# 韦来生《数理统计》抄写

BacktoBed

# Contents

**Contents** **i**

**1 绪论** **1**

1.1 引言 . . . . . 1
1.2 基本概念 . . . . . 1
1.2.1 总体和样本 . . . . . 1
1.2.2 样本空间和样本的两重性 . . . . . 1
1.2.3 样本分布 . . . . . 2
1.2.4 统计模型 . . . . . 2
1.2.5 统计推断 . . . . . 2
1.3 统计量 . . . . . 3
1.3.1 统计量的定义 . . . . . 3
1.3.2 若干常用统计量 . . . . . 3
1. 样本均值 . . . . . 3
2. 样本方差 . . . . . 4
3. 样本矩 . . . . . 4
4. 二维随机向量的样本矩 . . . . . 4
5. 次序统计量及其有关统计量 . . . . . 5
6. 样本变异系数 . . . . . 5
7. 样本偏度 . . . . . 5
8. 样本峰度 . . . . . 6
1.3.3 经验分布函数 . . . . . 6

**2 抽样分布及若干预备知识** **8**

2.1 引言 . . . . . 8
2.2 正态总体样本均值和样本方差的分布 . . . . . 8
2.2.1 正态变量线性函数的分布 . . . . . 8
2.2.2 正态样本均值和样本方差的分布 . . . . . 9
2.3 次序统计量的分布 . . . . . 10

2.3.1 单个次序统计量的分布 . . . . . . . . . . 10
2.3.2 次序统计量的联合分布 . . . . . . . . . . 11
2.3.3 极差的分布 . . . . . . . . . . 11
2.3.4 均匀分布情形 . . . . . . . . . . 12
2.4 $\chi^2$ 分布,$t$ 分布和 $F$ 分布 . . . . . . . . . . 13
2.4.1 $\chi^2$ 分布 . . . . . . . . . . 13
2.4.2 $t$ 分布 . . . . . . . . . . 14
2.4.3 $F$ 分布 . . . . . . . . . . 16
2.4.4 几个重要推论 . . . . . . . . . . 17
2.5 统计量的极限分布 . . . . . . . . . . 18
2.5.1 定义 . . . . . . . . . . 18
2.5.2 几个例子 . . . . . . . . . . 19
*2.6 指数族 . . . . . . . . . . 20
2.6.1 定义与例子 . . . . . . . . . . 20
2.6.2 指数族的自然形式及自然参数空间 . . . . . . . . . . 21
2.6.3 指数族的性质 . . . . . . . . . . 22
2.7 充分统计量 . . . . . . . . . . 23
2.7.1 引言与定义 . . . . . . . . . . 23
2.7.2 充分性的判别准则—因子分解定理 . . . . . . . . . . 24
*2.7.3 极小充分统计量 . . . . . . . . . . 25
*2.8 完全统计量 . . . . . . . . . . 26
2.8.1 定义和例子 . . . . . . . . . . 26
2.8.2 指数族中统计量的完全性 . . . . . . . . . . 27
2.8.3 有界完全统计量及其性质 . . . . . . . . . . 27

**3 点估计 29**

3.1 引言 . . . . . . . . . . 29
3.1.1 参数估计问题 . . . . . . . . . . 29
3.1.2 点估计 . . . . . . . . . . 29
3.1.3 点估计的优良性准则 . . . . . . . . . . 29
1. 无偏性 . . . . . . . . . . 29
2. 有效性 . . . . . . . . . . 30
3. 相合性 . . . . . . . . . . 30
3.2 矩估计 . . . . . . . . . . 31
3.2.1 矩法和距估计量 . . . . . . . . . . 31
3.2.2 若干例子 . . . . . . . . . . 32
3.2.3 矩估计的无偏性和渐近无偏性 . . . . . . . . . . 35
3.2.4 矩估计的相合性 . . . . . . . . . . 36

*3.2.5 矩估计的渐近正态性 . . . 37
3.3 极大似然估计 . . . 39
3.3.1 引言和定义 . . . 39
3.3.2 极大似然估计的求法及例 . . . 41
1. 用微积分中求极值的方法 . . . 41
2. 从定义出发 . . . 42
3.3.3 极大似然估计的性质 . . . 48
1. 极大似然估计的无偏性 . . . 48
2. 极大似然估计与充分统计量 . . . 48
3. 极大似然估计的相合性 . . . 48
*4. 极大似然估计的渐近相合正态性 . . . 48
*3.4 一致最小方差无偏估计 . . . 49
3.4.1 引言及定义 . . . 49
*3.4.2 零无偏估计法 . . . 51
3.4.3 充分完全统计量法 . . . 53
3.5 Cramer-Rao 不等式 . . . 55
3.5.1 引言 . . . 55
3.5.2 单参数 C-R 不等式 . . . 55
1.C-R 不等式及例 . . . 55
*2.C-R 不等式等号成立的条件 . . . 57
3.Fisher 信息函数 . . . 57
3.5.3 多参数 C-R 不等式简介 . . . 58
3.5.4 有效估计和估计的效率 . . . 58
*3.6 概率密度函数的核估计 . . . 59
3.6.1 概率密度函数的核估计 . . . 59
1. 概率密度函数的“自然”估计 . . . 59
2. 概率密度估计核函数的定义 . . . 59
3.6.2 概率密度函数导数的核估计 . . . 61
3.6.3 概率密度函数核估计的大样本性质 . . . 61
1. 若干定义 . . . 61
2. 大样本性质 . . . 61

**4 区间估计 64**
4.1 区间估计的基本概念 . . . 64
4.1.1 参数的区间估计问题 . . . 64
4.1.2 置信区间 . . . 64
1. 置信度 . . . 65
2. 精度 . . . 65

4.1.3 置信限 . . . 66
4.1.4 置信域 . . . 66
4.1.5 构造区间估计的方法 . . . 67
4.2 枢轴变量法—正态总体参数的置信区间 . . . 67
4.2.1 引言 . . . 67
4.2.2 单个正态总体的置信区间 . . . 68
1.$\sigma^2$ 已知, 求 $\mu$ 的置信区间 . . . 68
2.$\sigma^2$ 未知, 求 $\mu$ 的置信区间 . . . 69
3.$\mu$ 已知, 求 $\sigma^2$ 的置信区间 . . . 70
4.$\mu$ 未知, 求 $\sigma^2$ 的置信区间 . . . 70
4.2.3 两个正态总体参数的置信区间 . . . 71
1. 均值差 $b-a$ 的置信区间 . . . 71
2. 方差比 $\sigma_1^2/\sigma_2^2$ 的置信区间 . . . 74
4.3 枢轴变量法—非正态总体参数的置信区间 . . . 76
4.3.1 小样本方法 . . . 76
1. 指数分布参数的置信区间 . . . 76
2. 均匀分布参数的置信区间 . . . 77
4.3.2 大样本方法 . . . 78
1.Cauchy 分布未知参数的置信区间 . . . 78
2. 二项分布总体参数的置信区间 . . . 79
3.Poisson 分布参数的置信区间 . . . 80
*4. 一般情形 . . . 81
*5. 非参数情形 . . . 81
*4.4 Fisher 的信仰推断法 . . . 82
4.4.1 引言 . . . 82
4.4.2 信仰区间的求法 . . . 82
4.4.3 用 Fisher 方法解 Behrens-Fisher 问题 . . . 85
*4.5 容忍区间和容忍限 . . . 87
4.5.1 问题的提法及定义 . . . 87
4.5.2 正态总体的容忍区间和容忍限 . . . 89
4.5.3 非参数容忍限和容忍区间 . . . 91
1. 求 $F$ 的水平 $(1-\beta, 1-\gamma)$ 的容忍区间和容忍限 . . . 91
2. 特例 . . . 92

**5 参数假设检验** **94**
5.1 假设检验的若干基本概念 . . . 94
5.1.1 检验问题的提法 . . . 94
5.1.2 否定域, 检验函数和检验统计量 . . . 95

5.1.3 假设检验和功效函数 . . . . . . . . 97
5.1.4 检验水平和控制犯第一类错误概率的原则 . . . . . . . . 98
5.1.5 求解假设检验问题的一般步骤 . . . . . . . . 100
5.1.6 假设检验发展简史 . . . . . . . . 100
5.2 正态总体的假设检验 . . . . . . . . 100
5.2.1 单个正态总体均值的检验 . . . . . . . . 100
1. 方差 $\sigma^2$ 已知时的检验方法 . . . . . . . . 101
2. 方差未知时正态总体均值的检验方法 . . . . . . . . 102
5.2.2 单个正态总体方差的检验 . . . . . . . . 104
1. 均值未知时单个正态总体方差的检验方法 . . . . . . . . 104
2. 当均值 $\mu$ 已知时方差 $\sigma^2$ 的检验方法 . . . . . . . . 105
5.2.3 两个正态总体均值差的检验 . . . . . . . . 106
1. 当 $\sigma_1^2$ 和 $\sigma_2^2$ 已知时均值差的检验 . . . . . . . . 106
2. 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知时均值差的检验 . . . . . . . . 107
3. 当样本容量 $m=n$ 时均值差的检验 . . . . . . . . 108
5.2.4 两个正态总体方差比的检验 . . . . . . . . 110
1. 当 $\mu_1$ 和 $\mu_2$ 未知时方差比的检验方法 . . . . . . . . 110
2. 当 $\mu_1$ 和 $\mu_2$ 已知时方差比的检验方法 . . . . . . . . 111
5.2.5 极限分布为正态分布的检验 . . . . . . . . 112
1.Behrens-Fisher 检验问题的近似方法 . . . . . . . . 113
2. 二项分布参数的大样本检验 . . . . . . . . 115
3.Poisson 分布参数的大样本检验 . . . . . . . . 116
4. 两样本检验问题 . . . . . . . . 116
5.3 似然比检验 . . . . . . . . 117
5.3.1 似然比检验的定义 . . . . . . . . 117
5.3.2 若干例子 . . . . . . . . 118
*5.3.3 似然比的渐近分布 . . . . . . . . 124
*5.4 一致最优检验和无偏检验 . . . . . . . . 126
5.4.1 引言与定义 . . . . . . . . 126
5.4.2 Neyman-Pearson 引理 . . . . . . . . 127
5.4.3 利用 NP 引理求一致最优检验 . . . . . . . . 131
*5.4.4 无偏检验 . . . . . . . . 136
5.5 假设检验和区间估计 . . . . . . . . 137
5.5.1 如何由假设检验得到置信区间 . . . . . . . . 137
5.5.2 如何由置信区间得到假设检验 . . . . . . . . 139
*5.5.3 一致最精确置信集和一致最优检验 . . . . . . . . 139
1. 引言和定义 . . . . . . . . 139
2. 如何求 UMA 置信集 . . . . . . . . 140

5.5.4 假设检验和区间估计的比较 . . . . . . 140
5.5.5 检验的 $p$ 值 . . . . . . 141

**6 非参数假设检验** **144**
6.1 引言 . . . . . . 144
6.2 符号检验及符号秩和检验 . . . . . . 144
6.2.1 符号检验法 . . . . . . 144
1. 小样本方法 . . . . . . 145
2. 大样本方法 . . . . . . 146
6.2.2 符号秩和检验 . . . . . . 149
1. 小样本方法 . . . . . . 150
2. 大样本方法 . . . . . . 151
6.3 Wilcoxon 两样本秩和检验 . . . . . . 152
6.3.1 引言及定义 . . . . . . 152
6.3.2 Wilcoxon 两样本秩和检验—小样本方法 . . . . . . 154
6.3.3 Wilcoxon 两样本秩和检验—大样本方法 . . . . . . 156
6.4 拟合优度检验 . . . . . . 158
6.4.1 引言 . . . . . . 158
6.4.2 Pearson $\chi^2$ 检验: 理论分布完全已知的情况 . . . . . . 158
1. 随机变量 $X$ 为离散型, 且只取有限个不同值 $a_1, \cdots, a_r$ 的情形 . . . . . . 158
2. 理论分布为任一确定分布的情形 . . . . . . 160
6.4.3 Pearson $\chi^2$ 检验: 理论分布带有未知参数的情况 . . . . . . 161
*6.4.4 定理 6.4.1 的证明 . . . . . . 166
6.5 列联表中的独立性和齐一性检验 . . . . . . 166
6.5.1 列联表中的独立性检验—$\chi^2$ 检验的应用 . . . . . . 166
1. 问题的提法 . . . . . . 166
2. 检验方法 . . . . . . 167
6.5.2 列联表中的齐一性检验 . . . . . . 169
1. 问题的提法 . . . . . . 169
2. 检验的制定 . . . . . . 170
*6.6 其他的非参数检验方法 . . . . . . 172
6.6.1 柯尔莫哥洛夫检验 . . . . . . 172
6.6.2 斯米尔诺夫检验 . . . . . . 175
*6.6.3 正态性检验 . . . . . . 176
1.$W$ 检验 (Wilk 检验) . . . . . . 176
2.$D$ 检验 . . . . . . 178

**7 Bayes 方法和统计决策理论 180**

7.1 引言和若干基本概念 . . . 180

7.1.1 引言 . . . 180

7.1.2 先验分布和后验分布 . . . 182

7.1.3 古典学派和 Bayes 学派的争论 . . . 182

7.1.4 历史 . . . 182

7.2 先验分布的确定 . . . 183

7.2.1 主观概率 . . . 183

1. 引言和定义 . . . 183

2. 确定主观概率的方法 . . . 183

7.2.2 利用先验信息确定先验分布 . . . 183

1. 直方图法 . . . 183

2. 相对似然法 . . . 184

3. 选定先验密度函数的形式, 再估计超参数 . . . 185

7.2.3 无信息先验 . . . 186

1. 均匀分布与广义先验分布 . . . 187

2. 位置参数的无信息先验 . . . 187

3. 刻度参数的无信息先验 . . . 188

*4. 一般情形下的无信息先验 . . . 189

7.2.4 共轭先验分布 . . . 191

1. 共轭先验分布的概念 . . . 191

2. 后验分布的计算 . . . 192

7.3 Bayes 统计推断 . . . 196

7.3.1 点估计 . . . 196

1.Bayes 点估计方法 . . . 196

*2.Bayes 点估计的精度—估计的误差 . . . 198

*3. 多参数情形 . . . 200

7.3.2 区间估计 . . . 200

7.3.3 假设检验 . . . 202

1. 假设检验的一般方法 . . . 202

2. 多重假设检验 . . . 203

*3. 预测推断 . . . 204

7.4 Bayes 统计决策理论 . . . 205

7.4.1 统计决策理论的若干基本概念 . . . 205

1. 统计决策三要素和贝叶斯统计决策四要素 . . . 206

2. 风险函数和一致最优决策 . . . 207

3.Bayes 准则 . . . 207

7.4.2 后验风险最小的原则 . . . 208

1. 后验风险 . . . 208
2. 后验风险与 Bayes 风险的关系 . . . 208
3. 后验风险最小的原则 . . . 209
7.4.3 一般损失函数下的 Bayes 估计 . . . 209
1. 平方损失下的 Bayes 估计 . . . 210
2. 加权平方损失下的 Bayes 估计 . . . 211
3. 绝对值损失下的 Bayes 解 . . . 212
7.4.4 假设检验和有限行动 (分类) 问题 . . . 213
1. 假设检验问题 . . . 213
2. 多行动 (分类) 问题 . . . 214
*3. 统计决策中的区间估计问题 . . . 216
*7.5 minimax 准则 . . . 216
7.5.1 引言及定义 . . . 216
7.5.2 Minimax 解的求法 . . . 217
*7.6 同变估计及可容许性 . . . 219
7.6.1 同变估计及例子 . . . 219
7.6.2 决策函数的可容许性 . . . 221
1. 可容许性的概念及判别方法 . . . 221
2. James-Stein 估计 . . . 223

# Chapter 1

# 绪论

## 1.1 引言

没啥好说的.

## 1.2 基本概念

### 1.2.1 总体和样本

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**可以使用一个随机变量及其概率分布来描述.

实际上, 统计问题研究的是某种数量指标, 当这个数量指标的可能值有无限多的时候, 称为无限总体, 否则总体就是有限的了.

比如说, 用秤给物体称重, 则秤的读数的性质类似于随机变量, 样本总体为无限总体.

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 $X_1, \cdots, X_n$, 则常记为 $X_1, \cdots, X_n$ i.i.d. $\sim F$.

若 $F$ 有密度 $f$, 可记为 $X_1, \cdots, X_n$ i.i.d. $\sim f$.

当个体上的数量指标不止一项时, 用随机变量来表示总体.

### 1.2.2 样本空间和样本的两重性

1. 样本空间

样本空间定义:

**Definition 1.2.1.** 样本 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 的可能取值的全体.

2. 样本的两重性

样本的两重性指的是, 样本既可以看成式具体的数, 又可以看成随机变量 (的观测值).

3. 简单随机样本

抽样是指从总体中获得样本的方式与行为, 为借助样本对总体有较好的估计. 最常用的一种抽取方法称为"简单随机抽样", 它满足下列要求:

(1) 代表性. 样本中每个个体满足的分布都与总体的分布相同.

(2) 独立性. 样本中的各个个体都是彼此独立的随机变量.

由简单随机抽样获得的样本称为简单随机样本. 其定义如下:

**Example 1.2.1.** 设有一总体 $F$,$X_1, X_2, \cdots, X_n$ 为从 $F$ 中抽取的容量为 $n$ 的样本, 若

(1) $X_1, \cdots, X_n$ 相互独立,

(2) $X_1, \cdots, X_n$ 相同分布, 即同有分布 $F$,

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

### 1.2.3 样本分布

样本分布既然是随机变量, 就需要有一定的概率分布. 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最完整的表述.

**Example 1.2.2.** 有限多件产品中用简单随机抽样抽取一定量的样本, 样本中的赝品数量满足超几何分布, 也就是说可以用乘法原理进行计算.

### 1.2.4 统计模型

所谓一个问题的统计模型指的是研究该问题时所抽取样本的分布, 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.

1. 统计模型的确定

模型是对指定的样本而言的, 总体相同指标相同抽样方法不同, 产生的统计模型 (样本分布) 有可能不同. 从统计上来说, 只有规定了样本的分布, 问题才算是真正明确了.

2. 很多性质不同的问题, 可以归为同一个统计模型下.

3. 同一模型下能够提出不同的统计问题.

### 1.2.5 统计推断

1. 参数和参数空间

统计模型说的就是样本分布. 当样本分布不全已知时才存在统计推断问题.

统计学上把出现在样本分布中的未知常数称为**参数** (parameter).

在一些问题中参数虽然未知, 但是根据参数的性质可以给出参数的取值范围. 参数的取值范围称为**参数空间** (parametric space).

2. 样本分布族

样本分布既然包含位置参数, 则可能的样本分布就不止一个. 当参数取不同值的时候就得到不同的分布, 因此这些样本分布就构成一个分布族.

因此更进一步, 统计模型可以使用**样本分布族** (distribution family of the sample) 来描述. 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, 意味着可能做出更精确和更可靠的结论.

3. 统计推断

从总体中抽取一定大小的样本取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.

统计推断问题分为**参数统计推断问题**和**非参数统计推断问题**.

4. 概率论与数理统计的关系

本学年概率论后移一个学期, 数理统计老师鼓励我们退选课程等下学期再选.

可见一斑.

## 1.3 统计量

### 1.3.1 统计量的定义

数理统计的任务是用样本推断总体. 我们需要对样本进行处理获得一些有用的数量, 这些计算出的量称为统计量.

**Definition 1.3.1.** 由样本算出的量称为**统计量**. 或者说, 统计量是样本的函数.

对这一定义有如下的几点说明:

(1) 统计量只与样本有关, 不能与未知参数有关. 例如,$X \sim N(a,\sigma^2)$,$X_1,\cdots,X_n$ 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 $\sigma^2$ 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);

(3) 统计量的选取与具体问题相关.

### 1.3.2 若干常用统计量

#### 1. 样本均值

设 $X_1,\cdots,X_n$ 是从某总体 $X$ 中抽取的样本, 则称 $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n} X_i$ 为**样本均值** (sample mean). 它反映了总体均值的信息.

**2. 样本方差**

设 $X_1, \cdots, X_n$ 是从某总体 $X$ 中抽取的样本, 则称

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

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

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

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. 以后我们就使用 $S^2$(分母为 $n-1$) 定义的样本方差.

样本均值和样本方差是两个最常用的统计量, 它们具有下列性质:

(1) $\sum\limits_{i=1}^{n}(X_i - \overline{X}) = 0$.

(2) 设非零实数 $a$ 和 $b$ 为常数, 作变换 $Y_i = aX_i + b, i = 1, 2, \cdots, n$, 则 $Y_1, \cdots, Y_n$ 的样本均值 $\overline{Y} = a\overline{X} + b$, 其样本方差 $S_Y^2 = a^2 S_X^2$. 利用这一点可以简化样本中的一些计算.

(3) 对于任何常数 $c$, 有 $\sum\limits_{i=1}^{n}(X_i - c)^2 \geqslant \sum\limits_{i=1}^{n}(X_i - \overline{X})^2$, 且等号只在 $c = \overline{X}$ 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测量 $a$ 是最好的.

**3. 样本矩**

设 $X_1, \cdots, X_n$ 为从总体 $F$ 中抽取的样本, 则称

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

为**样本 $k$ 阶原点矩**. 特别 $k = 1$ 时,$a_{n,1} = \overline{X}$, 即**样本均值**. 称

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

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距统称为**样本矩** (sample moments).

**4. 二维随机向量的样本矩**

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

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2,$$

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$

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 $X_1, \cdots, X_n$ 为从总体 $F$ 中抽取的样本, 将其按大小排列为 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$, 则 $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ 称为样本 $(X_1, \cdots, X_n)$ 的**次序统计量** (order statistics),$(X_{(1)}, \cdots, X_{(n)})$ 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:

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

称为样本中位数. 它反映总体中位数的信息.

(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 $X_{(m)}$,$m = \lfloor (n+1)p \rfloor$, 此处 $\lfloor a \rfloor$ 表示实数 $a$ 的整数部分. 当 $p = 1/2$,$n$ 为奇数时, 此定义和 (1) 中的样本中位数相同.

(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 $X_1, \cdots, X_n$ 为从总体 $F$ 中抽取的样本, 则称

$$\widehat{\nu} = \frac{S_n}{\overline{X}} \tag{1.3.2}$$

为**样本变异系数** (sample coefficient 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步程度的量, 但这散布程度是以总体均值为单位来度量的.

**7. 样本偏度**

设 $X_1, \cdots, X_n$ 为从总体 $F$ 中抽取的样本, 则称

$$\widehat{\beta}_1 = \frac{m_{n,3}}{m_{n,2}^{3/2}} = \sqrt{n}\sum_{i=1}^{n}\frac{(X_i - \overline{X})^3}{\left(\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right)^{3/2}} \tag{1.3.3}$$

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 $\mu_i (i = 2, 3)$ 是总体的 $i$ 阶中心矩.

