可靠度, 127
可容许的决策函数, 321
可信区间, 300
可信上限, 300
可信下限, 300
刻度参数, 287
刻度同变估计, 320
- 两样本 t 检验, 183, 184, 187
两样本 U 检验, 181
临界值, 170
零假设, 168
零无偏估计法, 98, 100
- 拟合优度, 247
拟合优度检验, 248
频率学派, 279
- 强相合估计, 74, 82, 83
区间估计, 70, 127–129
- 容忍区间, 155, 156
容忍上限, 156
容忍下限, 156
容忍限, 156
弱相合估计, 74, 91
枢轴变量, 132
枢轴变量法, 131, 132, 141
- 双边检验, 175
斯米尔诺夫检验, 242, 268
似然比, 193, 207, 210
似然比检验, 194
似然函数, 84–86
随机化检验, 169, 170
损失函数, 306

- 统计大样本理论, 44
 统计量, 17
 统计模型, 14–16
 统计推断, 10, 15, 16
 完全统计量, 59, 61, 63, 103
 位置参数, 286
 位置同变估计, 320
 稳健性, 232
 无偏估计量, 71, 73
 无偏检验, 216
 无限总体, 8, 9
 无信息先验, 285
 无信息先验分布, 278
 先验分布, 278
 显著性水平, 172
 相合渐近正态估计, 82
 小样本性质, 44, 79
 信仰区间, 148, 149
 信仰系数, 148, 149
 行动空间, 306
 样本, 8
 样本 k 阶原点矩, 18, 75
 样本 k 阶中心矩, 19, 75
 样本 p 分位数, 19
 样本变异系数, 20, 78
 样本大小, 8
 样本的两重性, 10
 样本方差, 18, 19
 样本分布族, 15, 16
 样本峰度, 20, 78
 样本极差, 19, 30
 样本极值, 19
 样本矩, 19
 样本均值, 17–19
 样本空间, 9, 169, 306
 样本偏度, 20, 78
 样本容量, 8
 样本协方差, 19
 样本中位数, 3, 19
 一样本 t 检验, 177
 一样本 U 检验, 176
 一致最精确区间估计, 220
 一致最小方差无偏估计, 98, 99
 一致最优检验, 203
 一致最优解, 307
 一致最优决策函数, 307
 一致最优无偏检验, 216
 因子分解定理, 54
 有界完全统计量, 63
 有限总体, 8
 有效估计, 115
 有效性, 73
 原假设, 168
 指数族, 46, 48, 57, 87, 111
 指数族的自然形式, 48
 秩检验, 239
 秩统计量, 239
 置信度, 128
 置信区间, 129, 130
 置信上限, 130
 置信水平, 128–131
 置信系数, 128–131
 置信下限, 130
 置信域, 131
 主观概率, 279, 281
 自然参数空间, 48
 总体, 8, 9
 最优同变估计, 320

《中国科学技术大学数学教学丛书》

程 艺 主编

- | | |
|------------------|-----------------|
| ◆ 微积分五讲 | 龚 翌 |
| ◆ 线性代数五讲 | 龚 翌 |
| ◆ 微积分（上、下）（第二版） | 谢盛刚 李 娟 陈秋桂 |
| ◆ 微积分学习辅导 | 陈效群 陈秋桂 顾新身 |
| ◆ 复变函数 | 潘永亮 汪琥庭 汪芳庭 宋立功 |
| ◆ 数学物理方程（第二版） | 季孝达 薛兴恒 陆 英 |
| ◆ 随机过程（第三版） | 方兆本 缪柏其 |
| ◆ 数量经济分析（第二版） | 侯定丕 陈华友 陈 伟 |
| ◆ 概率论（第二版） | 苏 淳 |
| ◆ 数理统计（第二版） | 韦来生 |
| ◆ 组合数学 | 徐俊明 潘永亮 |
| ◆ 数值计算方法与算法（第二版） | 张韵华 奚梅成 陈效群 |

www.sciencep.com

ISBN 978-7-03-046573-3



9 787030 465733 >

定 价：46.00 元

高等教育出版中心 数理出版分社

联系电话：010-64034725

E-mail: mph@mail.sciencep.com