**8. 样本峰度**

设 $X_1,\cdots,X_n$ 为从总体 $F$ 中抽取的样本, 则称

$$\widehat{\beta}_2=\frac{m_{n,4}}{m_{n,2}^2}-3=n\sum_{n=1}^{n}\frac{(X_i-\overline{X})^4}{\left(\sum\limits_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^2}-3 \tag{1.3.4}$$

为**样本峰度** (sample kurtosis). **总体峰度**的定义是 $\beta_2=\mu_4/\mu_2^2-3$, 其中 $\mu_i(i=2,4)$ 是总体的 $i$ 阶中心矩.$\beta_2$ 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

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

$$F_n(x)=\begin{cases}0, & x\leqslant X_{(1)},\\ \dfrac{k}{n}, & X_{(k)}<x\leqslant X_{(k+1)}, k-1,2,\cdots,n-1,\\ 1, & X_{(n)}<x\end{cases} \tag{1.3.5}$$

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 $x=X_{(k)},k=1,2,\cdots,n$ 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.$F_n(x)$ 可以看成分布总体函数 $F(x)=P(X<x)$ 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

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

由定义可知 $F_n(x)$ 是仅依赖于样本 $X_1,X_2,\cdots,X_n$ 的函数, 因此, 它是统计量. 若记 $Y_i=I_{(-\infty,x)}(X_i),i=1,2,\cdots,n$, 则 $P(Y_i=1)=F(x),P(Y_i=0)=1-F(x)$, 且 $Y_1,Y_2,\cdots,Y_n$ i.i.d. $\sim b(1,F(A))$, 故 $nF_n(x)=\sum\limits_{i=1}^{n}Y_i\sim b(n,F(x))$, 因此对 $k=0,1,\cdots,n$ 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 $n \to \infty$ 时有

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

此处 $\xrightarrow{\mathscr{L}}$ 表示依分布收敛.

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

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

(3) 由 Borel 强大数定律, 则有

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

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

**Theorem 1.3.1.** 设 $F(x)$ 为随机变量 $X$ 的分布函数,$X_1, \cdots, X_n$ 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 $D_n = \sup\limits_{-\infty<x<\infty} |F_n(x) - F(x)|$, 则有

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

上述定理中的 $D_n$ 可以用来衡量 $F_n(x)$ 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.

# Chapter 2

# 抽样分布及若干预备知识

## 2.1 引言

没啥好说的.

## 2.2 正态总体样本均值和样本方差的分布

### 2.2.1 正态变量线性函数的分布

**Theorem 2.2.1.** 设随机变量 $X_1, \cdots, X_n$ 相互独立且 $X_k \sim N(a_k, \sigma_k^2)$,$k = 1, \cdots, n$. 令 $c_1, c_2, \cdots, c_n$ 为常数, 记 $T = \sum\limits_{k=1}^{n} c_k X_k$, 则 $T \sim N(\mu, \tau^2)$, 其中 $\mu = \sum\limits_{k=1}^{n} c_k a_k, \tau^2 = \sum\limits_{k=1}^{n} c_k^2 \sigma_k^2$.

*Proof.* 因 $X_k \sim N(a_k, \sigma_k^2)$,$k = 1, 2, \cdots, n$, 故其特征函数为

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

所以 $T$ 的特征函数为

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

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

利用此定理, 容易得到如下推论:

**Corollary 2.2.1.** 在定理2.2.1中, 若 $a_1 = \cdots = a_n = a,\sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2$, 则有

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

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

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

对独立同分布的正态变量的线性变换, 有如下结论:

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

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

则有

(1) $Y_1,\cdots,Y_n$ 也是正态随机变量, 且

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

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

(2) 特别, 当 $\boldsymbol{A} = (a_{ij})$ 为 $n$ 阶正交阵时,$Y_1,Y_2,\cdots,Y_n$ 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 $Y_1,Y_2,\cdots,Y_n$ 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.

*Proof.* 证明从此开始不再抄录. 钦此! □

### 2.2.2 正态样本均值和样本方差的分布

下述定理给出了正态样本均值和样本方差的分布以及它们的独立性.

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

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

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

(3) $\overline{X}$ 和 $S^2$ 独立.

此处 $\chi^2_{n-1}$ 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.

## 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 $X_1, X_2, \cdots, X_n$ i.i.d.$\sim$ $F$, 将其按大小排列为 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$, 则称 $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ 为样本 $(X_1, X_2, \cdots, X_n)$ 的**次序统计量**. 它的任何一部分, 如 $X_{(i)}$ 和 $(X_{(i)}, X_{(j)})\,(1 \leqslant i < j \leqslant n)$ 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 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ 几乎处处成立.

### 2.3.1 单个次序统计量的分布

设 $X_1, X_2, \cdots, X_n$i.i.d$\sim$ $F$,$f$ 为 $F$ 的密度, 求 $X_{(m)}$ 的分布, 此处 $1 \leqslant m \leqslant n$. 则可推得

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

从而其密度函数为

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

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

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

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

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

### 2.3.2 次序统计量的联合分布

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

令 $y_i=x_{(i)},i=1,\cdots,n$, 则

$$\begin{aligned}G(y_1,y_2,\cdots,y_n)&=P(X_{(1)}<y_1,X_{(2)}<y_2,\cdots,X_{(n)}<y_n)\\&=\begin{cases}n!P(X_{j_1}<y_1,X_{j_2}<y_2,\cdots,X_{j_n}<y_n), & y_1<y_2<\cdots<y_n,\\0, & \text{其他},\end{cases}\end{aligned}$$

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

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

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

$$\int\cdots\int_{a<x_1<\cdots<x_n<b} f(x_1)\cdots f(x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n=\frac{1}{n!}[F(b)-F(a)]^n,$$

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.

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

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

此处 $x=x_{(i)}$,$y=y_{(i)}$.

### 2.3.3 极差的分布

样本极差的定义在 1.3 节已经给出, 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 $J=\left|\frac{\partial(X_{(i)},X_{(j)})}{\partial V\partial Z}\right|=1$,$(X_{(i)},X_{(j)})$ 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

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

从而易知 $V$ 的密度为

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

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

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

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

### 2.3.4 均匀分布情形

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

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

设 $X_{(1)},X_{(2)},\cdots,X_{(n)}$ 为样本 $X_1,X_2,\cdots,X_n$ 的次序统计量, 由前面几节的结果可以知道次序统计量 $X_{(m)}$ 的密度函数为

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

$X_{(i)},X_{(j)}(i<j)$ 的联合密度为

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

$X_{(1)},X_{(2)},\cdots,X_{(n)}$ 的联合密度为

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

在 2.3.3 节中 $(V,Z)$ 的联合密度中, 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

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

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

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

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

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

## 2.4 $\chi^2$ 分布,$t$ 分布和 $F$ 分布

### 2.4.1 $\chi^2$ 分布

**Definition 2.4.1.** 设 $X_1,X_2,\cdots,X_n$ i.i.d. $\sim N(0,1)$, 则称 $\xi=\sum\limits_{i=1}^{n}{X_i}^2$ 是自由度为 $n$ 的 $\chi^2$ 变量, 其分布称为自由度为 $n$ 的 $\chi^2$ 分布, 记为 $\xi\sim\chi^2_n$.

$\chi^2$ 变量的概率密度函数由下面的定理给出.

**Theorem 2.4.1.** 设随机变量 $\xi$ 是自由度为 $n$ 的 $\chi^2$ 随机变量, 则其概率密度函数为

$$g_n(x)=\begin{cases}\dfrac{1}{2^{n/2}\gamma(n/2)}x^{n/2-1}e^{-x/2}, & x>0\\ 0, & x\leqslant 0.\end{cases}$$

$\chi^2_n$ 密度函数的支撑集 (即使密度函数为正的自变量的集合) 为 $(0,+\infty)$. 自由度越大, 其密度函数的图像越趋近于对称.$n=1,2$ 时单调下降趋于零,$n=3$ 开始有单峰.

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

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

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

$\chi^2$ 变量具有下列性质:

(1) 设 r.v. $\xi \sim \chi_n^2$, 则 $\xi$ 的特征函数为 $\varphi(t) = (1-2it)^{-n/2}$;

(2) r.v. $\xi$ 的均值和方差分别为 $E(\xi) = n$, $D(\xi) = 2n$.

(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 $Z_1 + Z_2 \sim \chi_{n_1+n_2}^2$.

### * 非中心 $\chi^2$ 分布

**Definition 2.4.2.** 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立,$X_i \sim N(a_i, 1)$,$a_i(i = 1, \cdots, n)$ 不全为 0. 记 $Y = \sum\limits_{i=1}^{n} X_i^2$, 则称 $Y$ 的分布是自由度为 $n$ 和非中心参数为 $\delta = \sqrt{\sum\limits_{i=1}^{n} a_i^2}$ 的非中心 $\chi^2$ 分布, 记为 $Y \sim \chi_{n,\delta}^2$. 特别当 $\delta = 0$ 时称为中心的 $\chi^2$ 分布, 即前面所述 $\chi_n^2$ 分布.

若 $Y \sim \chi_{n,\delta}^2$, 则其密度函数为

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

非中心的 $\chi^2$ 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 $Y_1, \cdots, Y_k$ 相互独立,$Y_i \sim \chi_{n_i,\delta_i}^2, i = 1, 2, \cdots, k$, 则 $\sum\limits_{i=1}^{k} Y_i \sim \chi_{n,\delta}^2$, 此处 $n = \sum\limits_{i=1}^{k} n_i$, $\delta = \sqrt{\delta_1^2 + \delta_2^2 + \cdots + \delta_k^2}$.

### 2.4.2 $t$ 分布

**Definition 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.

**Theorem 2.4.2.** 设随机变量 $T \sim t_n$, 则其密度函数为

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

$t$ 变量具有下面的性质:

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

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

特别当 $n \geqslant 2$ 时,$E(T) = 0$. 当 $n \geqslant 3$ 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 $n \to \infty$ 时 ,$t$ 变量的极限分布为 $N(0,1)$.

## * 非中心 $t$ 分布简介

**Definition 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 $\delta$ 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 $Z_n \xrightarrow{\mathscr{L}} N(\delta, 1)$;

(2) 若 $Z_n \sim t_{n,\delta}$, 则有

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

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

### 2.4.3 $F$ 分布

**Definition 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.

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

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

$F$ 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,

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

(3) 若 $T\sim t_n$, 则 $T^2\sim F_{1,n}$.

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

### * 非中心 $F$ 分布简介

**Definition 2.4.6.** 设 r.v $X\sim\chi_{m,\delta}^2, Y\sim\chi_n^2$ 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 $\delta$ 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

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

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 $Z_n \sim F_{m,n;\delta}$, $n = 1, 2, \cdots, \delta$ 固定, 则当 $n \to \infty$ 时 $Z_n \xrightarrow{\mathscr{L}} \chi^2_{m,\delta}/m$.

(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

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

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

特别当 $a_i = a, \sigma_i^2 = \sigma^2, i = 1, \cdots, n$ 时 $\sum\limits_{i=1}^{n} (X_i - a)^2 / \sigma^2 \sim \chi^2_n$.

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

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

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

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

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

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

**Corollary 2.4.4.** 设 $X_1, X_2, \cdots, X_m$ i.i.d.$\sim N(a_1, \sigma_1^2), Y_1, Y_2, \cdots, Y_n$ i.i.d.$\sim N(a_2, \sigma_2^2)$, 且样本 $X_1, X_2, \cdots, X_m$ 和 $Y_1, Y_2, \cdots, Y_n$ 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.

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

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

## 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.

**Definition 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.

**Definition 2.5.2.** 当样本容量 $n \to \infty$ 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 $n \to \infty$ 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

**Example 2.5.1.** 设 $X_1, \cdots, X_n$ i.i.d.$\sim F$, 其中总体 $F$ 有均值 $a_F$ 和方差 $\sigma_F^2$. 设 $0 < \sigma_F^2 < \infty$, $\overline{X}_n = \sum\limits_{i=1}^{n} X_i/n$ 为样本均值, 试讨论 $\overline{X}_n$ 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

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

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

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

$$\sqrt{n}(\overline{X}_n - a_F)/\sigma_F \xrightarrow{\mathscr{L}} N(0,1). \tag{2.5.2}$$

其中 $\xrightarrow{\mathscr{L}}$ 表示依分布收敛. 式 (2.5.2) 刻画了 $\overline{X}_n$ 的另一个**大样本性质**—渐近正态性. $\overline{X}_n$ 作为总体均值的估计与 $a_F$ 的偏差超过 $c$ 的概率 $P(|\overline{X}_n - a_F| > c)$, 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 $\sigma_F^2$ 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 $\overline{X}$ 的小样本性质有 $E(\overline{X}_n) = a_F$, 即估计量 $\overline{X}_n$ 的期望值等于被估计的未知参数 $a_F$. 这个性质称为 $\overline{X}_n$ 的无偏性 (即 $\overline{X}_n$ 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.

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

为证明此结果, 需要下述引理:

**Lemma 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 $n \to \infty$ 时 $X_n \xrightarrow{\mathscr{L}} X, Y_n \xrightarrow{P} c, -\infty < c < \infty$ 为常数, 则有 ①$X_n \pm Y_n \xrightarrow{\mathscr{L}} X \pm c$, ②$X_n Y_n \xrightarrow{\mathscr{L}} cX$, ③$X_n/Y_n \xrightarrow{\mathscr{L}} X/c (c \neq 0)$.

**Example 2.5.3.** 设总体 $X$ 有密度 $f, X_1, \cdots, X_n$ i.i.d.$\sim X$. 令总体 $X$ 的中位数为 $\xi_{1/2}$, 即 $\xi_{1/2}$ 满足 $\displaystyle\int_{-\infty}^{\xi_{1/2}} f(x)\mathrm{d}x = 1/2$, 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 $n \to \infty$ 时有

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

# *2.6 指数族

## 2.6.1 定义与例子

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

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

则称此分布为指数型分布族 (简称**指数族**,exponential family), 其中 $k$ 为正整数,$C(\theta) > 0$ 和 $Q_i(\theta)(i = 1, 2, \cdots, k)$ 都是定义在参数空间 $\Theta$ 上的函数,$h(x) > 0$ 和 $T_i(x)(i = 1, 2, \cdots, k)$ 都是定义在样本空间 $\mathscr{X}$ 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 $\theta$ 无关. 任一分布族若其支撑集与 $\theta$ 有关, 则族中分布不再有共同支撑集, 因而必不是**指数族**.

**Example 2.6.1.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 为从正态分布 $N(\mu, \sigma^2)$ 中抽取的简单样本, 则样本分布是指数族.

简单验证定义即可证明, 不作赘述.

**Example 2.6.2.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 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.

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

二项分布概率函数表达式可改写为

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

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.

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

泊松分布:

意义: 给定时间内事件发生的次数

基本假设: 在较短的时间段内, 事件出现的概率与等待的时间成正比 (比如在商店的营业中, 等待时间越长, 有顾客来买东西的可能性越大)

参数:$\lambda$(强度参数)

随机变量 $X$(非负整数) 服从参数为 $\lambda$ 的泊松分布:

$$P(X=x|\lambda)=\frac{e^{-\lambda}\lambda^x}{x!}(x=0,1,2,..),\mathrm{E}[X]=\lambda,\mathrm{Var}[X]=\lambda.$$

(摘自知乎, 不保证正确性.)

**Example 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.

它的密度函数的支撑集和参数是有关的, 故它不是指数族. 但若是 $\mu$ 已知, 则单参数指数分布族 $\mathrm{Exp}(1/\sigma)$ 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 $\varphi_i$ 代替 $Q_i(\theta)$, 而将 $C(\theta)$ 表成 $\varphi$ 的函数 $C^*(\varphi)$,$\varphi=(\varphi_1,\varphi_2,\cdots,\varphi_k)$, 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 $\varphi$ 为 $\theta_i$,$i=1,2,\cdots,k$, 得到如下**指数族的自然形式** (或称为标准形式) 的定义.

**Definition 2.6.2.** 如果指数族由下列形式:

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

则称它为指数族的**自然形式** (natural form). 此时集合

$$\Theta^* = \left\{(\theta_1, \theta_2, \cdots, \theta_k) : \int_{\mathscr{X}} \exp\left\{\sum_{i=1}^{k} \theta_i T_i(x)\right\} h(x)\mathrm{d}x < \infty\right\} \tag{2.6.11}$$

称为**自然参数空间** (natural parametric space).

**Example 2.6.6.** 把例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.

经过较为繁琐的计算可得其自然形式为

$$f(\boldsymbol{x}, \varphi) = C^*(\varphi) \exp\left\{\varphi_1 T_1(\boldsymbol{x}) + \varphi_2 T_2(\boldsymbol{x})\right\} h(\boldsymbol{x}), \tag{2.6.12}$$

其自然参数空间为

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

**Example 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.

经过不繁琐的计算可以获得其自然形式:

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

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

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

### 2.6.3 指数族的性质

自然参数空间具有下述重要性质.

**Theorem 2.6.1.** 在指数族的形式下, 自然参数空间为凸集.

**Theorem 2.6.2.** 设指数族的自然行驶中, 自然空间参数有内点, 其内点集为 $\Theta_0$. 设 $g(x)$ 为任一实函数, 使得积分

$$G(\theta) = \int_{\mathscr{X}} \exp\left\{\sum_{j=1}^{k} \theta_j T_j(x)\right\} h(x)\mathrm{d}x$$

在 $\Theta_0$ 内存在有限, 则 $G(\theta)$ 的任意阶偏导数在 $\Theta_0$ 内存在且可在积分号下求得, 即

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

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

这些定理的证明书上没写.

## 2.7 充分统计量

### 2.7.1 引言与定义

统计量是对样本信息的提取和简化. 它希望达到: ①简化程度高;②信息损失少. 一个统计量能够几种样本中的多少信息, 和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**.

这是个例子:

设 $\boldsymbol{X} = (X_1,\cdots,X_n)$ 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 $T(\boldsymbol{X}) = \sum\limits_{i=1}^{n} X_i$, 如果目的仅仅为了推断 $\theta$, 则 $T(\boldsymbol{X})$ 是充分统计量. 从直观上看, 知道了 $T(\boldsymbol{X})$ 推断 $\theta$ 的效果, 与知道样本 $(X_1,\cdots,X_n)$ 一样, 因为 $T(\boldsymbol{X})$ 表示事件 $A$ 在 $n$ 次独立的 Bernoulli 试验中成功的次数. 而 $(X_1,\cdots,X_n)$ 比 $T(\boldsymbol{X})$ 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 $\theta$ 不会有更多的用处. 故从直观上, 知道了 $T(\boldsymbol{X})$ 后与知道样本 $(X_1,\cdots,X_n)$ 对推断 $\theta$ 的效果相同.

一大堆原始资料, 加工成统计量 $T(\boldsymbol{X})$ 后, 一般来说在信息上会有损失. 但也有可能, 将样本 $\boldsymbol{X}$ 加工成 $T(\boldsymbol{X})$ 时抓住了问题的实质, 即 $T(\boldsymbol{X})$ 中保留了样本 $\boldsymbol{X}$ 中所含参数 $\theta$ 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 $\boldsymbol{X}$ 也能恢复参数 $\theta$ 的信息, 则称此统计量为充分的. 仍看上例, 样本 $\boldsymbol{X}$ 的分布为

$$P_\theta(\boldsymbol{X}=\boldsymbol{x}) = P_\theta(X_1=x_1,\cdots,X_n=x_n) = \theta^t(1-\theta)^{n-t}, \tag{2.7.1}$$

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

$$P_\theta(T(\boldsymbol{X})=t) = \binom{n}{t}\theta^t(1-\theta)^{n-t}, t=0,1,2,\cdots,n, 0<\theta<1. \tag{2.7.2}$$

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

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

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

故 $T(\boldsymbol{X})$ 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 $P_\theta(\boldsymbol{X}\in A|T=t)$ 与 $\theta$ 无关, 其中 $A$ 为任一事件. 因此得到如下的定义.

**Definition 2.7.1.** 设样本 $\boldsymbol{X}$ 的分布族为 $\{f(\theta,\boldsymbol{x}),\theta\in\Theta\}$,$\Theta$ 是参数空间. 令 $T=T(\boldsymbol{X})$ 为一统计量, 若在已知 $T$ 的条件下, 样本 $\boldsymbol{X}$ 的条件分布与 $\theta$ 无关, 则称 $T(\boldsymbol{X})$ 为 $\theta$ 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.

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

**Example 2.7.2.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从正态总体 $N(0,1)$ 中抽取的简单样本, 则 $T(\boldsymbol{X})=\displaystyle\sum_{i=1}^{n}\frac{X_i}{n}=\overline{X}$ 为 $\theta$ 的充分统计量.

**Example 2.7.3.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从指数分布 $\mathrm{Exp}(\theta)$ 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 $T(\boldsymbol{X})=\sum_{i=i}^{n}X_i$ 为 $\theta$ 的充分统计量.

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

### 2.7.2 充分性的判别准则—因子分解定理

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

$$f(\boldsymbol{x},\theta)=g(t(\boldsymbol{x}),\theta)h(\boldsymbol{x}) \tag{2.7.5}$$

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

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

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

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

**Example 2.7.6.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从总体 $b(1,\theta)$ 中抽取的简单样本, 则 $T(\boldsymbol{X})=\sum\limits_{i=1}^{n}X_i$ 是 $\theta$ 的充分统计量.

**Example 2.7.7.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从均匀分布 $U(0,\theta)$ 中抽取的简单样本, 则 $T(\boldsymbol{X})=X_n=\max\{X_1,\cdots,X_n\}$ 为 $\theta$ 的充分统计量.

**Example 2.7.8.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从均匀分布 $U(\theta-1/2,\theta+1/2)$ 中抽取的简单样本, 其中 $-\infty<\theta<+\infty$,$\theta$ 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 $\overline{X}$ 不是充分统计量.

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

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

### *2.7.3 极小充分统计量

一个分布族 $\mathscr{F}$ 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 $\theta$ 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 $T_1(\boldsymbol{X})=\left(\sum\limits_{i=1}^{m}X_i,\sum\limits_{i=m+1}^{n}X_i\right)$, 再进一步加工得到一维统计量 $T_2=\sum\limits_{i=1}^{n}X_i$. 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.

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

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必是极小充分的, 反之不必对.

# *2.8 完全统计量

## 2.8.1 定义和例子

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

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

的 $\varphi(T(X))$, 都有

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

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

由该定义可见, 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**.

**Remark 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

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

积分(2.8.3)形式上可看成 "$\varphi$ 与 $g_\theta$ 正交". 于是条件(2.8.3)⇒(2.8.2)可说成是 "若 $\varphi$ 与函数系 $\{g_\theta,\theta\in\Theta\}$ 正交, 则 $\varphi$ 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)⇒(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.

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

**Example 2.8.2.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从均匀分布 $U(0,\theta)$ 中抽取的简单样本，则 $T(\boldsymbol{X})=X_{(n)}=\max\{X_1,\cdots,X_n\}$ 是完全统计量.

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

### 2.8.2 指数族中统计量的完全性

**Theorem 2.8.1.** 设样本 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 的概率函数

$$f(\boldsymbol{x},\boldsymbol{\theta})=C(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^{k}\theta_iT_i(\boldsymbol{x})\right\}h(\boldsymbol{x}),\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)\in\Theta^*$$

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

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

**Example 2.8.5.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $(\overline{X},S^2)$ 为完全统计量，其中 $\overline{X},S^2$ 分布为样本均值和样本方差.

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

### 2.8.3 有界完全统计量及其性质

**Definition 2.8.2.** 若对任何满足

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

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

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

则称 $T(X)$ 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.

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

**Corollary 2.8.1.** 设样本 $\boldsymbol{X}$ 的分布族为指数族, 即

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

而自然参数空间 $\Theta^*$ 作为 $R_k$ 的子集有内点. 若 r.v. $V(\boldsymbol{X})$ 的分布族与 $\theta$ 无关, 则 $V(\boldsymbol{X})$ 与 $T(\boldsymbol{X})=(T_1(\boldsymbol{X}),\cdots,T_k(\boldsymbol{X}))$ 独立.

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

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

# Chapter 3

# 点估计

## 3.1 引言

### 3.1.1 参数估计问题

**参数估计** (parameter estimation) 问题通常有两类: 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.

### 3.1.2 点估计

**Definition 3.1.1.** 设 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 为从某个总体中抽取的样本,$\hat{g}(\boldsymbol{X}) = \hat{g}(X_1, \cdots, X_n)$ 是样本的函数, 用 $\hat{g}(\boldsymbol{X})$ 作为 $g(\theta)$ 的估计, 称为**点估计** (point estimation).

对于同一个未知参数 $\theta$, 我们在同一组样本中可以做出不同估计: 极值平均, 样本平均, 中位数, 等等. 这些估计当然有好有坏, 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.

### 3.1.3 点估计的优良性准则

**1. 无偏性**

我们希望估计量 $\hat{\theta}(\boldsymbol{X})$ 的平均值与 $\theta$ 越接近越好, 即 $E(\hat{\theta}(\boldsymbol{X}) - \theta)$ 越小越好. 这就是无偏性的定义的来源.

**Definition 3.1.2.** 设 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 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 $\Theta$ 上的已知函数.$\hat{g}(\boldsymbol{X})=\hat{g}(X_1,\cdots,X_n)$ 是 $g(\theta)$ 的一个估计量, 如果

$$E_\theta(\hat{g}(\boldsymbol{X}))=g(\theta), \qquad \theta\in\Theta,$$

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

$$\lim_{n\to\infty} E_\theta(\hat{g}_n(\boldsymbol{X}))=g(\theta), \qquad \theta\in\Theta,$$

则称 $\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 的**渐进无偏估计** (asymptotically unbiased estimation).

无偏性的含义有两个: 一是系统误差为零, 只会有随机误差. 二是在多次重复使用的情况下接近真值 $g(\theta)$ 的估计.

**Example 3.1.1.** 设 $X_1,\cdots,X_n$ 是取自期望为 $\mu$, 方差为 $\sigma^2$ 的总体的一个样本. 显然样本均值 $\overline{X}$ 是 $\mu$ 的无偏估计, 则样本方差 $S^2=\sum\limits_{i=1}^{n}(X_i-\overline{X})^2/(n-1)$ 是 $\sigma^2$ 的无偏估计.

**2. 有效性**

有时, 某参数的无偏估计可能不止一个. 那么选用哪一个无偏估计比较好呢? 这就要求我们讨论估计量的**有效性** (efficiency). 设 $\widehat{\theta}_1$ 和 $\widehat{\theta}_2$ 为 $\theta$ 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距—方差, 方差越小越好.

**Definition 3.1.3.** 设 $\hat{g}_1(\boldsymbol{X})=\hat{g}_1(X_1,\cdots,X_n)$ 和 $\hat{g}_2(\boldsymbol{X})=\hat{g}_2(X_1,\cdots,X_n)$ 为 $g(\theta)$ 的两个不同**无偏估计量**, 若

$$D_\theta(\hat{g}_1(\boldsymbol{X}))\leqslant D_\theta(\hat{g}_2(\boldsymbol{X})), \qquad \text{一切}\theta\in\Theta,$$

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

由这个定义可以看出, 在均值相等的情况下, 方差越小的估计量越有效.

**3. 相合性**

**Definition 3.1.4.** 设对每个自然数 $n$,$\hat{g}(\boldsymbol{X})=\hat{g}_n(X_1,\cdots,X_n)$ 是 $g(\theta)$ 的一个估计量, 若 $\hat{g}_n(\boldsymbol{X})$ 依概率收敛到 $g(\theta)$, 即对任何 $\theta\in\Theta$ 及 $\epsilon>0$ 有

$$\lim_{n\to\infty} P_\theta\left(|\hat{g}_n(\boldsymbol{X})-g(\theta)|\geqslant\varepsilon\right)=0,$$

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

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

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

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

则称 $\hat{g}_n(\boldsymbol{X})$ 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**.

强相合 $\Rightarrow$ 弱相合,$r>0$ 有 $r$ 阶矩相合 $\Rightarrow$ 弱相合. 强相合和 $r$ 阶矩相合之间并没有包含关系.

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

## 3.2 矩估计

### 3.2.1 矩法和距估计量

设 $X_1,\cdots,X_n$ 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**

$$a_{nk}=\frac{1}{n}\sum_{i=1}^n X_i^k,\quad k=1,2,\cdots \tag{3.2.1}$$

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 $a_{n1}=\overline{X}$.

**样本 $k$ 阶中心矩**

$$m_{nk}=\frac{1}{n}\sum_{i=1}^n\left(X_i-\overline{X}\right)^k,\quad k=2,3,\cdots \tag{3.2.2}$$

是总体 $k$ 阶中心矩 $\mu_k=E\left[X-E\left[X\right]\right]^k$ 的 "自然" 的**矩估计量**. 特别地,

总体方差 $\mu_2=E\left[X-E[X]\right]^2$ 的 "自然" 的**矩估计量**是 $m_{n2}=S_n^2=\sum_{i=1}^n\left(X_i-\overline{X}\right)^2/n$, 它与样本方差 $S^2=\sum_{i=1}^n\left(X_i-\overline{X}\right)^2/(n-1)$ 只差一个常数因子.

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

$$E\left[a_{nk}\right]=\frac{1}{n}\sum_{i=1}^n E(X_i^k)=\frac{1}{n}\sum_{i=1}^n\alpha_k=\alpha_k. \tag{3.2.3}$$

但用 $m_{nk}$ 估计 $\mu_k$, 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.

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

下面给出**矩法**即**矩估计量**的定义.

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

$$g(\theta) = G(\alpha_1, \cdots, \alpha_k; \mu_2, \cdots, \mu_s). \tag{3.2.7}$$

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

$$\hat{g}(\boldsymbol{X}) = G(a_{n1}, \cdots, a_{nk}; m_{n2}, \cdots, m_{ns}), \tag{3.2.8}$$

则 $\hat{g}(\boldsymbol{X})$ 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

**Example 3.2.2.** 设 $X_1, \cdots, X_n$ 是从具有成功概率 $\theta$ 的两点分布总体 $b(1, \theta)$ 中抽取的简单样本, 求 $\theta$ 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 $\theta$ 的据估计量就是 $\hat{\theta}(X_1, \cdots, X_n) = \overline{X}$, 按定义 $g(\theta)$ 的矩估计量是

$$\hat{g}(X_1, \cdots, X_n) = \overline{X}(1 - \overline{X}).$$

**Example 3.2.3.** 设 $X_1, \cdots, X_n$ 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 $\theta_1$ 和 $\theta_2$ 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

$$\theta_1 = \alpha_1 - \sqrt{3\mu_2}, \quad \theta_2 = \alpha_1 + \sqrt{3\mu_2}.$$

将上式中的 $\alpha_1$ 和 $\mu_2$ 分别用 $\overline{X}$ 和 $m_{n2}$ 代入得

$$\begin{aligned}\hat{\theta}_1(X_1,\cdots,X_n) =& \overline{X}-\sqrt{3m_{n2}}=\overline{X}-\sqrt{3}S_n,\\ \hat{\theta}_2(X_1,\cdots,X_n) =& \overline{X}+\sqrt{3}S_n,\end{aligned}$$

其中 $S_n^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(X_i-\overline{X}\right)^2$.

**Example 3.2.4.** 设总体分布有概率密度

$$f(x,\theta)=\begin{cases}2\sqrt{\dfrac{\theta}{\pi}}\exp\left\{-\theta x^2\right\}, & x>0\\ 0, & x\leqslant 0,\end{cases}$$

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 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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

$$\alpha_1=E(X)=2\sqrt{\frac{\theta}{\pi}}\int_0^{\infty}xe^{-\theta x^2}\mathrm{d}x=\frac{1}{\sqrt{\pi\theta}},$$

解得 $g(\theta)=1/\theta=\pi\alpha_1^2$, 将 $\alpha_1$ 用 $\overline{X}$ 代替得

$$\widehat{g}_1(\boldsymbol{X})=\pi\overline{X}^2.$$

另一方面, 由另一方程

$$\alpha_2=E_\theta(X^2)=2\sqrt{\frac{\theta}{\pi}}\int_0^{\infty}x^2e^{-\theta x^2}dx=\frac{1}{2\theta},$$

解得 $g(\theta)=1/\theta=2\alpha_2$, 将 $\alpha_2$ 用 $a_{n2}=\sum\limits_{i=1}^{n}{X_i}^2/n$ 代入, 得 $g(\theta)$ 的矩估计

$$\hat{g}_2(\boldsymbol{X})=2\,a_{n2}=\frac{2}{n}\sum_{i=1}^{n}X_i^2.$$

由此可见矩估计不唯一. 这两个矩估计量 $\hat{g}_1(X)$ 和 $\hat{g}_2(X)$ 中哪一个更好? 以后可以证明基于 $a_{n2}$ 的估计量 $\hat{g}_2(\boldsymbol{X})$, 在 $g(\theta)$ 一切无偏估计类中是方差最小者. 基于 $a_{n1}$ 的估计 $\hat{g}_1(\boldsymbol{X})$ 不是 $g(\theta)$ 的无偏估计.

**Example 3.2.5.** 设 $X_1,\cdots,X_n$ 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 $\alpha_1, \mu_2, \mu_3$ 和 $\mu_4$ 分别用它们的矩估计 $a_{n1}, m_{n2}, m_{n3}$ 和 $m_{n4}$ 代替, 得到 $\nu, \beta_1, \beta_2$ 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

$$\begin{aligned}
\hat{\nu}(X_1, \cdots, X_n) &= \frac{\sqrt{m_{n2}}}{a_{n1}}, \\
\hat{\beta}_1(X_1, \cdots, X_n) &= \frac{m_{n3}}{m_{n2}^{3/2}}, \\
\hat{\beta}_2(X_1, \cdots, X_n) &= \frac{m_{n4}}{m_{n2}^2} - 3.
\end{aligned}$$

**Example 3.2.6.** 设总体分布有概率密度

$$f(x, \theta) = \begin{cases} \dfrac{\theta_2}{\Gamma((1+\theta_1)/\theta_2)} x^{\theta_1} \exp\{-x^{\theta_2}\}, & x > 0, \\ 0, & x \leqslant 0, \end{cases}$$

参数 $\theta = (\theta_1, \theta_2)$ 的变化范围 $-1 < \theta_1 < \infty, \theta_2 > 0$. 设 $X = (X_1, \cdots, X_n)$ 为抽自此总体的简单随机样本, 求 $\theta_1$ 和 $\theta_2$ 的矩估计量.

**解:** 由简单计算得到总体分布的前两阶矩为

$$\begin{aligned}
\alpha_1 &= \Gamma\left(\frac{2+\theta_1}{\theta_2}\right) / \Gamma\left(\frac{1+\theta_1}{\theta_2}\right), \\
\alpha_2 &= \Gamma\left(\frac{3+\theta_1}{\theta_2}\right) / \Gamma\left(\frac{1+\theta_1}{\theta_2}\right).
\end{aligned} \tag{3.2.9}$$

按矩估计方法, 用 $a_{n1}$ 和 $a_{n2}$ 分别代替式(3.2.9)中的 $\alpha_1$ 和 $\alpha_2$, 用 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别代替 $\theta_1$ 和 $\theta_2$, 得到如下的方程组:

$$\begin{aligned}
a_{n1} &= \Gamma\left(\frac{2+\hat{\theta}_1}{\hat{\theta}_2}\right) \Big/ \Gamma\left(\frac{1+\hat{\theta}_1}{\hat{\theta}_2}\right), \\
a_{n2} &= \Gamma\left(\frac{3+\hat{\theta}_1}{\hat{\theta}_2}\right) \Big/ \Gamma\left(\frac{1+\hat{\theta}_1}{\hat{\theta}_2}\right).
\end{aligned}$$

其解就是 $\theta_1$ 和 $\theta_2$ 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.

矩估计方法也可用于多维样本, 请看下例.

**Example 3.2.7.** 设 $X_i, Y_i, i=1,2,\cdots,n$ 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 $\sigma_{12}$ 和相关系数 $\rho$ 的矩估计.

**解:** 按定义 $\sigma_{12}$ 的矩估计量是样本协方差, 即

$$m_{12} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}), \tag{3.2.10}$$

其中 $\overline{X} = \sum\limits_{i=1}^{n} X_i/n, \overline{Y} = \sum\limits_{j=1}^{n} Y_j/n$, 而 $\rho$ 的矩估计量是样本相关系数, 即

$$r = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum\limits_{i=1}^{n}(X_i - \overline{X})^2 \sum\limits_{j=1}^{n}(Y_i - \overline{Y})^2}}. \tag{3.2.11}$$

### 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. 特点: 直观性强, 简便易行, 不需要事先知道总体的分布, 在一定条件下具有相合性和渐近正态性.

缺点: 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, 矩估计量不具有唯一性.

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.

(1) 样本 $k$ 阶原点矩 $a_{nk}$ 是总体 $k$ 阶原点矩 $\alpha_k (k=1,2,\cdots)$ 的无偏估计, 式(3.2.3)已经给出证明.

(2) 对 $k \geqslant 2$, 样本 $k$ 阶中心矩不是总体 $k$ 阶中心矩的无偏估计.

(3) 当待估函数 $g(\theta)$ 为总体分布一些矩的函数时, 即

$$g(\theta) = G(\alpha_1, \cdots, \alpha_k; \mu_2, \cdots, \mu_s) \tag{3.2.14}$$

时, 矩估计

$$\hat{g}(X_1, \cdots, X_n) = G(a_{n1}, \cdots, a_{nk}; m_{n2}, \cdots, m_{ns}) \tag{3.2.15}$$

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

$$g(\theta) = h(\alpha_1, \cdots, \alpha_m) = \sum_{i=1}^{m} c_i \alpha_i$$

时, 其矩估计

$$\hat{g}(X_1, \cdots, X_n) = \sum_{i=1}^{m} c_i a_{ni}$$

是 $g(\theta)$ 的无偏估计.

(4) 矩估计一般具有渐近无偏性. $m_{n\nu}\,(\nu\geqslant 2)$ 是总体 $\nu$ 阶中心矩 $\mu_\nu$ 的渐近无偏估计, 如

$$\lim_{n\to\infty}E(m_{n3})=\lim_{n\to\infty}\frac{(n-1)(n-2)}{n^2}\mu_3=\mu_3.$$

**Example 3.2.8.** 设 $X_1,\cdots,X_n$ i.i.d. $\sim$ 指数分布 $\mathrm{Exp}(\lambda)$, 其密度为 $f(x,\lambda)=\lambda e^{-\lambda x}\cdot I_{[x>0]}$. 求 $\lambda$ 的矩估计量 $\hat{\lambda}$ 并讨论它的无偏性.

**解:** 由于 $\alpha_1=\int_0^\infty xf_\lambda(x)\mathrm{d}x=1/\lambda$, 解得 $\lambda=1/\alpha_1$, 将 $\alpha_1$ 用其矩估计量 $\overline{X}=a_{n1}$ 代入, 得到 $\lambda$ 的矩估计量为

$$\hat{\lambda}(X_1,\cdots,X_n)=\frac{1}{\overline{X}}.$$

由于 $Y=\sum_{i=1}^n X_i\sim$ Gamma 分布 $\Gamma(n,\lambda)$,

$$E(\hat{\lambda})=nE\left(\frac{1}{Y}\right)=n\int_0^\infty\frac{1}{y}\cdot\frac{\lambda^n}{\Gamma(n)}y^{n-1}e^{-\lambda y}\mathrm{d}y=\frac{n}{n-1}\lambda,$$

可见 $\hat{\lambda}(X_1,\cdots,X_n)=1/\overline{X}$ 不是 $\lambda$ 的无偏估计. 因此矩估计量不一定都具有无偏性. 但 $\lim\limits_{n\to\infty}E(\hat{\lambda})=\lim_{n\to\infty}[n/(n-1)\cdot\lambda]=\lambda$, 故 $\hat{\lambda}(X_1,\cdots,X_n)$ 是 $\lambda$ 的**渐近无偏估计**. 对 $\lambda$ 略作修正, 可得 $\lambda$ 的一个无偏估计

$$\hat{\lambda}^*(X_1,\cdots,X_n)=\frac{n-1}{n}\cdot\hat{\lambda}(X_1,\cdots,X_n)=\frac{n-1}{n}\cdot\frac{1}{\overline{X}}.$$

### 3.2.4 矩估计的相合性

估计量的几种相合性的定义在 3.1 节中已经给出.

(1) 样本 $k$ 阶原点矩是总体 $k$ 阶原点矩的强相合估计. 设 $X_1,\cdots,X_n$ 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,$\alpha_k$ 为总体的 $k$ 阶原点矩. 由独立同分布场合的科尔莫戈洛夫强大数定律可知 $a_{nk}\xrightarrow{a.s.}\alpha_k$,$k=1,2,\cdots$, 即

$$P\left(\lim_{n\to\infty}a_{nk}=\alpha_k\right)=1,\quad k=1,2,\cdots.$$

(2) 样本 $k$ 阶中心矩是总体 $k$ 阶中心距的强相合估计, 这一结论的证明用到下列结论.

**Lemma 3.2.1.** 设函数 $f(y_1,\cdots,y_k)$ 在 $(c_1,c_2,\cdots,c_k)$ 处连续, 若 $y_{n1}\xrightarrow{\text{a.s.}}c_1,\cdots,y_{nk}\xrightarrow{\text{a.s.}}c_k$, 则 $f(y_{n1},\cdots,y_{nk})\xrightarrow{\text{a.s.}}f(c_1,c_2,\cdots,c_k)$.

证明剩余部分略.

(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.

**Theorem 3.2.1.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 为从总体 $F$ 中抽取的简单随机样本, 待估函数 $g(\theta) = G(\alpha_1, \cdots, \alpha_k, \mu_2, \cdots, \mu_s)$,
其矩估计量为 $\hat{g}_n(\boldsymbol{X}) = G(a_{n1}, \cdots, a_{nk}, m_{n2}, \cdots, m_{ns})$, 且 $G$ 为其变元的连续函数, 则称 $\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 之**强相合估计**.

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 $\overline{X}$ 和样本方差 $S^2$ 分别是 $a$ 和 $\sigma^2$ 的强相合估计. 也不难证明 $S^2$ 是 $\sigma^2$ 的均方相合估计. 其实对任何 $r > 0, S^2$ 是 $\sigma^2$ 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.

**Example 3.2.9.** 证明例3.2.3中 $\theta_1$ 和 $\theta_2$ 的矩估计量分别是 $\theta_1$ 和 $\theta_2$ 的强相合估计.

**证:** 在例3.2.3中已求出 $\theta_1$ 和 $\theta_2$ 的矩估计量分别为

$$\hat{\theta}_1 = \hat{\theta}_1(X_1, \cdots, X_n) = \overline{X} - \sqrt{3}S_n,$$
$$\hat{\theta}_2 = \hat{\theta}_2(X_1, \cdots, X_n) = \overline{X} + \sqrt{3}S_n.$$

由强大数定律可知 $\overline{X}$ 和 $S_n^2 = \sum\limits_{i=1}^{n}(X_i - \overline{X})^2/n$ 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 $\overline{X}$ 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

我们尝试给出矩估计是**相合渐近正态估计**. 下面首先给出定义.

**Definition 3.2.2.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从总体 $\{f(x, \theta), \theta \in \Theta\}$ 中抽取的简单样本,$\hat{g}_n(\boldsymbol{X})$ 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 $\Theta$ 上的函数 $A_n(\theta)$ 和 $B_n(\theta)$, 其中 $B_n(\theta)$ 在 $\Theta$ 上处处大于 0, 使当 $n \to \infty$ 时,

$$\frac{\hat{g}_n(\boldsymbol{X}) - A_n(\theta)}{B_n(\theta)} \xrightarrow{\mathscr{L}} N(0, 1),$$

且 $\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 的弱相合估计, 则称 $\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 的**相合渐近正态估计** (consistent asymptotic normal estimation, CAN 估计).

也就是说 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 $X_1,\cdots,X_n$ 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 $\Theta$ 上的实函数, 它可以表示为形式:$g(\theta)=G(\alpha_1,\cdots,\alpha_k)$ (若 $G$ 是 $\alpha_1,\cdots,\alpha_k,\mu_2,\cdots,\mu_s$ 的函数, 不妨令 $s\leqslant k$, 可将 $\mu_j$ 用 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 表出, 则 $G$ 仍可表为 $\alpha_1,\cdots,\alpha_k$ 的函数), 而 $\hat{g}(\boldsymbol{X})=\hat{g}(X_1,\cdots,X_n)=G(a_{n1},\cdots,a_{nk})$ 为 $g(\theta)$ 的矩估计. 再设总体的 $2k$ 阶原点矩存在, 且 $G$ 对其各变量的一个偏导数存在且连续, 令

$$b_{ij}=\alpha_{i+j}-\alpha_i\alpha_j,i,j=1,2,\cdots,k;\boldsymbol{B}=(b_{ij})\text{为 } k\times k \text{ 的方阵},$$
$$d_i=\frac{\partial G(\alpha_1,\cdots,\alpha_k)}{\partial\alpha_i},i=1,2,\cdots,k;\boldsymbol{d}=(d_1,\cdots,d_k)',$$
$$b^2=\boldsymbol{d}'\boldsymbol{B}\boldsymbol{d}.$$

**Theorem 3.2.2.** 在上述记号和条件下,$\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 的 CAN 估计, 即 $\hat{g}_n(X)$ 为 $g(\theta)$ 的弱相合估计, 且有

$$\sqrt{n}(\hat{g}_n(X)-G(\alpha_1,\cdots,\alpha_k))\xrightarrow{\mathscr{L}}N(0,b^2),n\to\infty.$$

**证:** 由定理3.2.1已知 $\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)=G(\alpha_1,\cdots,\alpha_k)$ 的强相合估计, 显然它也是弱相合估计.

(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 $\alpha_1$, 这样得到的 $g(\theta)$ 的表达式比较简单. 若将中心矩用原点矩表出,$g(\theta)$ 的表达式则会显得复杂, 因此有必要给出这种情形下渐进正态性的结果. 一般地, 将 $g(\theta)$ 表达成如下形式:

$$g(\theta)=H(\alpha_1,\mu_{t_2},\cdots,\mu_{t_r}),\tag{3.2.16}$$

其矩估计量为 $H(\overline{X},m_{nt_2},\cdots,m_{nt_r})$, 使用与定理3.2.2基本相同的证明方法, 可得如下结果.

**Theorem 3.2.3.** 设式(3.2.16)中的函数 $H$ 在点 $\alpha_1,\mu_{t_2},\cdots,\mu_{t_r}$ 的邻域内有一阶偏导数, 且此偏导数在点 $(\alpha_1,\mu_{t_2},\cdots,\mu_{t_r})$ 处连续, 则有

$$\sqrt{n}\left(H(\overline{X},m_{nt_2},\cdots,m_{nt_r})-H(\alpha_1,\mu_{t_2},\cdots,\mu_{t_r})\right)\xrightarrow{\mathscr{L}}N(0,b^2).\tag{3.2.17}$$

此处

$$b^2=\sum_{i=1}^{r}\sum_{j=1}^{r}\sigma_{ij}H_iH_j,\tag{3.2.18}$$

其中

$$H_1=\frac{\partial H}{\partial\alpha_1},\quad H_i=\frac{\partial H}{\partial\mu_{t_i}},\quad i=2,3,\cdots,r;$$
$$\sigma_{11}=\mu_2,\quad \sigma_{1i}=\sigma_{i1}=\mu_{t_i+1}-t_i\mu_{t_i-1}\mu_2,i=2,\cdots,r;$$
$$\begin{aligned}\sigma_{ij}=&\mu_{t_i+t_j}-t_i\mu_{t_i-1}\mu_{t_j+1}-t_j\mu_{t_i+1}\mu_{t_j-1}-\mu_{t_i}\mu_{t_j}\\&+t_it_j\mu_2\mu_{t_i-1}\mu_{t_j-1},\quad i,j=2,3,\cdots,r.\end{aligned}$$

如果 $g(\theta)$ 有 $H(\mu_{t_2},\cdots,\mu_{t_r})$ 的形状, 即与 $\alpha_1$ 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 $\sum_{i=2}^{r}\sum_{j=2}^{r}\sigma_{ij}H_iH_j$ 即可.

**Example 3.2.10.** 继续考虑例3.2.5. 被估计的量 $g(\theta)$ 为偏度 $\beta_1$, 峰度 $\beta_2$ 和变异系数 $\nu$, 其矩估计量由例3.2.5可知为

$$\hat{\beta}_1=\frac{m_{n3}}{m_{n2}^{3/2}},\quad \hat{\beta}_2=\frac{m_{n4}}{m_{n2}^2}-3, \hat{\nu}=\sqrt{m_{n2}}/\overline{X}.$$

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

$$\begin{aligned}&\sqrt{n}(\hat{\beta}_1-\beta_1)\xrightarrow{\mathscr{L}}N(0,6),\\&\sqrt{n}(\hat{\beta}_2-\beta_2)\xrightarrow{\mathscr{L}}N(0,24),\\&\sqrt{n}(\hat{\nu}-\nu)\xrightarrow{\mathscr{L}}N(0,\nu^2/2+\nu^4).\end{aligned}$$

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 $\beta_1,\beta_2$ 的大样本推断有用.

## 3.3 极大似然估计

### 3.3.1 引言和定义

极大似然法是在参数分布族场合下常用的参数估计方法.

**Definition 3.3.1.** 设 $f(\boldsymbol{x},\theta)=f(x_1,\cdots,x_n,\theta)$ 为样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 的概率函数. 当 $\boldsymbol{x}$ 固定时, 把 $f(\boldsymbol{x},\theta)$ 看成 $\theta$ 的函数, 称为**似然函数** (likelihood function), 记为

$$L(\theta,\boldsymbol{x})=f(\boldsymbol{x},\theta),\quad \theta\in\Theta,\boldsymbol{x}\in\mathscr{X},\tag{3.3.1}$$

其中 $\Theta$ 为参数空间,$\mathscr{X}$ 为样本空间. 称 $\log L(\theta,\boldsymbol{x})$ 为**对数似然函数**, 记为 $l(\theta,\boldsymbol{x})$.

**Note 3.3.1.** 似然函数和概率密度是同一表达式(3.3.1), 但表示两种不同含义.

当把 $\theta$ 固定, 将其看成定义在样本空间 $\mathscr{X}$ 上的函数时, 称为概率密度; 当把 $\boldsymbol{x}$ 固定, 为其看成定义在参数空间 $\Theta$ 上的函数时, 称为似然函数. 这是两个不同的概念.

**Example 3.3.1.** 设罐子里有许多的黑球和红球. 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是 1/4 或 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:** 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 $\theta$, 此处 $\theta$ 只取可能的两个值 $\theta_1 = 1/4$ 和 $\theta_2 = 3/4$ 之一. 记 $X = \sum\limits_{i=1}^{n} X_i$, 则样本分布族 $\mathscr{F} = \{f(x,\theta_1), f(x,\theta_2)\}$, 其中 $f(x,\theta_1)$ 为 $b(n,x,\theta_1)$,$f(x,\theta_2)$ 为 $b(n,x,\theta_2)$. 要根据抽样结果对 $\theta$ 作出估计, 即 $\theta$ 取值为 1/4 还是 3/4? 或者说样本来自总体 $f(x,\theta_1)$ 还是 $f(x,\theta_2)$?

显然, 当样本 $X$ 给定时, 似然函数为

$$L(\theta, x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, 2, \cdots, n.$$

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 $x = \sum\limits_{i=1}^{3} x_i$ 取值为 $0, 1$ 时认为样本来自总体 $f(x,\theta_1)$, 即取参数 $\theta$ 的估计值为 $\theta_1 = 1/4$; 当 $x = 2, 3$ 时认为样本来自总体 $f(x,\theta_2)$, 即取 $\theta$ 的估计值为 $\theta_2 = 3/4$.

上例说明了"似然估计"的运作模式: 我们获取样本后针对该样本进行计算, 在该样本取值下计算参数的分布函数 (此时的参数才是我们的估计中的随机变量), 最终选定可能性最大的参数. 换句话说, 就是比较真参数取值的"似然性".

更一般地, 若样本 $\boldsymbol{X}$ 的分布族 $\mathscr{F} = \{f(x,\theta), \theta \in \Theta\}$, 参数空间 $\Theta$ 为 $\mathbb{R}$ 的有限子集或无限子集. 当样本 $\boldsymbol{x}$ 给定时, 若 $\hat{\theta}^*$ 使似然函数 $L(\hat{\theta}^*, x)$ 为似然函数的集合 $\{L(\theta, x), \text{一切}\theta \in \Theta\}$ 中最大者, 即参数 $\theta$ 的真值为 $\hat{\theta}^*$ 的"似然性"比参数空间 $\theta$ 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 $\theta$ 的估计值, 这一方法得到的参数 $\theta$ 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.

**Definition 3.3.2.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从参数分布族 $\mathscr{F}=\{f(x,\theta),\theta\in\Theta\}$ 中抽取的简单随机样本,$L(\theta,\boldsymbol{x})$ 是**似然函数**, 若存在统计量 $\hat{\theta}^*=\hat{\theta}^*(\boldsymbol{X})$, 满足条件

$$L(\hat{\theta}^*,\boldsymbol{x})=\sup_{\theta\in\Theta}L(\theta,\boldsymbol{x}),\boldsymbol{x}\in\mathscr{X}, \tag{3.3.2}$$

或等价的使得

$$L(\hat{\theta}^*,\boldsymbol{x})=\sup_{\theta\in\Theta}l(\theta,\boldsymbol{x}),\boldsymbol{x}\in\mathscr{X}, \tag{3.3.3}$$

则称 $\hat{\theta}^*$ 为 $\theta$ 的**极大似然估计** (maximum likelihood estimation, MLE). 若待估函数是 $g(\theta)$, 则定义 $g\left(\hat{\theta}^*\right)$ 为 $g(\theta)$ 的 MLE.

### 3.3.2 极大似然估计的求法及例

获得参数 $\theta$ 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**

设 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)$ 为参数向量 (特别当 $k=1$,$\boldsymbol{\theta}$ 为参数). 若 $l(\boldsymbol{\theta},\boldsymbol{x})$ 的极大值在参数空间 $\Theta$ 的内点 (而非边界点) 达到, 则此点必为似然方程组

$$\frac{\partial l(\boldsymbol{\theta},\boldsymbol{x})}{\partial\theta_i}=0,i=1,2,\cdots,k$$

的解.

因此求极大似然估计首先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 $\Theta$ 内部达到;②似然方程组只有唯一解, 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 $\boldsymbol{X}=(X_1,...,X_n)$ 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

$$f(x,\boldsymbol{\theta})=C(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^{k}\theta_iT_i(x)\right\}h(x),\quad \boldsymbol{\theta}\in\Theta,$$

其中 $\Theta$ 为自然参数空间,$\Theta_0$ 为 $\Theta$ 之内点, 这时 $\boldsymbol{X}$ 的联合密度为

$$L(\boldsymbol{\theta},\boldsymbol{x})=C^n(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}T_i(x_j)\right\}h(\boldsymbol{x}),$$

此处 $h(\boldsymbol{x})=\prod_{i=1}^{n}h(x_i)$. 对上式取对数得

$$l(\boldsymbol{\theta},\boldsymbol{x})=\log L(\boldsymbol{\theta},\boldsymbol{x})=n\log C(\boldsymbol{\theta})+\sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}T_i(x_j)+\log h(\boldsymbol{x}).$$

**Theorem 3.3.1.** 若对任何样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$, 方程组

$$\frac{n}{C(\boldsymbol{\theta})}\frac{\partial C(\boldsymbol{\theta})}{\partial\theta_i}=-\sum_{j=1}^{n}T_i(X_j), i=1,2,\cdots,k$$

在 $\Theta_0$ 内有解, 则解必唯一且为 $\boldsymbol{\theta}$ 的 MLE.

**2. 从定义出发**

当似然函数 $L(\theta,\boldsymbol{x})$ 对参数 $\theta$ 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 $\theta$ 的极大似然估计.

**Example 3.3.2.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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

$$L(p,\boldsymbol{x})=p^{\sum\limits_{i=1}^{n}x_i}(1-p)^{n-\sum\limits_{i=1}^{n}x_i},$$

故有

$$l(p,\boldsymbol{x})=\log L(p,\boldsymbol{x})=\left(\sum_{i=1}^{n}x_i\right)\log p+\left(n-\sum_{i=1}^{n}x_i\right)\log(1-p).$$

对数似然方程为

$$\frac{\partial l(p,\boldsymbol{x})}{\partial p}=\frac{1}{p}\sum_{i=1}^{n}x_i-\frac{1}{1-p}\left(n-\sum_{i=1}^{n}x_i\right)=0,$$

解得

$$\hat{p}^*=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}.$$

由于两点分布族为指数族, 当 $\overline{X}\neq 0,1$ 时, $\hat{p}^*=\overline{X}$ 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

$$\hat{g}^*(\boldsymbol{X})=\hat{p}^*(1-\hat{p}^*)=\overline{X}(1-\overline{X}).$$

**Note 3.3.2.** 当 $0<\overline{X}<1$ 时, 易知 $\hat{p}^*=\overline{X}$ 为 $p$ 唯一的 MLE. 而当 $\sum\limits_{i=1}^{n}X_i=0$, 或 $n$ 时 $\overline{X}=0,1$ 不在 $\Theta=(0,1)$ 内, 严格意义上来讲, 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 $\lim\limits_{n\to\infty}\hat{p}_n^*=\hat{p}^*$, 且 $\hat{p}_n^*\in\Theta$,$n=1,2,\cdots$, 则 $\hat{p}^*$ 也称为 $p$ 的 MLE. 在给出上述补充定义后, 本题中当 $\overline{X}=0,1$ 时, 可认为 $\hat{p}^*=\overline{X}$ 是 $p$ 的 MLE.

**Example 3.3.3.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从 Poisson 分布族 $\{P(\lambda):\lambda>0\}$ 中抽取的简单样本, 求 $\lambda$ 和 $g(\lambda)=e^{-\lambda}$ 的 MLE.

**解:** 似然函数, 即样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 的分布为

$$L(\lambda,\boldsymbol{x})=P(X_1=x_1,\cdots,X_n=x_n)=\frac{\lambda^{\sum\limits_{i=1}^{n}x_i}e^{-n\lambda}}{x_1!\cdots x_n!},\quad \lambda>0.$$

故对数似然函数为

$$l(\lambda,\boldsymbol{x})=\left(\sum_{i=1}^{n}x_i\right)\log\lambda-n\lambda-\sum_{i=1}^{n}\log x_i!.$$

由对数似然方程

$$\frac{\partial l(\lambda,\boldsymbol{x})}{\partial\lambda}=\frac{1}{\lambda}\sum_{i=1}^{n}x_i-n=0,$$

解得

$$\hat{\lambda}^*=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

由于 Poisson 分布族是指数族, 当 $\overline{X}\neq0$ 时,$\hat{\lambda}^*=\overline{X}$ 属于自然参数空间 $\Theta^*=(0,\infty)$ 的内点集, 故 $\hat{\lambda}^*$ 为 $\lambda$ 的 MLE, 它与 $\lambda$ 的矩估计量相同. 当 $\overline{X}=0$ 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

$$\hat{g}^*(X)=e^{-\overline{X}}.$$

**Example 3.3.4.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\mu$,$\sigma^2$ 和 $g(\theta)-\mu/\sigma^2$ 的 MLE.

**解:** 样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 的分布为

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

对数似然函数为

$$l(\theta,\boldsymbol{x})=\log f(\boldsymbol{x},\theta)=-\frac{n}{2}\log2\pi-\frac{n}{2}\log\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

由对数似然方程组

$$\frac{\partial l(\theta,\boldsymbol{x})}{\partial\mu}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0,$$
$$\frac{\partial l(\theta,\boldsymbol{x})}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0$$

解得
$$\hat{\mu}^* = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}, \quad \hat{\sigma}_*^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S_n^2.$$
由于正态分布族为指数族, 且 $\hat{\mu}^* = \overline{X}, \hat{\sigma}_*^2 = S_n^2$ 属于自然参数空间 $\Theta^* = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ 的内点集, 因此 $\hat{\mu}^*$ 和 $\hat{\sigma}_*^2$ 分别是 $\mu$ 和 $\sigma^2$ 的 MLE, 它们也分别是 $\mu$ 和 $\sigma^2$ 的矩估计量. 前者是 $\mu$ 的无偏估计, 后者不是 $\sigma^2$ 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
$$\hat{g}^*(\boldsymbol{X}) = \frac{\overline{X}}{S_n^2}.$$

**Example 3.3.5.** 设元件的寿命 $X$ 服从下列指数分布 $\mathrm{Exp}(\lambda)$, 其密度函数为
$$f(x,\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leqslant 0. \end{cases}$$
设 $X_1, \cdots, X_n$ 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 $r (r \leqslant n)$ 个元件失效时就停止了, 以 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(r)}$ 记这 $r$ 个元件的寿命, 即只观察到了样本 $X_1, \cdots, X_n$ 的前 $r$ 个次序统计量 $X_{(1)}, \cdots, X_{(r)}$. 基于这前 $r$ 个次序统计量, 求 $\lambda$ 和 $g(\lambda) = 1/\lambda$ 的 MLE.

**解:** 为叙述方便, 记 $t_i = x_{(i)}, i = 1, 2, \cdots, n$, 则有 $t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n$. 设 $F(x)$ 为 $X$ 的分布函数, 易知 $F(x) = 1 - e^{-\lambda x}$. 下面利用前面的结果和 2.3.2 节介绍的方法求 $X_{(1)}, X_{(2)}, \cdots, X_{(r)}$ 的密度如下:
$$\begin{aligned} p(t_1, \cdots, t_r) &= \int\cdots\int_{t_r < t_{r+1} < \cdots < t_n < \infty} n! f(t_1, \lambda) \cdots f(t_n, \lambda) \mathrm{d}t_{r+1} \cdots \mathrm{d}t_n \\ &= \frac{n!}{(n-r)!} f(t_1, \lambda) \cdots f(t_r, \lambda) \left[1 - F(t_r)\right]^{n-r} \\ &= \frac{n!}{(n-r)!} \lambda^r \exp\left\{-\lambda\left(\sum_{i=1}^{r} t_i + (n-r)t_r\right)\right\}. \end{aligned}$$

记 $T = \sum\limits_{i=1}^{r} t_i + (n-r)t_r$, 故 $\lambda$ 的似然函数为
$$L(\lambda, t_1, \cdots, t_r) = c \cdot \lambda^r \exp\{-\lambda T\},$$
其中 $c = n!/(n-r)!$. 对数似然函数为
$$l(\lambda, t_1, \cdots, t_r) = \log c + r \log \lambda - \lambda T.$$
对 $\lambda$ 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 $L\left(\lambda, x_{(1)}, \cdots, x_{(r)}\right)$ 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 $\lambda$ 的 MLE. 由定义可知 $g(\lambda) = 1/\lambda$ 的 MLE 为

$$\hat{g}^*(X_{(1)}, \cdots, X_{(r)}) = \frac{T}{r} = \frac{1}{r}\left[\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}\right].$$

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**.

**Example 3.3.6.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从均匀分布族 $\{U(0,\theta) : \theta > 0\}$ 中抽取的简单样本.

(1) 求 $\theta$ 的 MLE $\hat{\theta}^*$;

(2) 说明 $\hat{\theta}^*$ 是否为 $\theta$ 的无偏估计. 若不然, 作适当修正获得 $\theta$ 的无偏估计 $\hat{\theta}_1^*$;

(3) 试将 $\hat{\theta}_1^*$ 与 $\theta$ 的矩估计 $\hat{\theta}_1$ 比较, 看哪一个有效?

(4) 证明 $\theta$ 的极大似然估计 $\hat{\theta}^*$ 是 $\theta$ 的弱相合估计.

**解:** (1) 样本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 的联合密度为

$$f(\boldsymbol{x}, \theta) = \begin{cases} \frac{1}{\theta^n}, & 0 < x_1, \cdots, x_n < \theta, \\ 0, & \text{其他}. \end{cases} \tag{3.3.4}$$

因为均匀分布 $U(0,\theta)$ 的支撑集依赖于 $\theta$, 似然函数 $L(\theta, \boldsymbol{x}) = f(\boldsymbol{x}, \theta)$ 作为 $\theta$ 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 $\theta$ 的 MLE. 只能从 MLE 的定义出发来讨论.

为使 $L(\theta, \boldsymbol{x})$ 达到极大, 由式(3.3.4)可见, 应使分母上的 $\theta$ 尽可能地小, 但 $\theta$ 又不能太小以致 $L$ 为 0. 这个界限就在

$$\hat{\theta}^* = \max(X_1, \cdots, X_n) = X_{(n)}.$$

故当 $\theta > \hat{\theta}^*$ 时 $L > 0$ 且为 $\theta^{-n}$, 当 $\theta < \hat{\theta}^*$ 时 $L = 0$. 因此 $\hat{\theta}^*$ 是唯一使 $L$ 达到最大的 $\theta$ 值, 即 $\theta$ 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

$$g(t, \theta) = \begin{cases} \frac{nt^{n-1}}{\theta^n}, & 0 \leqslant t \leqslant \theta, \\ 0, & \text{其他}, \end{cases} \tag{3.3.5}$$

故有

$$E[\hat{\theta}^*] = E[T] = \int_0^\theta \frac{nt^n}{\theta^n} \mathrm{d}t = \frac{n}{n+1}\theta,$$

故 $\hat{\theta}^* = X_{(n)}$ 不是 $\theta$ 的无偏估计. 显见

$$\hat{\theta}^* = \frac{n+1}{n} X_{(n)}$$

为 $\theta$ 的无偏估计.

(3) $\theta$ 的矩估计 $\hat{\theta}_1 = 2\overline{X}$ 是 $\theta$ 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 $n \geqslant 2$ 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

$$\begin{aligned}
& P(|\hat{\theta}^* - \theta| \geqslant \varepsilon) \\
=& 1 - P(|\hat{\theta}^* - \theta| < \varepsilon) = 1 - P(\theta - \varepsilon < T < \theta + \varepsilon) \\
=& 1 - \int_{\theta-\varepsilon}^{\theta} \frac{n t^{n-1}}{\theta^n} \mathrm{d}t = 1 - \frac{1}{\theta^n}\left[\theta^n - (\theta-\varepsilon)^n\right] = \left(1 - \frac{\varepsilon}{\theta}\right)^n.
\end{aligned}$$

因此有

$$\lim_{n\to\infty} P(|\hat{\theta}^* - \theta| \geqslant \varepsilon) = \lim_{n\to\infty}\left(1 - \frac{\varepsilon}{\theta}\right)^n = 0,$$

故知 $\hat{\theta}^* = X_{(n)}$ 为 $\theta$ 的弱相合估计.

**Example 3.3.7.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从均匀分布族 $\{U(\theta, \theta+1) : -\infty < \theta < +\infty\}$ 中抽取的简单样本, 求 $\theta$ 的 MLE.

**解:** 给定样本 $\boldsymbol{x}$ 时,$\theta$ 的似然函数为

$$L(\theta, \boldsymbol{x}) = \begin{cases} 1, & \theta < x_{(1)} \leqslant x_{(n)} < \theta + 1 \\ 0, & \text{其他} \end{cases} = \begin{cases} 1, & x_{(n)} - 1 < \theta < x_{(1)} \\ 0, & \text{其他}. \end{cases}$$

这是, 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 $\theta$ 的 MLE 不止一个, 如

$$\hat{\theta}_1^*(\boldsymbol{X}) = X_{(1)}, \quad \hat{\theta}_2^*(\boldsymbol{X}) = X_{(n)} - 1$$

都是 $\theta$ 的 MLE. 事实上对任给的 $0 \leqslant \lambda \leqslant 1$,

$$\hat{\theta}^* = \lambda \hat{\theta}_1^*(\boldsymbol{X}) + (1-\lambda)\hat{\theta}_2^*(\boldsymbol{X}) = \lambda \boldsymbol{X}_{(1)} + (1-\lambda)(X_{(n)} - 1)$$

都是 $\theta$ 的 MLE, 故知 $\theta$ 的 MLE 有无穷多个.

**Example 3.3.8.** 设 $k$ 个事件 $A_1, A_2, \cdots, A_k$ 构成完备事件群, 事件 $A_i$ 发生的概率为 $0 < p_i < 1, i = 1, 2, \cdots, k$ 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, $i = 1, 2, \cdots, k$, 则 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 服从多项分布 $M(n, p_1, \cdots, p_k)$. 求 $p_1, \cdots, p_k$ 的 MLE.

**解:** 记 $\boldsymbol{p} = (p_1, \cdots, p_k)$. 给定样本 $\boldsymbol{x}$ 时, $\boldsymbol{p}$ 的似然函数为

$$L(\boldsymbol{p}, \boldsymbol{x}) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{x_k}.$$

对数似然函数为

$$l(\boldsymbol{p}, \boldsymbol{x}) = \log n! - \sum_{i=1}^{k} \log x_i! + \sum_{i=1}^{k-1} x_i \log p_i + x_k \log \left(1 - \sum_{i=1}^{k-1} p_i\right). \tag{3.3.6}$$

对 $p_i$ 求偏导数, 得似然方程组

$$\frac{\partial l(\boldsymbol{p}, \boldsymbol{x})}{\partial p_i} = \frac{x_i}{p_i} - \frac{x_k}{p_k} = 0, \ i = 1, 2, \cdots, k-1.$$

若令

$$\frac{x_1}{p_1} = \frac{x_2}{p_2} = \cdots = \frac{x_k}{p_k} = \lambda,$$

则有

$$x_i = \lambda p_i, \ i = 1, 2, \cdots, k. \tag{3.3.7}$$

将这 $k$ 个等式分别相加得到

$$n = \sum_{i=1}^{k} x_i = \lambda \sum_{i=1}^{k} p_i = \lambda.$$

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:

$$\hat{p}_i^* = \frac{X_i}{n}, \quad i = 1, 2, \cdots, k.$$

在有些问题中, $p_1, \cdots, p_k$ 都是另一些参数 $\theta_1, \cdots, \theta_r (r \leqslant k)$ 的函数, 这时去掉式(3.3.6)中与 $\theta_1, \cdots, \theta_r$ 无关的部分后, 得对数似然函数为

$$l(\theta_1, \cdots, \theta_r; \boldsymbol{x}) = \sum_{i=1}^{k} x_i \log p_i(\theta_1, \cdots, \theta_r),$$

对 $\theta_i$ 求偏导数得似然方程组

$$\frac{\partial l}{\partial \theta_i} = 0, i = 1, 2, \cdots, r. \tag{3.3.8}$$

若似然方程之解 $\hat{\theta}_1^*, \cdots, \hat{\theta}_r^*$ 为 $\theta_1, \cdots, \theta_r$ 的 MLE, 则 $p_1, \cdots, p_k$ 的 MLE 分别是 $\hat{p}_1(\hat{\theta}_1^*, \cdots, \hat{\theta}_r^*)$, $\cdots$, $\hat{p}_k(\hat{\theta}_1^*, \cdots, \hat{\theta}_r^*)$. 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**

例3.3.4已经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, 也可以是有偏的, 取决于具体情况.

**2. 极大似然估计与充分统计量**

设 $\boldsymbol{X}=(X_1,\ldots,X_n)$ 为自总体 $\{f(x,\theta),\theta\in\Theta\}$ 中抽出的简单随机样本, $T=T(X_1,\ldots,X_n)$ 是参数 $\theta$ 的充分统计量, 如果 $\theta$ 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 $\boldsymbol{X}$ 的概率函数, 即似然函数可表示为:

$$L(\theta,\boldsymbol{x})=\prod_{i=1}^{n}f(x_i,\theta)=g(T(\boldsymbol{x}),\theta)h(\boldsymbol{x}).$$

若 $\theta$ 的 MLE 存在, 记为 $\hat{\theta}^*$, 则

$$L(\hat{\theta}^*,\boldsymbol{x})=\sup_{\theta\in\Theta}L(\theta,\boldsymbol{x})\Longleftrightarrow g(T(\boldsymbol{x}),\hat{\theta}^*)=\sup_{\theta\in\Theta}g(T(\boldsymbol{x}),\theta),$$

因此使 $\sup\limits_{\theta}L(\theta,\boldsymbol{x})$ 达到上确界之点 $\hat{\theta}^*$, 即为使 $\sup\limits_{\theta}g(T(\boldsymbol{x}),\theta)$ 达到上确界之点, 它必为 $T(\boldsymbol{x})$ 的函数.

此性质说明 $\theta$ 的极大似然估计 $\hat{\theta}^*=\hat{\theta}^*(X_1,\ldots,X_n)$ 可表示为充分统计量 $T(\boldsymbol{X})$ 的函数, 即 $\hat{\theta}^*=\hat{\theta}^*(T(X_1,\ldots,X_n))$, 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**

1949 年 Wald 首次证明极大似然估计的强相合性, 但所需要的条件很复杂.

***4. 极大似然估计的渐近相合正态性**

只考虑参数 $\theta$ 为一维的情形. 设 $\mathscr{F}=\{f(x,\theta),\theta\in\Theta\}$ 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 $\theta\in\Theta$, 由 $f(x,\theta)>0$ 当 $x\in\mathscr{X}$;$f(x,\theta)=0$, 当 $x\notin\mathscr{X}$; 且对一切 $\theta\in\Theta$,$x\in\mathscr{X}$ 时, 偏导数

$$\frac{\partial\log f(x,\theta)}{\partial\theta},\frac{\partial^2\log f(x,\theta)}{\partial\theta^2},\frac{\partial^3\log f(x,\theta)}{\partial\theta^3}$$

存在.

(2) 存在定义于样本空间 $\mathscr{X}$ 上的函数 $F_1(x)$,$F_2(x)$ 和 $H(x)$, 使对一切 $\theta\in\Theta$ 和 $x\in\mathscr{X}$ 有

$$\left|\frac{\partial f(x,\theta)}{\partial\theta}\right|<F_1(x),\quad\left|\frac{\partial^2 f(x,\theta)}{\partial\theta^2}\right|<F_2(x),\quad\left|\frac{\partial^3 f(x,\theta)}{\partial\theta^3}\right|\leqslant H(x),$$

其中

$$\int_{-\infty}^{\infty}F_i(x)\mathrm{d}x<\infty,i=1,2;\quad\int_{-\infty}^{\infty}H(x)f(x,\theta)\mathrm{d}x<M,\theta\in\Theta,$$

此处 $M$ 与 $\theta$ 无关.

(3) 对一切 $\theta\in\Theta$, 有

$$0<I(\theta)=E\left[\left(\frac{\partial\log f(X,\theta)}{\partial\theta}\right)^2\right]=\int_{\infty}^{\infty}\left(\frac{\partial\log f(x,\theta)}{\partial\theta}\right)^2 f(x,\theta)\mathrm{d}x<\infty.$$

关于极大似然估计的相合渐进正态 (CAN) 性, 有下列结果.

**Theorem 3.3.2.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\hat{\theta}^*=\hat{\theta}^*(X_1,\cdots,X_n)$, 则 $\hat{\theta}^*$ 为 $\theta$ 的 CAN 估计, 即

$$\sqrt{n}(\hat{\theta}^*-\theta)\xrightarrow{\mathscr{L}}N\left(0,\frac{1}{I(\theta)}\right),\quad\theta\in\Theta,$$

且 $\hat{\theta}$ 为 $\theta$ 的弱相合估计.

**Example 3.3.9.** 设 $X_1,\cdots,X_n$ 是取自正态分布 $N(\mu,\sigma^2)$ 的样本. 证明 $\mu$ 和 $\sigma^2$ 的 MLE 分别具有渐近正态性.

## *3.4 一致最小方差无偏估计

### 3.4.1 引言及定义

**Definition 3.4.1.** 设 $\hat{g}(\boldsymbol{X})$ 为 $g(\theta)$ 的估计量, 则称 $E_\theta(\hat{g}(\boldsymbol{X})-g(\theta))^2$ 为 $\hat{g}(\boldsymbol{X})$ 的均方误差 (mean square error,MSE).

设 $\hat{g}_1(\boldsymbol{X})$ 和 $\hat{g}_2(\boldsymbol{X})$ 为 $g(\theta)$ 的两个不同的估计量, 若

$$E_\theta\left(\hat{g}_1(\boldsymbol{X})-g(\theta)\right)^2\leqslant E_\theta\left(\hat{g}_2(\boldsymbol{X})-g(\theta)\right)^2,\quad 一切\theta\in\Theta,$$

且不等号至少对某个 $\theta\in\Theta$ 成立, 则称在 MSE 准则下 $\hat{g}_1(\boldsymbol{X})$ 优于 $\hat{g}_2(\boldsymbol{X})$.

若存在 $\hat{g}^*(\boldsymbol{X})$, 使得对 $g(\theta)$ 的任一估计量 $\hat{g}(\boldsymbol{X})$, 都有

$$E_\theta\left(\hat{g}^*(\boldsymbol{X})-g(\theta)\right)^2\leqslant E_\theta\left(\hat{g}(\boldsymbol{X})-g(\theta)\right)^2,\quad 一切\theta\in\Theta,$$

则称 $\hat{g}^*(\boldsymbol{X})$ 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, 使最合适这种最优性准则的估计一般都能够存在.

在无偏估计类中, 估计量的均方误差就变为其方差, 即当 $\hat{g}(\boldsymbol{X})$ 为 $g(\theta)$ 的无偏估计时,MSE$(\hat{g}(\boldsymbol{X}))$ $= D_\theta(\hat{g}(\boldsymbol{X}))$, 此处 $D_\theta(\hat{g}(\boldsymbol{X}))$ 表示 $\hat{g}(\boldsymbol{X})$ 的方差.

**Example 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.

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.

**Definition 3.4.2.** 设 $\mathscr{F} = \{f(x,\theta), \theta \in \Theta\}$ 是一个参数分布族, 其中 $\Theta$ 为参数空间,$g(\theta)$ 为定义在 $\Theta$ 上的可估函数. 设 $\hat{g}^*(\boldsymbol{X}) = \hat{g}^*(X_1, \cdots, X_n)$ 为 $g(\theta)$ 的一个无偏估计, 若对 $g(\theta)$ 的任一无偏估计 $\hat{g}(\boldsymbol{X}) = \hat{g}(X_1, \cdots, X_n)$, 都有

$$D_\theta(\hat{g}^*(\boldsymbol{X})) \leqslant D_\theta(\hat{g}(\boldsymbol{X})), \text{一切}\theta \in \Theta,$$

则称 $\hat{g}^*(\boldsymbol{X})$ 是 $g(\theta)$ 的**一致最小方差无偏估计** (uniformly minimum variance unbiased estimation, UMVUE).

下面的引理可以简化一些问题.

**Lemma 3.4.1.** 设 $T = T(\boldsymbol{X})$ 是一个充分统计量, 而 $\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的一个无偏估计, 则

$$h(T) = E(\hat{g}(X)|T)$$

是 $g(\theta)$ 的无偏估计, 并且

$$D_\theta(h(T)) \leqslant D_\theta(\hat{g}(\boldsymbol{X})), \quad \text{一切}\theta \in \Theta, \tag{3.4.1}$$

其中等号当且仅当 $P_\theta(\hat{g}(\boldsymbol{X}) = h(T)) = 1$, 即 $\hat{g}(\boldsymbol{X}) = h(T)$, a.s.$P_\theta$ 成立.

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 $\hat{g}(\boldsymbol{X})$ 对充分估计量 $T(\boldsymbol{X})$ 的条件期望 $E\{\hat{g}(\boldsymbol{X})|T\}$ 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 $\hat{g}(\boldsymbol{X})$ 的方差. 若原估计 $\hat{g}(\boldsymbol{X})$ 不是 $T(\boldsymbol{X})$ 的函数, 则新的无偏估计 $E(\hat{g}(\boldsymbol{X})|T)$ 一定比原估计 $\hat{g}(\boldsymbol{X})$ 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.

**Example 3.4.2.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 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,$T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ 是 $p$ 的充分统计量, 试利用 $T = T(\boldsymbol{X})$ 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:

$$\begin{aligned} h(t) =& E(X_1|T=t) = 1\cdot P(X_1=1|T=t) + 0\cdot P(X_1=0|T=t) \\ =& \frac{P(X_1=1,T=t)}{P(T=t)} = \frac{P(X_1=1,X_2+\cdots+X_n=t-1)}{P(T=t)} \\ =& \frac{p\cdot\binom{n-1}{t-1}p^{t-1}(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}} = \frac{t}{n} = \overline{x}. \end{aligned}$$

显然, 样本均值 $h(T)=\overline{X}$ 的方差为 $p(1-p)/n$, 而 $X_1$ 的方差为 $p(1-p)$, 当 $n\geqslant 2$ 时 $\overline{X}$ 的方差更小.

## *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.

**Theorem 3.4.1.** 设 $\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的一个无偏估计,$D_\theta(\hat{g}(\boldsymbol{X}))<\infty$, 对任何 $\theta\in\Theta$ 若对任何满足条件 "$E_\theta l(\boldsymbol{X})=0$, 对一切 $\theta\in\Theta$" 的统计量 $l(\boldsymbol{X})$, 必有

$$\operatorname{Cov}_\theta(\hat{g}(\boldsymbol{X}),l(\boldsymbol{X})) = E_\theta\left[\hat{g}(\boldsymbol{X})\cdot l(\boldsymbol{X})\right] = 0, \text{一切}\theta\in\Theta, \tag{3.4.2}$$

则 $\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的 UMVUE.

**Note 3.4.1.** 从形式上看, 条件 "$E_\theta l(\boldsymbol{X})=0,\theta\in\Theta$" 可解释为 "$l(\boldsymbol{X})$ 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 $\hat{g}(\boldsymbol{X})$ 为 $g(\theta)$ 的一个无偏估计,$D_\theta(\hat{g}(\boldsymbol{X}))<\infty,\theta\in\Theta$, 则 $\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的 UMVUE 的充分必要条件是: 对任何满足条件 "$E_\theta l(\boldsymbol{X})=0$, 一切 $\theta\in\Theta$" 的统计量 $l(\boldsymbol{X})$, 必有式(3.4.2)成立.

**Corollary 3.4.1.** 设 $T=T(\boldsymbol{X})$ 为 $\theta$ 的充分统计量,$h(T(\boldsymbol{X}))$ 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 "对一切 $\theta\in\Theta$ 都有 $E_\theta\delta(T)=0$" 的统计量 $\delta(T)$, 必有

$$\operatorname{Cov}\left(h(T),\delta(T)\right) = E_\theta\left[h(T)\cdot\delta(T)\right] = 0, \text{一切}\theta\in\Theta, \tag{3.4.3}$$

则 $h(T)$ 是 $g(\theta)$ 的 UMVUE.

**Example 3.4.3.** 求例3.4.2中 $g(p)=p$ 的 UMVUE.

**解:** 过程略, 结果为 $h(T)=\overline{X}=T/n$ 为 $p$ 的 UMVUE.

**Example 3.4.4.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从指数分布 $\mathrm{Exp}(\lambda)$(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:** 过程略. 结果为 $h(T)=T/n$ 为 $g(\lambda)=1/\lambda$ 的 UMVUE.

**Example 3.4.5.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从均匀分布 $U(0,\theta)$ 抽取的简单样本, 求 $\theta$ 的 UMVUE.

**解:** 过程略, 结果为 $h(T)=(n+1)X_{(n)}/n$ 为 $g(\theta)=\theta$ 的 UMVUE.

***Example 3.4.6.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从正态分布 $N(a,\sigma^2)$ 中抽取的简单随机样本, 求 $a$ 和 $\sigma^2$ 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 $T_1=\overline{X}$,$T_2=\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$, 又 $T_1$ 和 $T_2$ 独立且 $T_1\sim N(a,\sigma^2/n)$,$T_2/\sigma^2\sim\chi^2_{n-1}$. 因此 $(T_1,T_2)$ 的联合密度为

$$f_\theta(t_1,t_2)=\begin{cases}\frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-\frac{n(t_1-a)^2}{2\sigma^2}}\cdot\frac{1}{2^{\frac{n-1}{2}}\Gamma((n-1)/2)\sigma^{n-1}}t_2^{\frac{n-1}{2}-1}e^{-\frac{t_2}{2\sigma^2}}, & -\infty<t_1<+\infty,t_2>0\\ 0 & \text{其他.}\end{cases} \tag{3.4.4}$$

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

$$E_\theta(\delta(T))=\int_0^\infty\int_{-\infty}^\infty\delta(t_1,t_2)f_\theta(t_1,t_2)\mathrm{d}t_1\mathrm{d}t_2=0,$$

此处 $-\infty<a<\infty$,$\sigma>0$. 将上式两边对 $a$ 求导数, 得

$$\int_0^\infty\int_{-\infty}^\infty\delta(t_1,t_2)(t_1-a)\cdot t_2^{\frac{n-3}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left[n(t_1-a)^2+t_2\right]\right\}\mathrm{d}t_1\mathrm{d}t_2=0,$$

按式(3.4.4), 上式等价于

$$\mathrm{Cov}_\theta(h_1(T),\delta(T))=E_\theta(h_1(T)\cdot\delta(T_1,T_2))=0,-\infty<a<+\infty,\sigma>0,$$

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.

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 $\overline{X}=\sum\limits_{i=1}^{n}X_i/n$ 皆为 $g(\theta)=\theta/2$ 的无偏估计, 但 $(n+1)X_{(n)}/(2n)$ 为 $g(\theta)$ 的 UMVUE, 它比 $\overline{X}$ 更好, 而在例3.4.6中正好相反. 此例中因 $\overline{X}$ 为 $g_1(\theta)=a$ 的 UMVUE, 故 $\overline{X}$ 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

**Theorem 3.4.2** (Lehmann-Scheffè 定理). 设 $X \sim \{f(x,\theta), \theta \in \Theta\}$,$\Theta$ 为参数空间. 令 $X_1, \cdots, X_n$ 为从总体 $X$ 中抽取的简单样本,$g(\theta)$ 为定义于参数空间 $\Theta$ 上的可估函数,$T(\boldsymbol{X})$ 为一个充分完全统计量. 若 $\hat{g}(T(\boldsymbol{X}))$ 为 $g(\theta)$ 的一个无偏估计, 则 $\hat{g}(T(\boldsymbol{X}))$ 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).

**Corollary 3.4.2.** 设样本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 的指数分布族函数

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp\left\{\sum_{j=1}^{k} \theta_j T_j(\boldsymbol{x})\right\} h(\boldsymbol{x}), \boldsymbol{\theta} = (\theta_1, \cdots, \theta_k \in \Theta^*).$$

令 $T(\boldsymbol{X}) = (T_1(\boldsymbol{X}), \cdots, T_k(\boldsymbol{X}))$, 若自然参数空间 $\Theta^*$ 作为 $R_k$ 的子集有内点, 且 $h(T(\boldsymbol{X}))$ 为 $g(\boldsymbol{\theta})$ 的无偏估计, 则 $h(T(\boldsymbol{X}))$ 为 $g(\boldsymbol{\theta})$ 的唯一的 UMVUE.

**Example 3.4.7.** 证明例3.4.3中获得的 $p$ 的无偏估计 $h(T) = T/n = \overline{X}$ 为 $p$ 的 UMVUE.

**Example 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 $T(\boldsymbol{X})$ 为充分完全统计量, 求 $g(p) - p(1-p)$ 的 UMVUE.

**解:** 过程略, 解答为 $\delta(T) = \frac{T(n-T)}{n(n-1)}$.

**Example 3.4.9.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 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:** 过程略,$T = T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ 为充分完全统计量. (1) $h_1(T) = T/n$, (2) $h_2(T) = \dfrac{T(T-1)\cdots(T-r+1)}{n^r}$, 当 $T = 0, 1, 2, \cdots$ (3) $h_3(T) = \dbinom{T}{x}\left(\dfrac{1}{n}\right)^x\left(1 - \dfrac{1}{n}\right)^{T-r}$.

**Example 3.4.10.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 为从指数分布 $\mathrm{Exp}(\lambda)$ 中抽取的简单样本, 求 (1) $g(\lambda) = \lambda$; (2) $g_2(\lambda) = 1 - e^{-\lambda x_0}$ 的 UMVUE, 其中 $x_0$ 已知.

**解:** 过程略,$T = T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ 为充分完全统计量, 且 $T(\boldsymbol{X}) \sim \Gamma(n, \lambda)$. (1)$h(T) = \frac{n-1}{T}$ 为 $\lambda$ 的 UMVUE. (2) $h(T) = 1 - \frac{(T-x_0)^{n-1}}{T^{n-1}} I_{(x_0,\infty)}(T)$ 为 $g_2(\lambda) = 1 - e^{-\lambda x_0}$ 的 UMVUE.

**Example 3.4.11.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\sigma^2$; (2) $g_1(\theta)=\sigma^r, r>0$; (3) $g_2(\theta)=a/\sigma^2$ 的 UMVUE.

**解:** (1) 由于 $h_1(T)=\overline{X}=T_1$ 和 $h_2(T)=T_2/(n-1)$ 分别为 $a$ 和 $\sigma^2$ 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 (定理3.4.2) 可知它们分别是 $a$ 和 $\sigma^2$ 的 UMVUE.

(2) 由于 $Y=T_2/\sigma^r\sim\chi^2_{n-1}$, 故 $\sigma^r$ 的无偏估计与 $T_2$ 的幂函数有关. 先计算下式:

$$\begin{aligned}\frac{1}{\sigma^r}E\left(T_2^{\frac{r}{2}}\right)=&E\left(\frac{T_2}{\sigma^2}\right)^{r/2}\\=&\int_0^\infty y^{\frac{r}{2}}\cdot\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}y^{\frac{n-1}{2}-1}e^{-\frac{y}{2}}\mathrm{d}y\\=&\frac{2^{r/2}\Gamma\left(\frac{n+r-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\triangleq\frac{1}{K_{n-1,r}}.\end{aligned}$$

由上式可知

$$E(K_{n-1,r}\cdot T_2^{\frac{n}{2}})=\sigma^r.$$

因此估计量

$$h_3(T)=K_{n-1,r}T_2^{r/2}=\frac{\Gamma\left(\frac{n+1}{2}\right)}{2^{(r/2)}\Gamma\left(\frac{n+1}{2}\right)}T_2^{r/2}$$

是 $\sigma^r$ 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 $q(\theta)=\sigma^r$ 的 UMVUE.

(3) 由下列事实: 若 $X\sim\chi^2_n$, 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 $q_2(\theta)$ 的 UMVUE.

**Example 3.4.12.** 用 Lehmann-Scheffè 定理求例3.4.5中均匀分布 $U(0,\theta)$ 中的参数 $\theta$ 的 UMVUE.

**解:** 由 2.7 节和 2.8 节可知 $T=T(\mathbf{X})=\max(X_1,\cdots,X_n)=X_{(n)}$ 为充分完全统计量，由例3.3.6已知 $h(T)=(n+1)T/n$ 为 $\theta$ 的无偏估计，故由 L-S 定理立得 $h(T)$ 为 $\theta$ 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

**Definition 3.5.1.** 若单参数概率函数族 $\mathscr{F}=\{f(x,\theta),\theta\in\Theta\}$ 满足下列条件:

(1) 参数空间 $\Theta$ 是直线上的某个开区间;

(2) 对任何 $x\in\mathscr{X}$ 及 $\theta\in\Theta,f(x,\theta)>0$, 即分布族具有共同支撑;

(3) 对任何 $x\in\mathscr{X}$ 及 $\theta\in\Theta,\dfrac{\partial f(x,\theta)}{\partial\theta}$ 存在;

(4) 概率函数 $f(x,\theta)$ 的积分与微分运算可交换, 即

$$\frac{\partial}{\partial\theta}\int f(x,\theta)\mathrm{d}x=\int\frac{\partial}{\partial\theta}f(x,\theta)\mathrm{d}x,$$

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).(“C-R” 为 “Cramer-Rao” 的简写)

### 3.5.2 单参数 C-R 不等式

#### 1.C-R 不等式及例

**Theorem 3.5.1.** 设 $\mathscr{F}=\{f(x,\theta),\theta\in\Theta\}$ 是 C-R 正则分布族,$g(\theta)$ 是定义于参数空间 $\Theta$ 上的可微函数. 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是由总体 $f(x,\theta)\in\mathscr{F}$ 中抽取的简单随机样本,$\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的任一无偏估计, 且满足下列条件:

(6) 积分

$$\int\cdots\int\hat{g}(\boldsymbol{x})f(\boldsymbol{x},\theta)\mathrm{d}\boldsymbol{x}$$

可在积分号下对 $\theta$ 求导数, 此处 $\mathrm{d}\boldsymbol{x}=\mathrm{d}x_1\cdots\mathrm{d}x_n$, 则有

$$D_\theta[\hat{g}(\boldsymbol{X})]\geqslant\frac{(g'(\theta))^2}{nI(\theta)},\text{一切}\theta\in\Theta. \tag{3.5.1}$$

特别当 $g(\theta)=\theta$ 时, 式(3.5.1)变为

$$D_\theta[\hat{g}(\boldsymbol{X})]\geqslant\frac{1}{nI(\theta)},\text{一切}\theta\in\Theta. \tag{3.5.2}$$

当 $f(x,\theta)$ 为离散 r.v. $X$ 的概率分布时, 式(3.5.1)变为

$$D_\theta[\hat{g}(\boldsymbol{X})] \geqslant \frac{[g'(\theta)]^2}{n\sum\limits_i \left\{\left[\frac{\partial \log f(x_i,\theta)}{\partial\theta}\right]^2 f(x_i,\theta)\right\}}, \quad 一切\theta\in\Theta. \tag{3.5.3}$$

**证:** 由于 $X_1,\dots,X_n$ 为 i.i.d. 样本, 故有 $f(\boldsymbol{x},\boldsymbol{\theta})=\prod\limits_{i=1}^n f(x_i,\boldsymbol{\theta})$. 记

$$S(\boldsymbol{x},\theta)=\frac{\partial \log f(\boldsymbol{x},\theta)}{\partial\theta}=\sum_{i=1}^n\frac{\partial \log f(x_i,\theta)}{\partial\theta},$$

因此由正则条件 (3) 和 (4) 可知

$$\begin{aligned}E_\theta\{S(\boldsymbol{X},\theta)\}=&\sum_{i=1}^n E_\theta\left\{\frac{\partial \log f(X_i,\theta)}{\partial\theta}\right\}=\sum_{i=1}^n\int\frac{1}{f(x_i,\theta)}\frac{\partial f(x_i,\theta)}{\partial\theta}\cdot f(x_i,\theta)\mathrm{d}x_i\\=&\sum_{i=1}^n\int\frac{\partial f(x_i,\theta)}{\partial\theta}\mathrm{d}x_i=\sum_{i=1}^n\frac{\partial}{\partial\theta}\int f(x_i,\theta)\mathrm{d}x=0.\end{aligned}$$

由 $\hat{g}(\boldsymbol{x})$ 为 $g(\theta)$ 的无偏估计和正则条件 (5) 和 (6) 可知

$$\begin{aligned}\mathrm{Cov}_\theta(\hat{g}(\boldsymbol{X}),S(\boldsymbol{X},\theta))=&E_\theta\{\hat{g}(\boldsymbol{X})\cdot S(\boldsymbol{X},\theta)\}\\=&\int\cdots\int\hat{g}(\boldsymbol{x})\left[\frac{1}{f(\boldsymbol{x},\theta)}\frac{\partial f(\boldsymbol{x},\theta)}{\partial\theta}\right]f(x,\theta)\mathrm{d}x\\=&\int\cdots\int\hat{g}(\boldsymbol{x})\frac{\partial f(\boldsymbol{x},\theta)}{\partial\theta}\mathrm{d}x\\=&\frac{\partial}{\partial\theta}\int\cdots\int\hat{g}(\boldsymbol{x})f(\boldsymbol{x},\theta)\mathrm{d}\boldsymbol{x}=\frac{\partial g(\theta)}{\partial\theta}=g'(\theta),\end{aligned}$$

$$\begin{aligned}D_\theta(S(\boldsymbol{X},\theta))=&\sum_{i=1}^n D_\theta\left\{\frac{\partial \log f(X_i,\theta)}{\partial\theta}\right\}\\=&\sum_{i=1}^n E_\theta\left\{\frac{\partial \log f(X_i,\theta)}{\partial\theta}\right\}^2=nI(\theta).\end{aligned}\tag{3.5.4}$$

由 Cauchy-Schwarz 不等式, 得

$$D_\theta\{\hat{g}(\boldsymbol{X})\}\cdot D_\theta\{S(\boldsymbol{X},\theta)\}\geqslant[\mathrm{Cov}_\theta(\hat{g}(\boldsymbol{X}),S(\boldsymbol{X},\theta))]^2=[g'(\theta)]^2,$$

将式(3.5.4)代入上式得

$$D_\theta\left[\hat{g}(\boldsymbol{X})\right]\geqslant\frac{[g'(\theta)]^2}{nI(\theta)}, 一切\theta\in\Theta.$$

定理得证.

不等式(3.5.1)称为 **Cramer-Rao 不等式**, 简称 C-R 不等式. 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 $g(\boldsymbol{X})$ 的方差 $D_\theta(\hat{g}(\boldsymbol{X}))$, 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 $D_\theta(\hat{g}(\boldsymbol{X}))$ 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,$g(\boldsymbol{X})$ 是 $g(\theta)$ 的 UMVUE, 但其方差大于 C-R 下界. 后面将给出这一例子.

**Example 3.5.1.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\overline{X}=\sum\limits_{i=1}^{n}X_i/n$ 为 $p$ 的 UMVUE.

**Example 3.5.2.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\overline{X}=\sum\limits_{i=1}^{n}X_i/n$ 为 $\lambda$ 的 UMVUE.

**Example 3.5.3.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 从指数分布 $\mathrm{Exp}(\lambda)$ 中抽取的简单样本, 用 C-R 不等式验证 $\hat{g}(\boldsymbol{X})=\overline{X}$ 为 $g(\lambda)=1/\lambda$ 的 UMVUE.

**Example 3.5.4.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从 $N(a,\sigma^2)$ 中抽取的简单随机样本, 其中 $\sigma^2$ 已知, 用 C-R 不等式验证 $\overline{X}$ 为 $a$ 的 UMVUE.

#### *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 $f(\boldsymbol{x},\theta)=C(\theta)\exp\{Q(\theta)T(\boldsymbol{x})\}h(\boldsymbol{x})$, 也不是对任何 $g(\theta)$ 都能够找到无偏估计 $\hat{g}(\boldsymbol{X})$, 使其方差处处达到 C-R 下界. 唯有在 $g(\theta)=E_\theta(aT(\boldsymbol{X})+b)$ 时才有, 即 $\hat{g}(\boldsymbol{X})=aT(\boldsymbol{X})+b$ 的情形才可, 此处 $a\neq 0$ 与 $b$ 与 $\boldsymbol{X}$ 无关, 但可以是 $\theta$ 的函数.

**Example 3.5.5.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为自 Poisson 分布 $P(\lambda)$ 中抽取的简单随机样本, 证明只有 $g(\lambda)$ 是 $\lambda$ 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.

#### 3.Fisher 信息函数

我们将 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**).

### 3.5.3 多参数 C-R 不等式简介

本分节内容仅了解即可.

以上讨论的都是参数 $\theta$ 为一维的情形, 对 $\theta$ 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 $A \geqslant B$. 这时必有 $a_{ii} \geqslant b_{ii}$, 对一切 $i$.

现设 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)$, 总体概率函数记为 $f(x,\boldsymbol{\theta})$,$\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从总体中抽取的简单随机样本. 设 $\hat{\boldsymbol{\theta}}=\hat{\theta}(\boldsymbol{X})=(\hat{\theta}_1,\cdots,\hat{\theta}_k)$ 是 $\boldsymbol{\theta}$ 的一个无偏估计. 以 $\mathrm{Cov}_{\boldsymbol{\theta}}(\hat{\theta})$ 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 $E_\theta\left[(\hat{\theta}_i-\theta_i)(\hat{\theta}_j-\theta_j)\right]$, 则在类似于 $\theta$ 为一维的正则条件下, 可以证明

$$\mathrm{Cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \geqslant (n\boldsymbol{I}(\boldsymbol{\theta}))^{-1}, \tag{3.5.5}$$

其中 $I(\theta)=(I_{ij}(\theta))$ 是一个 $k$ 阶正定方阵, 且

$$I_{ij}(\boldsymbol{\theta})=E_{\boldsymbol{\theta}}\left[\left(\frac{\partial \log f(X,\boldsymbol{\theta})}{\partial \theta_i}\right)\left(\frac{\partial \log f(X,\boldsymbol{\theta})}{\partial \theta_j}\right)\right], i,j=1,2,\cdots,k, \tag{3.5.6}$$

式(3.5.5)就是多维的 C-R 不等式. 若记 $(\boldsymbol{I}(\boldsymbol{\theta}))^{-1}=\boldsymbol{I}^*(\boldsymbol{\theta})=\left(\boldsymbol{I}_{ij}^*(\boldsymbol{\theta})\right)$, 故由式 (3.5.5)可得

$$D_{\boldsymbol{\theta}}\left(\hat{\theta}_i\right) \geqslant \frac{I_{ii}^*(\boldsymbol{\theta})}{n}, \quad i=1,2,\cdots,k. \tag{3.5.7}$$

这给出了 $\boldsymbol{\theta}$ 的每个分量 $\theta_i$ 的无偏估计 $\hat{\theta}_i$ 的方差的下限.

**Example 3.5.6.** 设 $X=(X_1,\cdots,X_n)$ 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 $\theta_1$ 和 $\theta_2$ 的无偏估计 $\overline{X}$ 和 $S^2$ 的方差进行比较.

略, 因为本节不要求深究.

### 3.5.4 有效估计和估计的效率

**Definition 3.5.2.** 设 $\hat{g}(\boldsymbol{X})$ 为 $g(\theta)$ 的无偏估计, 比值

$$e_{\hat{g}}(\theta)=\frac{[g'(\theta)]^2/(nI(\theta))}{D_\theta[\hat{g}(\boldsymbol{X})]}$$

称为无偏估计 $\hat{g}(\boldsymbol{X})$ 的效率 (efficiency). 显然 $0<e_{\hat{g}}(\theta) \leqslant 1$, 当 $e_{\hat{g}}(\theta)=1$ 时, 称 $\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的**有效估计** (effective estimation). 若 $\hat{g}(\boldsymbol{X})$ 不是 $g(\theta)$ 的有效估计, 但 $\lim\limits_{n\to\infty} e_{\hat{g}}(\theta)=1$, 则称 $\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的**渐近有效估计** (asymptotically effective estimation).

**Example 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.

**Example 3.5.8.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从 $N(a,\sigma^2)$ 中抽取的简单随机样本,

(1) 当 $a$ 未知时, 样本方差 $S^2$ 不是 $\sigma^2$ 的有效估计, 但是是渐近有效估计.

(2) 当 $a$ 已知时,$\sigma^2$ 的有效估计为 $S_a^2=\sum\limits_{i=1}^{n}(X_i-a)^2/n$.

**Example 3.5.9.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从下列含有位置参数的指数分布族中抽取的简单样本,

$$f(x,a)=e^{-(x-a)I_{(a,\infty)}(x)},-\infty<a<+\infty,$$

$a$ 的 UMVUE 为 $X_{(1)}-\frac{1}{n}$.

# *3.6 概率密度函数的核估计

## 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

先介绍密度函数的一种“自然”估计.

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

$$f(x)=\lim_{h\to 0}\frac{F(x+h)-F(x-h)}{2h}$$

当 $h=h_n$ 充分小时近似有 $f(x)\approx\dfrac{F(x+h)-F(x-h)}{2h}$, 将其中 $F(\cdot)$ 用经验分布函数 $F_n(\cdot)$ 代替就得到

$$f_n(x)=\frac{F_n(x+h_n)-F_n(x-h_n)}{2h_n},\tag{3.6.1}$$

则 $f_n(x)$ 就称为 $f(x)$ 的一个“自然”估计.

### 2. 概率密度估计核函数的定义

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 $I_A$ 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 $X_1,\cdots,X_n$ 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.

**Definition 3.6.1.** 概率密度函数 $f(x)$ 的下述估计量

$$
f_n(x)=\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{x-\boldsymbol{X}_i}{h_n}\right) \tag{3.6.2}
$$

称为核估计 (kernel estimation), 此处 $h_n$ 称为窗宽, 满足条件:$0<h_n\to 0$, 当 $n\to\infty$, 而核函数 $K(\cdot)$ 通常是一个适当的概率密度函数.

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) $\sup\limits_{x\in\mathbb{R}}\{K(x)\}\leqslant M<\infty$, $\lim\limits_{|x|\to\infty}|x|K(x)=0$.

(2) $K(x)=K(-x)$,$x\in\mathbb{R}$, 即 $K(x)$ 对称, 且 $\int_{-\infty}^{\infty}x^2K(x)\mathrm{d}x<\infty$.

(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) $K_1(x)=\begin{cases}\frac{1}{2}, & -1<x\leqslant 1,\\ 0, & \text{其他};\end{cases}$

(ii) $K_2(x)=\begin{cases}1-|x|, & |x|\leqslant 1,\\ 0, & |x|>1;\end{cases}$

(iii) $K_3(x)=(2\pi)^{-1/2}\exp\left\{-\frac{x^2}{2}\right\}$,$x\in\mathbb{R}$, 这是标准正态分布的密度;

(iv) $K_4(x)=\left[\pi(1+x^2)\right]^{-1}$, $x\in\mathbb{R}$, 这是柯西分布 $C(0,1)$ 的密度;

(v) $K_5(x)=\begin{cases}\frac{1}{2\pi}\left[\frac{\sin(x/2)}{x/2}\right]^2, & x\neq 0,\\ \frac{1}{2\pi}, & x=0;\end{cases}$

(vi) $K_6(x)=\begin{cases}\frac{3}{4}\lambda^{-3}(\lambda^2-x^2), & x^2\leqslant\lambda^2,\\ 0, & x^2>\lambda^2,\lambda>0.\end{cases}$

当然还有其他形式的核函数, 他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

概率密度的导函数 $f^{(p)}(x)$ 的核估计与密度函数 $f(x)$ 的核估计的定义并无本质的差别. 现定义如下:

**Definition 3.6.2.** 设 $K_j(\cdot)(j=0,1,2,\cdots)$ 为一列核函数, 则称

$$f_n^{(j)}(x)=\frac{1}{nh_n^{1+j}}\sum_{i=1}^{n}K_j\left(\frac{x-X_j}{h_n}\right),\quad j=0,1,\cdots \tag{3.6.3}$$

为 $f^{(j)}(x)$ 的核估计. 此处 $0<h_n\to 0$, 当 $n\to\infty$.

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.

**Definition 3.6.3.** 设 $\mathscr{F}$ 为一元概率密度族,$X_1,\cdots,X_n,\cdots$ 为从 $\mathscr{F}$ 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 $\mathscr{X}$ 中的每一个 $x$ 和一切 $f\in\mathscr{F}$ 有

(1) 若 $\lim\limits_{n\to\infty}E[f_n(x)]=f(x)$, 则称 $f_n(x)$ 为 $f(x)$ 的**渐近无偏估计**.

(2) 若 $\lim\limits_{n\to\infty}E[f_n(x)-f(x)]^2=0$, 则称 $f_n(x)$ 为 $f(x)$ 的**均方相合估计**.

(3) 当 $n\to\infty$ 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 $n\to\infty$ 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.

设核函数 $K(\cdot)$ 满足下列条件 (I):

(a) $K(\cdot)$ 有界;

(b) $\displaystyle\int_{-\infty}^{\infty}|K(u)|\mathrm{d}u<\infty$;

(c) $|u||K(u)|\to 0$, 当 $|u|\to\infty$;

关于 $f_n(x)$ 的渐近无偏性有下列结果.

**Theorem 3.6.1.** 若 $K(\cdot)$ 满足条件 (I) 中的 (a)-(b), 且 $f(\cdot)$ 在 $x$ 处连续, $\lim\limits_{n\to\infty} h_n = 0$, 则有

$$\lim_{n\to\infty} E[f_n(x)] = f(x).$$

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.

**Theorem 3.6.2.** 若 $K(\cdot)$ 满足条件 (I) 中的 (a)-(c),$f(\cdot)$ 在 $x$ 处连续, 且

$$\lim_{n\to\infty} h_n = 0, \lim_{n\to\infty} nh_n = \infty,$$

则有

(1) $E|f_n(x) - f(x)|^2 \to 0$, 当 $n \to \infty$.

(2) $f_n(x) \xrightarrow{P} f(x)$, 当 $n \to \infty$.

关于 $f_n(x)$ 的强相合性有下列结果.

**Theorem 3.6.3.** 若 $K(\cdot)$ 满足条件 (I) 中的 (a)-(c),$f(\cdot)$ 在 $x$ 连续, 则有

$$f_n(x) \xrightarrow{\text{a.s.}} f(x), 当 n \to \infty.$$

关于 $f_n(x)$ 的渐近正态性有下列结果.

**Theorem 3.6.4.** 若 $K(\cdot)$ 满足条件 (I) 中的 (a)-(c), 且 $\lim\limits_{n\to\infty} h_n = 0, \lim\limits_{n\to\infty} nh_n = \infty$, 则有

$$\frac{f_n(x) - E[f_n(x)]}{\sqrt{\operatorname{Var}(f_n(x))}} \xrightarrow{\mathscr{L}} N(0,1), 当 n \to \infty.$$

**Note 3.6.1.** 概率密度函数核估计的大样本性质对于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) 中的 (a)-(c) 即可.

**Example 3.6.1.** 设 $X_1, \cdots, X_n$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

$$\begin{aligned} E[f_n(x)] =& \frac{1}{h_n} E\left[K\left(\frac{x-X_1}{h_n}\right)\right] = \frac{1}{h_n}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h_n}\right) f(y)\mathrm{d}y \\ =& \int_{-\infty}^{\infty} K(u) f(x - h_n u)\mathrm{d}u. \end{aligned} \tag{3.6.4}$$

由于概率密度函数在数轴 $\mathbb{R}$ 上有界, 故有 $|K(u)f(x-h_nu)| \leqslant c|K(u)|$, 由定理3.6.1条件 (I) 中 (b) 可知 $\int |K(u)|\mathrm{d}u < \infty$. 故积分(3.6.4)关于 $x \in \mathbb{R}$ 上一致收敛, 由参变量积分的性质可知积分(3.6.4)在 $\mathbb{R}$ 上关于变量 $x$ 连续. 因此有

$$\begin{aligned}\lim_{n\to\infty}(E[f_n(x)]-f(x)) &= \lim_{n\to\infty}\int_{-\infty}^{\infty}K(u)[f(x-h_n(u))-f(x)]\mathrm{d}u \\ &= \int_{-\infty}^{\infty}K(u)\lim_{h_n\to 0}[f(x-h_nu)-f(x)]\mathrm{d}u = 0,\end{aligned}$$

即证明了 $\lim\limits_{n\to\infty}E[f_n(x)] = f(x)$,$f_n(x)$ 的渐近无偏性成立.

# Chapter 4

# 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 $\mathscr{F}=\{f(x,\theta),\theta\in\Theta\}$, 其中 $\Theta$ 是参数空间.$\boldsymbol{X}=(X_1,\cdots,X_n)$ 为取自分布族中某总体 $f(x,\theta)$ 的样本,$g(\theta)$ 为定义在 $\Theta$ 上的一个已知函数, 要利用样本 $\boldsymbol{X}$ 对 $g(\theta)$ 的值做出估计, 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经在第三章讨论过了. 在那里使用样本函数 $\hat{g}(\boldsymbol{X})$ 去估计 $g(\theta)$ 的, 称为点估计. 这种估计的缺点是: 单从 $\hat{g}(\boldsymbol{X})$ 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 $d(\boldsymbol{X})$, 把估计写成 $\hat{g}(\boldsymbol{X})\pm d(\boldsymbol{X})$ 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 $[\hat{g}(\boldsymbol{X})-d(\boldsymbol{X}),\hat{g}(\boldsymbol{X})+d(\boldsymbol{X})]$ 之内. 将其一般化, 给出**区间估计**的下列定义.

**Definition 4.1.1.** 设有一个参数分布族 $\mathscr{F}=\{f(x,\theta),\theta\in\Theta\}$,$g(\theta)$ 是定义在参数空间 $\Theta$ 上的一个已知函数,$\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从分布族某总体 $f(x,\theta)$ 中抽取的样本, 令 $\hat{g}_1(\boldsymbol{X})$ 和 $\hat{g}_2(\boldsymbol{X})$ 为定义在样本空间 $\mathscr{X}$ 上, 取值在 $\Theta$ 上的连个统计量, 且 $\hat{g}_1(\boldsymbol{X})\leqslant\hat{g}_2(\boldsymbol{X})$, 则称随机区间 $[\hat{g}_1(\boldsymbol{X}),\hat{g}_2(\boldsymbol{X})]$ 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 $\hat{g}_1\leqslant\hat{g}_2$ 的统计量 $\hat{g}_1,\hat{g}_2$ 都可以构成 $g(\theta)$ 的一个区间估计 $[\hat{g}_1,\hat{g}_2]$. 评价区间估计优劣有两个标准: 可靠度与精度 (精确度). **可靠度**指的是待估参数 $g(\theta)$ 被包含在 $[\hat{g}_1,\hat{g}_2]$ 内的可能性大小. **精度**可以使用随机区间的平均长度来衡量.

### 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 $\theta$ 自身, 这与一般情况没有原则上的区别.

**1. 置信度**

设 $\boldsymbol{X}$ 为样本,$[\hat{\theta}_1(\boldsymbol{X}), \hat{\theta}_2(\boldsymbol{X})]$ 是 $\theta$ 的一个**区间估计**. 由于 $\theta$ 未知且样本随机, 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 $\theta$, 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 $\theta$ 的概率 $P_\theta(\hat{\theta}_1 \leqslant \theta \leqslant \hat{\theta}_2)$ 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 $\theta$ 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 $\Theta$ 中的任一 $\theta$, 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.

**Definition 4.1.2.** 设随机区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 为参数 $\theta$ 的一个区间估计, 则 $[\hat{\theta}_1, \hat{\theta}_2]$ 包含 $\theta$ 的概率 $P_\theta(\hat{\theta}_1 \leqslant \theta \leqslant \hat{\theta}_2)$ 称为此区间估计的**置信水平** (confidence level). 置信水平在参数空间 $\Theta$ 上的下确界

$$\inf_{\theta \in \Theta} P_\theta\left(\hat{\theta}_1 \leqslant \theta \leqslant \hat{\theta}_2\right)$$

称为该区间估计的**置信系数** (confidence coefficient).

**2. 精度**

精度的概念在前面已经说过. **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.

**Example 4.1.1.** 设样本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 来自正态总体 $N(\mu, \sigma^2)$, 其中 $\mu \in \mathbb{R}$,$\sigma^2 > 0$.$\mu$ 和 $\sigma^2$ 的估计量分别是样本均值 $\overline{X}$ 和样本方差 $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$, 用 $[\overline{X} - kS/\sqrt{n}, \overline{X} + kS/\sqrt{n}]$ 作为总体均值 $\mu$ 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

$$\begin{aligned}
&P_\theta\left(\overline{X} - kS/\sqrt{n} \leqslant \mu \leqslant \overline{X} + kS/\sqrt{n}\right) \\
=&P_\theta\left(\left|\sqrt{n}(\overline{X} - \mu)/S\right| \leqslant k\right) \\
=&P(|T| \leqslant k),
\end{aligned}$$

其中 $T = \sqrt{n}(\overline{X} - \mu)/S \sim t_{n-1}$, 其分布与 $\theta$ 无关, 因而区间估计的置信系数为 $P(|T| \leqslant k)$. 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

$$l_k = \frac{2kE(S)}{\sqrt{n}} = \frac{2\sqrt{2}k\sigma\Gamma(n/2)}{\sqrt{n(n-1)}\Gamma((n-1)/2)}.$$

显然 $k$ 越大, 区间也越长, 精度就越差.

**Definition 4.1.3.** 设 $[\hat{\theta}_1(\boldsymbol{X}),\hat{\theta}_2(\boldsymbol{X})]$ 是参数 $\theta$ 的一个**区间估计**, 若对于给定的 $0<\alpha<1$, 有

$$P_\theta\left(\hat{\theta}_1(\boldsymbol{X})\leqslant\theta\leqslant\hat{\theta}_2(\boldsymbol{X})\right)\geqslant 1-\alpha,\theta\in\Theta,$$

则称 $[\hat{\theta}_1(\boldsymbol{X}),\hat{\theta}_2(\boldsymbol{X})]$ 是 $\theta$ 的**置信水平** (confidence level) 为 $1-\alpha$ 的**置信区间** (confidence interval). 而 $\inf\limits_{\theta\in\Theta}P_\theta(\hat{\theta}_1(\boldsymbol{X})\leqslant\theta\leqslant\hat{\theta}_2(\boldsymbol{X}))$ 称为 $\hat{\theta}_1(\boldsymbol{X}),\hat{\theta}_2(\boldsymbol{X})$ 的**置信系数**.

### 4.1.3 置信限

有时我们只关心参数的置信上限或者是置信下限.

**Definition 4.1.4.** 设 $\hat{\theta}_U(\boldsymbol{X})$ 和 $\hat{\theta}_L(\boldsymbol{X})$ 是定义在样本空间 $\mathscr{X}$ 上, 在参数空间 $\theta$ 上取值的两个统计量, 若对给定的 $0<\alpha<1$, 有

$$\begin{aligned}P_\theta\left(\theta\leqslant\hat{\theta}_U(\boldsymbol{X})\right)\geqslant 1-\alpha,\quad \text{一切}\theta\in\Theta,\\ P_\theta\left(\hat{\theta}_L(\boldsymbol{X})\leqslant\theta\right)\geqslant 1-\alpha,\quad \text{一切}\theta\in\Theta,\end{aligned}$$

则分别称 $\hat{\theta}_U(\boldsymbol{X})$ 和 $\hat{\theta}_L(\boldsymbol{X})$ 是 $\theta$ 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 $\Theta$ 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.

**Lemma 4.1.1.** 设 $\hat{\theta}_L(\boldsymbol{X})$ 和 $\hat{\theta}_U(\boldsymbol{X})$ 分别是参数 $\theta$ 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 $\boldsymbol{X}$ 都有 $\hat{\theta}_L(\boldsymbol{X})\leqslant\hat{\theta}_U(\boldsymbol{X})$, 则 $[\hat{\theta}_L(\boldsymbol{X}),\hat{\theta}_U(\boldsymbol{X})]$ 是 $\theta$ 的置信水平为 $1-(\alpha_1+\alpha_2)$ 的双侧置信区间.

### 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 $\theta$ 是一维的, 如果将其推广到参数 $\theta$ 是 $k$ 维 $(k\geqslant 2)$ 的情形, 就得如下定义的置信域.

**Definition 4.1.5.** 设有一个参数分布族 $\mathscr{F}=\{f(x,\boldsymbol{\theta}),\boldsymbol{\theta}\in\Theta\}$,$\Theta$ 是参数空间, 其中 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)\in\Theta\subset\mathbb{R}^k$, $k\geqslant 2.\boldsymbol{X}=(X_1,\cdots,X_n)$ 是来自分布族中某总体 $f(x,\boldsymbol{\theta})$ 的样本. 若统计量 $S(\boldsymbol{X})$ 满足

(1) 对任一样本 $\boldsymbol{X},S(\boldsymbol{X})$ 是 $\Theta$ 的一个子集;

(2) 对给定的 $0<\alpha<1, P_{\boldsymbol{\theta}}(\boldsymbol{\theta}\in S(\boldsymbol{X}))\geqslant 1-\alpha$, 一切 $\boldsymbol{\theta}\in\Theta$, 则称 $S(\boldsymbol{X})$ 是 $\boldsymbol{\theta}$ 的置信水平为 $1-\alpha$ 的**置信域** (confidence region) 或置信集, 而 $\inf\limits_{\boldsymbol{\theta}\in\Theta} P_{\boldsymbol{\theta}}(\boldsymbol{\theta}\in S(\boldsymbol{X}))$ 称为**置信系数**.

在多维场合, 置信域 $S(\boldsymbol{X})$ 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.

**Example 4.2.1.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从总体 $N(\mu,\sigma^2)$ 中抽取的简单随机样本, 此处 $\sigma^2$ 已知, 求 $\mu$ 的置信系数为 $1-\alpha$ 的置信区间.

**解:** 显然,$\mu$ 的一个良好的点估计是 $\overline{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$, 其分布为 $\overline{X}\sim N(\mu,\sigma^2/n)$, 将其标准化得

$$U=\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\sim N(0,1),$$

其分布与 $\mu$ 无关. 由正态分布的对称性, 可得

$$P_\mu\left(\left|\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\right|\leqslant u_{\alpha/2}\right)=1-\alpha,$$

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

$$P_\mu\left(\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right)=1-\alpha.$$

因此 $\left[\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2},\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right]$ 为 $\mu$ 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 $\mu$ 的一个良好点估计, 此例中这个点估计是 $T(\boldsymbol{X})=\overline{X}$.

(2) 构造一个 $T(\boldsymbol{X})$ 和 $\mu$ 的函数 $\varphi(T,\mu)$, 使其满足

(i) 其表达式与待估参数 $\mu$ 有关,

(ii) 其分布与待估参数 $\mu$ 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 $U=\sqrt{n}(\overline{X}-\mu)/\sigma$, 它的表达式与 $\mu$ 有关, 但其分布 $N(0,1)$ 与 $\mu$ 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

$$P_\mu(a\leqslant\varphi(T,u)\leqslant b)=1-\alpha.$$

解括号中的不等式得到 $\hat{\mu}_1(\boldsymbol{X})\leqslant\mu\leqslant\hat{\mu}_2(\boldsymbol{X})$, 则有

$$P_\mu(\hat{\mu}_1(\boldsymbol{X})\leqslant\mu\leqslant\hat{\mu}_2(\boldsymbol{X}))=1-\alpha.$$

这表明 $[\hat{\mu}_1(\boldsymbol{X}),\hat{\mu}_2(\boldsymbol{X})]$ 是 $\mu$ 的置信水平为 $1-\alpha$ 的单侧置信上下限.

类似的步骤可获得 $\mu$ 的置信水平为 $1-\alpha$ 的单侧置信上下限.

4.2.1中的 $\mu$ 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 $\mu$ 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 $\mu$ 和 $\sigma^2$ 的置信区间是实际中长遇到的问题. 下面分几种情况加以讨论. 这里假设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从正态总体 $N(\mu,\sigma^2)$ 中抽取的简单随机样本. 记

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

即 $\overline{X}$ 和 $S^2$ 分别为样本均值和样本方差.

**1.$\sigma^2$ 已知, 求 $\mu$ 的置信区间**

这就是例4.2.1讨论过的问题.$\mu$ 的置信系数为 $1-\alpha$ 置信区间为

$$\left[\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2},\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right], \tag{4.2.1}$$

其单侧置信上, 下限分别为 $\overline{X}+\sigma u_\alpha/\sqrt{n}$ 和 $\overline{X}-\sigma u_\alpha/\sqrt{n}$.

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) $\sigma$ 越大, 则 $l_n$ 越大, 精度越低. 这是因为方差越大, 随机影响也越大, 精度就会低下来.

(3) 置信系数 $1-\alpha$ 越大, 则 $\alpha$ 越小, 从而 $u_{\alpha/2}$ 就越大,$l_n$ 越长, 精度就越低.

由此可见，在 $\sigma$ 和 $\alpha$ 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 $l_n \leqslant l_0$,$l_0$ 为给定的常数，则 $n \geqslant \left[(2\sigma u_{\alpha/2}/l_0)^2\right]+1$，其中 $[x]$ 表示实数 $x$ 的整数部分.

**Example 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 $\mu$ 的 95% 的置信区间.

**解：** 由样本观察值算得 $\overline{X}=13, n=4, \sigma=0.3$，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,$\mu$ 的 95% 的置信区间为

$$\left[\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2}, \overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right]=[12.71, 13.29].$$

**2.$\sigma^2$ 未知，求 $\mu$ 的置信区间**

在这种情况下,$\mu$ 的良好的点估计仍为 $\overline{X}$，由于 $\sigma^2$ 未知，例4.2.1中的随机变量 $U=\sqrt{n}(\overline{X}-\mu)/\sigma$ 在此不能作为枢轴变量，将其中的 $\sigma$ 用 $S$($S^2$ 是样本方差) 代替，得到

$$T=\frac{\sqrt{n}(\overline{X}-\mu)}{S}.$$

由推论2.4.2可知 $T \sim t_{n-1}$. 可见 $T$ 的表达式与 $\mu$ 有关，而其分布与 $\mu$ 无关，故取 $T$ 为枢轴变量. 由于 $T$ 分布关于原点对称，令

$$P(|T| \leqslant c)=P\left(-c \leqslant \frac{\sqrt{n}(\overline{X}-\mu)}{S} \leqslant c\right)=1-\alpha,$$

则 $c=t_{n-1}(\alpha/2)$. 将括号中的不等式经过等价变形得 $\mu$ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left[\overline{X}-\frac{S}{\sqrt{n}}t_{n-1}(\alpha/2), \overline{X}+\frac{S}{\sqrt{n}}t_{n-1}(\alpha/2)\right], \tag{4.2.2}$$

其单侧置信上，下限为 $\overline{X}+St_{n-1}(\alpha)/\sqrt{n}, \overline{X}-St_{n-1}(\alpha)/\sqrt{n}$.

**Example 4.2.3.** 为测得某种溶液中的甲醛浓度，取样得四个独立测定值的平均值 $\overline{X}=8.34\%$，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 $\mu$ 的 95% 的置信区间为

$$\left[\overline{X}-\frac{S}{\sqrt{n}}t_{n-1}(\alpha/2), \overline{X}+\frac{S}{\sqrt{n}}t_{n-1}(\alpha/2)\right]=[8.292\%, 8.388\%].$$

**3.$\mu$ 已知，求 $\sigma^2$ 的置信区间**

当 $\mu$ 已知时,$\sigma^2$ 的一个良好的无偏估计为 $S_\mu^2 = \sum\limits_{i=1}^{n}(X_i-\mu)^2/n$, 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 $\sigma^2$ 有关, 但其分布与 $\sigma^2$ 无关, 找 $c_1$ 和 $c_2$ 使得

$$P_{\sigma^2}\left(c_1 \leqslant \frac{nS_\mu^2}{\sigma^2} \leqslant c_2\right) = 1-\alpha.$$

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 $\chi^2$ 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

$$P_{\sigma^2}\left(\chi_n^2\left(1-\frac{\alpha}{2}\right) \leqslant \frac{nS_\mu^2}{\sigma^2} \leqslant \chi_n^2\left(\frac{\alpha}{2}\right)\right) = 1-\alpha.$$

最后利用不等式的等价变形, 得到 $\sigma^2$ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left[\frac{nS_\mu^2}{\chi_n^2(\alpha/2)}, \frac{nS_\mu^2}{\chi_n^2(1-\alpha/2)}\right], \tag{4.2.3}$$

其单侧置信下上限为 $nS_\mu^2/\chi_n^2(\alpha)$ 和 $nS_\mu^2/\chi_n^2(1-\alpha)$.

**Example 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 $\sigma^2$ 的置信水平为 0.95 的置信区间.

**解:** 此处 $n=6,\mu=30$, 易得出 $\sum\limits_{i=1}^{6}(X_i-\mu)^2 = 0.35, \alpha=0.05$, 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

$$\hat{\sigma}_1^2 = \frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\chi_n^2(\alpha/2)} = \frac{0.35}{14.4494} = 0.0242,$$
$$\hat{\sigma}_2^2 = \frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\chi_n^2(1-\alpha/2)} = \frac{0.35}{1.2375} = 0.2828.$$

因此 $\sigma^2$ 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.$\mu$ 未知，求 $\sigma^2$ 的置信区间**

记 $\theta=(\mu,\sigma^2)$. 此时 $S^2 = \sum\limits_{i=1}^{n}(X_i-\overline{X})^2/(n-1)$ 是 $\sigma^2$ 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 $\sigma^2$ 有关, 而其分

布与 $\sigma^2$ 无关. 找 $d_1$ 和 $d_2$, 使得

$$P_\theta\left(d_1 \leqslant \frac{(n-1)S^2}{\sigma^2} \leqslant d_2\right) = 1-\alpha.$$

类似于 3 中确定 $c_1$ 和 $c_2$ 的理由和方法, 取 $d_1 = \chi^2_{n-1}(1-\alpha/2)$,$d_2 = \chi^2_{n-1}(\alpha/2)$, 故有

$$P_\theta\left(\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right) \leqslant \frac{(n-1)S^2}{\sigma^2} \leqslant \chi^2_{n-1}(\alpha/2)\right) = 1-\alpha.$$

最后再利用不等式的等价变形, 得出 $\sigma^2$ 的置信系数为 $1-\alpha$ 得到置信区间为

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)}, \frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)}\right]. \tag{4.2.4}$$

若把这个随机区间的两个端点开平方, 得到 $\sigma$ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left[\left(\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)}\right)^{1/2}, \left(\frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)}\right)^{1/2}\right].$$

类似可得 $\sigma^2$ 的置信系数为 $1-\alpha$ 单侧置信下上限为 $(n-1)S^2/\chi^2_{n-1}(\alpha)$ 和 $(n-1)S^2/\chi^2_{n-1}(1-\alpha)$.

**Example 4.2.5.** 求例4.2.3中总体方差 $\sigma^2$ 及 $\sigma$ 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 $\chi^2_3(0.025)=9.348$,$\chi^2_3(0.975)=0.216$,$S^2=0.0009$.

由式(4.2.4)可知,$\sigma^2$ 的置信系数为 95% 的置信区间为

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)}, \frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)}\right] = [0.00029, 0.0125],$$

$\sigma$ 的置信系数为 95% 的置信区间为

$$\left[\left(\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)}\right)^{1/2}, \left(\frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)}\right)^{1/2}\right] = [0.017, 0.112].$$

### 4.2.3 两个正态总体参数的置信区间

设 $X_1,\cdots,X_m$ 是自正态总体 $N(a,\sigma_1^2)$ 抽取的简单随机样本,$Y_1,\cdots,Y_n$ 是自正态总体 $N(b,\sigma_2^2)$ 抽取的简单随机样本, 且 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 独立. 设 $\overline{X}$,$\overline{Y}$ 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 $S_1^2=\sum\limits_{i=1}^{m}(X_i-\overline{X})^2/(m-1)$,$S_2^2=\sum\limits_{i=1}^{n}(Y_i-\overline{Y})^2/(n-1)$. 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 $Z_i=Y_i-X_i$, $i=1,2,\cdots,n$, 且记 $\tilde{\mu}=b-a$,$\tilde{\sigma}^2=\sigma_1^2+\sigma_2^2$, 则有

$$Z_i\sim N(\tilde{\mu},\tilde{\sigma}^2),\quad i=1,2,\cdots,n.$$

这就转化为单个正态总体当 $\tilde{\sigma}^2$ 未知, 求其均值 $\tilde{\mu}$ 的置信区间问题. 显见 $\overline{Z}=\overline{Y}-\overline{X}$ 是 $\tilde{\mu}$ 的一个良好的无偏估计, 枢轴变量

$$T_Z=\frac{\sqrt{n}(\overline{Z}-\tilde{\mu})}{S_Z}\sim t_{n-1},$$

此处 $S_Z^2=\sum\limits_{i=1}^{n}(Z_i-\overline{Z})^2/(n-1)$,$T_Z$ 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

$$\left[\overline{Z}-\frac{S_Z}{\sqrt{n}}t_{n-1}(\alpha/2),\overline{Z}+\frac{S_Z}{\sqrt{n}}t_{n-1}(\alpha/2)\right],\tag{4.2.5}$$

其单侧置信下上限分别为 $\overline{Z}-S_Zt_{n-1}(\alpha)/\sqrt{n}$ 和 $\overline{Z}+S_Zt_{n-1}(\alpha)/\sqrt{n}$.

(2) 当 $\sigma_1^2$ 和 $\sigma_2^2$ 已知时, 易知 $\overline{Y}-\overline{X}$ 为 $b-a$ 的一个良好的无偏估计, 且 $\overline{Y}-\overline{X}\sim N(b-a,\sigma_1^2/m+\sigma_2^2/n)$, 将 r.v.$\overline{Y}-\overline{X}$ 标准化, 可知

$$U=\frac{\overline{Y}-\overline{X}-(b-a)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\sim N(0,1).$$

$U$ 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

$$P_{a,b}\left(\left|\frac{\overline{Y}-\overline{X}-(b-a)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\right|\leqslant u_{\alpha/2}\right)=1-\alpha.$$

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

$$\left[\overline{Y}-\overline{X}-u_{\alpha/2}\sqrt{\sigma_1^2/m+\sigma_2^2/n},\overline{Y}-\overline{X}+u_{\alpha/2}\sqrt{\sigma_1^2/m+\sigma_2^2/n}\right].\tag{4.2.6}$$

其单侧置信上下限为 $(\overline{Y}-\overline{X})\pm u_\alpha\sqrt{\sigma_1^2/m+\sigma_2^2/n}$, 其中 "+" 号是置信上限.

(3) 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知时, 令

$$\begin{aligned}S_\omega^2=&\frac{1}{m+n-2}\left[(m-1)S_1^2+(n-1)S_2^2\right]\\=&\frac{1}{m+n-2}\left[\sum_{i=1}^{m}(X_i-\overline{X})^2+\sum_{i=1}^{n}(Y_i-\overline{Y})^2\right],\end{aligned}$$

显然 $\overline{Y}-\overline{X}$ 是 $b-a$ 的无偏估计, 由推论2.4.3可知

$$T_\omega=\frac{\overline{Y}-\overline{X}-(b-a)}{S_\omega}\sqrt{\frac{mn}{m+n}}\sim t_{n+m-2},$$

$T_\omega$ 的表达式与 $b-a$ 有关, 但其分布与 $b-a$ 无关, 取 $T_\omega$ 为枢轴变量, 故有

$$P\left(\left|\frac{\overline{Y}-\overline{X}-(b-a)}{S_\omega}\right|\sqrt{\frac{mn}{m+n}}\leqslant t_{m+n-2}(\alpha/2)\right)=1-\alpha,$$

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

$$\left[\overline{Y}-\overline{X}-S_\omega t_{m+n-2}\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{m}+\frac{1}{n}},\overline{Y}-\overline{X}+S_\omega t_{m+n-2}\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{m}+\frac{1}{n}}\right], \tag{4.2.7}$$

其单侧置信上下限为 $(\overline{Y}-\overline{X})\pm S_\omega t_{m+n-2}(\alpha)\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}$, 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 Scheffè 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

$$\frac{\overline{Y}-\overline{X}-(b-a)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\sim N(0,1), \tag{4.2.8}$$

且当 $m\to\infty$ 时 $S_1^2\xrightarrow{P}\sigma_1^2$, 当 $n\to\infty$ 时 $S_2^2\xrightarrow{P}\sigma_2^2$, 将式(4.2.8)中的 $\sigma_1^2$ 和 $\sigma_2^2$ 分别用 $S_1^2$ 和 $S_2^2$ 代入, 利用引理2.5.1可知, 当 $m,n\to\infty$ 时, 有

$$\widetilde{U}=\frac{\overline{Y}-\overline{X}-(b-a)}{\sqrt{S_1^2/m+S_2^2/n}}=\frac{\overline{Y}-\overline{X}-(b-a)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\cdot\frac{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}{\sqrt{S_1^2/m+S_2^2/n}}\xrightarrow{\mathscr{L}}N(0,1).$$

因此, 取 $\widetilde{U}$ 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

$$\left[\overline{Y}-\overline{X}-u_{\alpha/2}\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}},\overline{Y}-\overline{X}+u_{\alpha/2}\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}\right]. \tag{4.2.9}$$

其单侧置信上下限为 $(\overline{Y}-\overline{X})\pm u_\alpha\sqrt{S_1^2/m+S_2^2/n}$, 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

$$T_*=\frac{\overline{Y}-\overline{X}-(b-a)}{S_*}.$$

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:

$$r=S_*^4\bigg/\left[\frac{S_1^4}{m^2(m-1)}+\frac{S_2^4}{n^2(n-1)}\right].$$

$r$ 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

$$T_*=\frac{\overline{Y}-\overline{X}-(b-a)}{S_*}\sim t_r.$$

运用与前面类似步骤可得 $b-a$ 置信系数近似为 $1-\alpha$ 的置信区间为

$$\left[\overline{Y}-\overline{X}-S_*t_r(\alpha/2),\overline{Y}-\overline{X}+S_*t_r(\alpha/2)\right]. \tag{4.2.10}$$

其单侧置信上下限为 $(\overline{Y}-\overline{X})\pm S_*t_r(\alpha)$, 其中 “+” 号是置信上限.

**Example 4.2.6.** 某公司利用两条自动化流水线灌装瓶装矿泉水. 现从生产线上抽取样本 $X_1,\cdots,X_{12}$ 和 $Y_1,\cdots,Y_{17}$, 它们是每瓶矿泉水的体积 (单位:ml). 算得样本均值 $\overline{X}=501.1$ 和 $\overline{Y}=499.7$; 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:** $\overline{Y}-\overline{X}=-1.4, S_\omega^2=\dfrac{(m-1)S_1^2+(n-1)S_2^2}{n+m-2}=\dfrac{11\times 2.4+16\times 4.7}{12+17-2}=3.763$, 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 $[-2.9,0.1]$.

**Example 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 $X_1,\cdots,X_{120}$ 和 $Y_1,\cdots,Y_{60}$, 其均值分别为 $\overline{X}=3.32, \overline{Y}=3.76, S_1^2=2.18, S_2^2=5.76$. 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 $\overline{X}-\overline{Y}=-0.44$, $\sqrt{S_1^2/m+S_2^2/n}=\sqrt{2.18/120+5.76/60}=0.338$, 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 $\sigma_1^2/\sigma_2^2$ 的置信区间**

(1) 若 $a$ 和 $b$ 已知, 记 $S_a^2=\sum_{i=1}^{m}(X_i-a)/m, S_b^2=\sum_{i=1}^{n}(Y_i-a)^2/n$, 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 $\sigma_1^2$ 的无偏估计,$S_b^2$ 为 $\sigma_2^2$ 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 $\sigma_1^2/\sigma_2^2$ 有关, 但其分布与 $\sigma_1^2/\sigma_2^2$ 无关, 故取 $F$ 为枢轴变量. 找 $c_1$ 和 $c_2$, 使得

$$P\left(c_1\leqslant\frac{S_a^2}{S_b^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}\leqslant c_2\right)=1-\alpha.$$

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

$$P\left(F_{m,n}(1-\alpha/2) \leqslant \frac{S_a^2}{S_b^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leqslant F_{m,n}(\alpha/2)\right) = 1-\alpha.$$

再利用不等式的等价变形, 得到 $\sigma_1^2/\sigma_2^2$ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left[\frac{S_a^2}{S_b^2} \cdot \frac{1}{F_{m,n}(\alpha/2)}, \frac{S_a^2}{S_b^2} \cdot \frac{1}{F_{m,n}(1-\alpha/2)}\right]. \tag{4.2.11}$$

注意到由于 $\alpha$ 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):

$$F_{m,n}(1-\alpha/2) = \frac{1}{F_{m,n}(\alpha/2)}, \tag{4.2.12}$$

可以将其通过查 $F_{n,m}(\alpha/2)$ 算得.

由引理4.1.1可知,$\sigma_1^2/\sigma_2^2$ 的置信系数为 $1-\alpha$ 的单侧置信下上限为

$$\frac{S_a^2}{S_b^2} \cdot \frac{1}{F_{m,n}(\alpha)} \text{和} \frac{S_a^2}{S_b^2} \cdot \frac{1}{F_{m,n}(1-\alpha)} = \frac{S_a^2}{S_b^2} \cdot F_{n,m}(\alpha).$$

(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 $\sigma_1^2$ 和 $\sigma_2^2$ 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 $\sigma_1^2/\sigma_2^2$ 有关, 但其分布与 $\sigma_1^2/\sigma_2^2$ 无关, 故取 $F$ 为枢轴变量. 找 $d_1$ 和 $d_2$, 使得

$$P\left(d_1 \leqslant \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leqslant d_2\right) = 1-\alpha.$$

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

$$P\left(F_{m-1,n-1}(1-\alpha/2) \leqslant \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leqslant F_{m-1,n-1}(\alpha/2)\right) = 1-\alpha.$$

最后利用不等式的等价变形, 得到 $\sigma_1^2/\sigma_2^2$ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left[\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{m-1,n-1}(\alpha/2)}, \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{m-1,n-1}(1-\alpha/2)}\right]. \tag{4.2.13}$$

由于当 $\alpha$ 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,$\sigma_1^2/\sigma_2^2$ 的置信系数为 $1-\alpha$ 的单侧置信下、上限为

$$\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{m-1,n-1}(\alpha)} \quad \text{和} \quad \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{m-1,n-1}(1-\alpha)} = \frac{S_1^2}{S_2^2} \cdot F_{n-1,m-1}(\alpha).$$

**Example 4.2.8.** 求例4.2.7中方差比 $\sigma_1^2/\sigma_2^2$ 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 $\sigma_1^2/\sigma_2^2$ 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

### 4.3.1 小样本方法

**1. 指数分布参数的置信区间**

设 $X_1,\cdots,X_n$ 为从指数分布 $\mathrm{Exp}(\lambda)$ 中抽取的简单随机样本,

其密度函数 $f(x,\lambda)=\lambda e^{-\lambda x}I_{(0,\infty)}(x)$, $\lambda>0$, 要求 $\lambda$ 的置信系数为 $1-\alpha$ 的置信区间.

因为 $\overline{X}=\sum\limits_{i=1}^{n}X_i/n$ 是 $1/\lambda$ 的一个无偏估计 (且是 UMVUE), 设想 $\lambda$ 的置信区间可通过 $\overline{X}$ 表示. 枢轴变量可取为 $T=2\lambda n\overline{X}$, 由推论2.4.5可知

$$2\lambda n\overline{X}=2\lambda\sum_{i=1}^{n}X_i\sim\chi^2_{2n}. \tag{4.3.1}$$

确定 $a,b$ 使得

$$P(a\leqslant 2\lambda n\overline{X}\leqslant b)=1-\alpha.$$

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 $a,b$ 的表达式复杂, 不易求得, 引用上也不方便. 通常采取下列方法, 令

$$P(2\lambda n\overline{X}<a)=\frac{\alpha}{2}, P(2\lambda n\overline{X}>b)=\frac{\alpha}{2}.$$

由式(4.3.1)可知,$a=\chi^2_{2n}(1-\alpha/2)$,$b=\chi^2_{2n}(\alpha/2)$, 这样找到的 $a,b$ 虽然不能使置信区间的精度最高, 但表达式简单, 可通过查 $\chi^2$ 分布的上侧 $\alpha$ 分位数表求得, 应用上很方便. 因此有

$$P\left(\chi^2_{2n}(1-\alpha/2)\leqslant 2\lambda n\overline{X}\leqslant\chi^2_{2n}(\alpha/2)\right)=1-\alpha.$$

利用不等式的等价变形, 可得 $\lambda$ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left[\frac{\chi^2_{2n}(1-\alpha/2)}{2n\overline{X}},\ \frac{\chi^2_{2n}(\alpha/2)}{2n\overline{X}}\right], \tag{4.3.2}$$

其单侧置信下、上限为 $\chi^2_{2n}(1-\alpha)/(2n\overline{X})$ 和 $\chi^2_{2n}(\alpha)/(2n\overline{X})$.

**Example 4.3.1.** 设某电子产品的寿命服从指数分布 $\mathrm{Exp}(\lambda)$, 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 $\overline{X}=59, 2n\overline{X}=1062$. 查表求得

$$\chi^2_{18}(0.05)=28.869,\quad \chi^2_{18}(0.95)=9.390,$$
$$\chi^2_{18}(0.10)=25.989,\quad \chi^2_{18}(0.90)=10.865,$$

则 $g(\lambda)=1/\lambda$ 的置信系数为 90% 的置信区间由下式确定:

$$P_\lambda\left(\chi^2_{18}(0.95)\leqslant 2n\lambda\overline{X}\leqslant \chi^2_{18}(0.05)\right)=0.90.$$

解括号内不等式得 $g(\lambda)=1/\lambda$ 的 90% 的置信区间为

$$\left[\frac{2n\overline{X}}{\chi^2_{18}(0.05)}, \frac{2n\overline{X}}{\chi^2_{18}(0.95)}\right]=[36.787, 113.099].$$

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

$$\hat{g}_U=\frac{2n\overline{X}}{\chi^2_{18}(0.90)}=97.745\text{千小时},$$
$$\hat{g}_L=\frac{2n\overline{X}}{\chi^2_{18}(0.10)}=40.863\text{千小时},$$

**2. 均匀分布参数的置信区间**

设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为自均匀分布总体 $U(0,\theta)$ 中抽取的简单随机样本, 求 $\theta$ 的置信系数为 $1-\alpha$ 的置信区间.

设 $T(\boldsymbol{X})=X_{(n)}=\max\{X_1,\cdots,X_n\}, T(\boldsymbol{X})$ 为充分统计量, 由例3.3.6可知 $(n+1)T/n$ 是 $\theta$ 的无偏估计 (也是 UMVUE 的). 设想枢轴变量一定与 $T$ 有关. 由于 $Y_i=X_i/\theta\sim U(0,1), i=1,2,\cdots,n$, 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 $\theta$ 有关, 其密度函数为

$$\frac{\theta}{T}\sim g(z)=nz^{-(n+1)}I_{(1,\infty)}(z). \tag{4.3.3}$$

$g(z)$ 与 $\theta$ 无关. 确定 $d_1$,$d_2$, $1\leqslant d_1<d_2\leqslant\infty$ 使得

$$\begin{aligned} P_\theta\left(d_1\leqslant\frac{\theta}{T}\leqslant d_2\right)=&P\left(d_1T\leqslant\theta\leqslant d_2T\right)\\ =&\int_{d_1}^{d_2}nz^{-n-1}\mathrm{d}z=\frac{1}{d_1^n}-\frac{1}{d_2^n}=1-\alpha. \end{aligned} \tag{4.3.4}$$

于是 $[d_1T,d_2T]$ 为 $\theta$ 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 $\theta$ 的置信系数为 $1-\alpha$ 的置信区间为

$$[d_1T,d_2T]=\left[T,\frac{T}{\sqrt[n]{\alpha}}\right]. \tag{4.3.5}$$

### 4.3.2 大样本方法

**1.Cauchy 分布未知参数的置信区间**

设 $X_1,\cdots,X_n$ 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

$$f(x,\theta)=\frac{1}{\pi\left[1+(x-\theta)^2\right]},-\infty<x<+\infty,-\infty<\theta<+\infty,$$

要求位置参数 $\theta$ 的置信区间.

以 $m_n$ 记 $X_1,\cdots,X_n$ 的样本中位数,$\theta$ 是总体的中位数, 因此 $m_n$ 作为 $\theta$ 的点估计是合适的. 由于 Cauchy 分布关于 $\theta$ 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 $\theta$ 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

$$P_\theta(|m_n-\theta|\leqslant c)=\int_{-c}^{c}f_n(x)\mathrm{d}x=1-\alpha, \tag{4.3.6}$$

由此得出 $\theta$ 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 $n\to\infty$ 时有

$$\frac{2\sqrt{n}(m_n-\theta)}{\pi}\xrightarrow{\mathscr{L}}N(0,1), \tag{4.3.7}$$

因此有

$$P\left(\frac{2\sqrt{n}|m_n-\theta|}{\pi}\leqslant u_{\alpha/2}\right)\approx 1-\alpha,$$

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,$\theta$ 的置信系数近似为 $1-\alpha$ 的置信区间为

$$\left[m_n-\frac{\pi u_{\alpha/2}}{2\sqrt{n}},m_n+\frac{\pi u_{\alpha/2}}{2\sqrt{n}}\right], \tag{4.3.8}$$

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 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $S_n=\sum\limits_{i=1}^{n}X_i$, 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

$$\frac{S_n-np}{\sqrt{np(1-p)}}=\frac{\sqrt{n}(\overline{X}-p)}{\sqrt{p(1-p)}}\xrightarrow{\mathscr{L}}N(0,1),\quad n\to\infty. \tag{4.3.9}$$

这表明当 $n$ 充分大时, 随机变量 $T=\sqrt{n}(\overline{X}-p)/\sqrt{p(1-p)}$ 的极限分布是 $N(0,1)$, 与位置参数 $p$ 无关. 取这样的 $T$ 为枢轴变量, 因此当 $n$ 充分大时有

$$P(|T|\leqslant u_{\alpha/2})=P\left(\left|\frac{\sqrt{n}(\overline{X}-p)}{\sqrt{p(1-p)}}\right|\leqslant u_{\alpha/2}\right)\approx 1-\alpha. \tag{4.3.10}$$

解上式括号中的不等式 $|T|\leqslant\alpha/2$, 则由式(4.3.10)得到

$$P(|T|\leqslant u_{\alpha/2})=P(c_1\leqslant p\leqslant c_2)\approx 1-\alpha,$$

其中 $c_1=c_1(\boldsymbol{X})$,$c_2=c_2(\boldsymbol{X})$ 与 $p$ 无关. 故 $[c_1(\boldsymbol{X}),c_2(\boldsymbol{X})]$ 可作为 $p$ 的置信系数近似为 $1-\alpha$ 的置信区间. 记 $\hat{p}=\overline{X}$,$\gamma=u_{\alpha/2}$, 则式(4.3.10)括号中的不等式等价于

$$(\hat{p}-p)^2\leqslant\frac{\gamma^2p(1-p)}{n},$$

将不等式两边展开, 整理合并同类项, 得到 $p$ 的二次三项式

$$p^2(n+\gamma^2)-p(2n\hat{p}+\gamma^2)+n\hat{p}^2\leqslant 0.$$

上式左端 $p$ 的二次三项式的判别式大于 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

$$c_1,c_2=\frac{n}{n+\gamma^2}\left[\hat{p}+\frac{\gamma^2}{2n}\pm\gamma\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{\gamma^2}{4n^2}}\right], \tag{4.3.11}$$

其中 $c_1$ 相应于 "$-$" 号. 因此 $[c_1(\boldsymbol{X}),c_2(\boldsymbol{X})]$ 就是 $p$ 的置信系数近似为 $1-\alpha$ 的置信区间.

在实用中, 可采用以下更为简单的方法: 由 $\hat{p}=S_n/n\xrightarrow{\mathrm{P}}p$(即当 $n\to\infty$ 时依概率收敛到 $p$) 及式(4.3.9)可知, 当 $n\to\infty$ 时, 有

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

故由引理2.5.1可知

$$\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}}=\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{p(1-p)}}\cdot\frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}(1-\hat{p})}}\xrightarrow{\mathscr{L}}N(0,1), \tag{4.3.12}$$

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

$$P\left(\left|\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}}\right|\leqslant u_{\alpha/2}\right)\approx 1-\alpha.$$

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

$$\left[\hat{p}-u_{\alpha/2}\cdot\sqrt{\hat{p}(1-\hat{p})/n}, \hat{p}+u_{\alpha/2}\cdot\sqrt{\hat{p}(1-\hat{p})/n}\right], \tag{4.3.13}$$

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 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为抽自 Poisson 总体 $P(\lambda)$ 的简单随机样本, 要求 $\lambda$ 的置信区间. 记 $S_n=\sum\limits_{i=1}^n X_i$, 则 $S_n$ 为服从参数为 $n\lambda$ 的 Poisson 分布, 即

$$P(S_n=k)=\frac{e^{-n\lambda}(n\lambda)^k}{k!},\quad k=0,1,2,\cdots.$$

当 $n$ 充分大时, 由中心极限定理可知

$$\frac{S_n-n\lambda}{\sqrt{n\lambda}}=\frac{\sqrt{n}(\overline{X}-\lambda)}{\sqrt{\lambda}}\xrightarrow{\mathscr{L}}N(0,1),\quad 当 n\to\infty, \tag{4.3.14}$$

将随机变量 $T=\sqrt{n}(\overline{X}-\lambda)/\sqrt{\lambda}$ 作为枢轴变量, 其极限分布与未知参数 $\lambda$ 无关. 记 $\hat{\lambda}=\overline{X}$, 令

$$P(|T|\leqslant u_{\alpha/2})=P\left(\left|\frac{\sqrt{n}(\hat{\lambda}-\lambda)}{\sqrt{\lambda}}\right|\leqslant u_{\alpha/2}\right)\approx 1-\alpha.$$

仿照前一段的方法得到 $\lambda$ 的置信区间 $[d_1,d_2]$, 其中 $d_1=d_1(\boldsymbol{X}),d_2=d_2(\boldsymbol{X})$ 为二次方程

$$(\hat{\lambda}-\lambda)^2=\frac{\lambda u_{\alpha/2}^2}{n}$$

的两根, 即有

$$d_1,d_2=\hat{\lambda}+\frac{u_{\alpha/2}^2}{2n}\pm u_{\alpha/2}\sqrt{\frac{u_{\alpha/2}^2}{4n^2}+\frac{\hat{\lambda}}{n}}, d_1 \text{ 相应于 “}-\text{” 号}, \tag{4.3.15}$$

因此 $\lambda$ 的置信系数近似为 $1-\alpha$ 的置信区间为 $[d_1(\boldsymbol{X}),d_2(\boldsymbol{X})]$.

实用上, 可采用下列更简单的方法: 由 $\hat{\lambda}=\overline{X}\xrightarrow{P}\lambda\,(当 n\to\infty 时)$, 因此当 $n\to\infty$ 时有

$$\frac{\sqrt{n}(\hat{\lambda}-\lambda)}{\sqrt{\lambda}}\xrightarrow{\mathscr{L}}N(0,1), \frac{\sqrt{\lambda}}{\sqrt{\hat{\lambda}}}\xrightarrow{P}1.$$

故由引理2.5.1可知

$$\frac{\sqrt{n}(\hat{\lambda}-\lambda)}{\sqrt{\hat{\lambda}}}=\frac{\sqrt{n}(\hat{\lambda}-\lambda)}{\sqrt{\lambda}}\cdot\frac{\sqrt{\lambda}}{\sqrt{\hat{\lambda}}}\xrightarrow{\mathscr{L}}N(0,1). \tag{4.3.16}$$

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 $\lambda$ 无关. 给定置信系数 $1-\alpha$, 则有

$$P\left(\frac{\left|\sqrt{n}(\hat{\lambda}-\lambda)\right|}{\sqrt{\hat{\lambda}}}\leqslant u_{\alpha/2}\right)\approx 1-\alpha.$$

由不等式的等价变形, 得到 $\lambda$ 的置信系数近似为 $1-\alpha$ 的置信区间为

$$\left[\hat{\lambda}-u_{\alpha/2}\sqrt{\hat{\lambda}/n}, \hat{\lambda}+u_{\alpha/2}\sqrt{\hat{\lambda}/n}\right] \tag{4.3.17}$$

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 $\theta$ 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 $\theta$ 的极大似然估计 $\hat{\theta}_n^* = \hat{\theta}_n^*(X_1, \cdots, X_n)$ 有渐近正态分布 $\sqrt{n}(\hat{\theta}_n^* - \theta) \xrightarrow{\mathscr{L}} N(0, 1/I(\theta))$, 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 $\theta$, 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 $n \to \infty$ 时, 有

$$\frac{\sqrt{n}(\hat{\theta}_n^* - \theta)}{\sigma(\hat{\theta}_n^*)} = \frac{\sqrt{n}(\hat{\theta}_n^* - \theta)}{\sigma(\theta)} \cdot \frac{\sigma(\theta)}{\sigma(\hat{\theta}_n^*)} \xrightarrow{\mathscr{L}} N(0,1). \tag{4.3.18}$$

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 $\theta$ 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 $\theta$ 的置信系数近似为 $1-\alpha$ 的置信区间

$$\left[\hat{\theta}_n^* - \frac{u_{\alpha/2}}{\sqrt{n}}\,\sigma(\hat{\theta}_n^*),\ \hat{\theta}_n^* + \frac{u_{\alpha/2}}{\sqrt{n}}\,\sigma(\hat{\theta}_n^*)\right]. \tag{4.3.19}$$

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 $\theta_F$ 和方差 $\sigma_F^2$, 且 $\theta_F$ 和 $\sigma_F^2$ 皆未知. 从这一总体中抽取简单随机样本 $X_1, \cdots, X_n$, 要求 $\theta_F$ 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 $\sqrt{n}(\overline{X} - \theta_F)/\sigma_F \xrightarrow{\mathscr{L}} N(0,1)$, 当 $n \to \infty$ 时. 但此处 $\sigma_F$ 未知, 仍不能以 $\sqrt{n}(\overline{X} - \theta_F)/\sigma_F$ 作为枢轴变量. 因为 $n$ 充分大, 样本标准差 $S$ 是 $\sigma_F$ 的一个相合估计, 故可近似地用 $S$ 代替 $\sigma_F$, 利用引理2.5.1可证, 当 $n \to \infty$ 时, 有

$$\frac{\sqrt{n}(\overline{X} - \theta_F)}{S} = \frac{\sqrt{n}(\overline{X} - \theta_F)}{\sigma_F} \cdot \frac{\sigma_F}{S} \xrightarrow{\mathscr{L}} N(0,1). \tag{4.3.20}$$

此时可将 $T = \sqrt{n}(\overline{X} - \theta_F)/S$ 作为枢轴变量, 它的极限分布与 $\theta_F$ 无关, 令

$$P\left(\left|\frac{\sqrt{n}(\overline{X} - \theta_F)}{S}\right| \leqslant u_{\alpha/2}\right) \approx 1-\alpha,$$

解上式括号中的不等式, 得到 $\theta_F$ 的置信系数近似为 $1-\alpha$ 的置信区间为

$$\left[\overline{X} - \frac{S}{\sqrt{n}}u_{\alpha/2}, \overline{X} + \frac{S}{\sqrt{n}}u_{\alpha/2}\right], \tag{4.3.21}$$

其单侧置信上下限为 $\overline{X} \pm S\, u_\alpha/\sqrt{n}$, 其中 “+” 号是置信上限.

## *4.4 Fisher 的信仰推断法

### 4.4.1 引言

本章前几节讨论了关于构造 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科尔莫戈洛夫公理体系下的概率论的, 在实用上允许给出频率的解释. 在本章的后面两节介绍另外两种区间估计的方法.

因此如果能通过样本对这种信仰分布 (概率) 作一数量上的刻画 (即利用样本描述参数分布于不同区间的可能性), 则就可利用它作出 $\theta$ 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

**Example 4.4.1.** 设样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为自 $N(\theta,1)$ 中抽取的简单随机样本, 其中 $-\infty<\theta<+\infty$, 求 $\theta$ 的信仰区间.

**解:** 由于 $\overline{X}$ 是 $\theta$ 的充分统计量, 所以可基于 $\overline{X}$ 来考虑该问题. 因为 $\overline{X}\sim N(\theta,1/n)$, 若记随机误差 $e\sim N(0,1)$, 则等价地有

$$\overline{X}=\theta+\frac{1}{\sqrt{n}}e. \tag{4.4.1}$$

移项可得

$$\theta=\overline{X}-\frac{1}{\sqrt{n}}e. \tag{4.4.2}$$

由式(4.4.1)到式(4.4.2), 从通常概率的观点来看啊, 不过是写法不同, 但是 Fisher 赋予了式(4.4.2)以全新的解释: 有了样本 $\boldsymbol{X}$ 之后继而得到 $\overline{X}$(因而 $\overline{X}$ 视为已知数), 把 $\theta$ 看成随机变量, 则 $\theta\sim N(\overline{X},1/n)$,Fisher 称这个分布为信仰分布.

记 $\widetilde{P}$ 表示信仰概率,$P$ 为通常意义下的概率, 利用信仰分布(4.4.2), 对任意的 $a$,$b(a<b)$, 算出事件 $\{a\leqslant\theta\leqslant b\}$ 的信仰概率为

$$\widetilde{P}(a\leqslant\theta\leqslant b)=\int_a^b\frac{\sqrt{n}}{\sqrt{2\pi}}e^{\frac{-n(y-\overline{x})^2}{2}}\mathrm{d}y. \tag{4.4.3}$$

由(4.4.2)可知 $\sqrt{n}(\theta-\overline{X})=-e\sim N(0,1)$,$e$ 是关于 0 对称分布, 故 $-e$ 与 $e$ 同分布. 令

$$\widetilde{P}\left(\left|\sqrt{n}(\theta-\overline{X})\right|\leqslant c\right)=P\left(|e|\leqslant c\right)=1-\alpha,$$

则 $c=u_{\alpha/2}$. 等价地有

$$\widetilde{P}\left(\overline{X}-\frac{u_{\alpha/2}}{\sqrt{n}}\leqslant\theta\leqslant\overline{X}+\frac{u_{\alpha/2}}{\sqrt{n}}\right)=1-\alpha,$$

其中 $1-\alpha$ 称为**信仰系数** (fiducial coefficient).$\left[\overline{X}-u_{\alpha/2}/\sqrt{n},\ \overline{X}+u_{\alpha/2}/\sqrt{n}\right]$ 称为 $\theta$ 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 $\theta$ 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, 虽然它们的解释完全不同. 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.

**Example 4.4.2.** 设样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\theta$ 和 $\sigma^2$ 的信仰区间.

**解:** 设 $\overline{X}$ 为样本均值. 记

$$Q^2=(n-1)S^2=\sum_{i=1}^{n}(X_i-\overline{X})^2, \tag{4.4.4}$$

其中 $S^2$ 为样本方差, 由于 $(\overline{X},Q^2)$ 是 $(\theta,\sigma^2)$ 的充分统计量, 所以可以基于 $(\overline{X},Q^2)$ 来考虑问题. 因 $\overline{X}\sim N(\theta,\sigma^2/n),Q^2/\sigma^2\sim\chi^2_{n-1}$, 并且 $\overline{X}$ 和 $Q^2$ 相互独立. 若记 $e_1\sim N(0,1)$,$e_2\sim\chi^2_{n-1}$ 且 $e_1,e_2$ 相互独立, 则等价地有

$$\begin{cases}\overline{X}=\theta+\frac{\sigma}{\sqrt{n}}\cdot e_1,\\ Q^2=\sigma^2\cdot e_2.\end{cases} \tag{4.4.5}$$

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

$$\begin{cases}\theta=\overline{X}-\frac{1}{\sqrt{n}}\frac{Q}{\sqrt{e_2}}\cdot e_1,\\ \sigma^2=\frac{Q^2}{e_2}.\end{cases} \tag{4.4.6}$$

在有了样本 $\boldsymbol{X}$, 从而有了 $(\overline{X},Q^2)$ 后, 可诱导出 $(\theta,\sigma^2)$ 的联合信仰分布. 由于 $e_1\sim N(0,1)$,$e_2\sim\chi^2_{n-1}$ 且 $e_1$ 和 $e_2$ 相互独立, 则 $(\theta,\sigma^2)$ 联合信仰分布为

$$\begin{aligned}p(\theta,\tau)=c\cdot\exp\left\{-\frac{n(\theta-\overline{X})^2+Q^2}{2\tau}\right\}\cdot(\tau)^{-\frac{n+2}{2}},\\ -\infty<\mu<+\infty,\tau>0,\end{aligned} \tag{4.4.7}$$

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 $\theta$ 和 $\sigma^2$ 的边缘分布. 当然, 也可以直接由式(4.4.6)分别导出 $\theta$ 和 $\sigma^2$ 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

$$\frac{\sqrt{n}(\theta-\overline{X})}{S}=-\frac{e_1}{\sqrt{e_2/(n-1)}}\sim t_{n-1}.$$

因此 $\theta$ 的边缘信仰分布为自由度 $n-1$ 的 $t$ 分布. 若令

$$\widetilde{P}\left(\left|\frac{\sqrt{n}(\theta-\overline{X})}{S}\right|\leqslant c\right)=P\left(\left|\frac{e_1}{\sqrt{e_2/(n-1)}}\right|\leqslant c\right)=1-\alpha,$$

则 $c=t_{n-1}(\alpha/2)$. 上式等价于

$$\widetilde{P}\left(\overline{X}-\frac{S}{\sqrt{n}}\,t_{n-1}(\alpha/2)\leqslant\theta\leqslant\overline{X}+\frac{S}{\sqrt{n}}\,t_{n-1}(\alpha/2)\right)=1-\alpha.$$

因此 $\theta$ 的信仰系数为 $1-\alpha$ 的置信区间为

$$\left[\overline{X}-\frac{S}{\sqrt{n}}t_{n-1}(\alpha/2),\ \overline{X}+\frac{S}{\sqrt{n}}t_{n-1}(\alpha/2)\right].$$

这与 4.2 节中求得的 $\theta$ 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

$$\frac{\sigma^2}{Q^2}=\frac{1}{e_2},\quad e_2\sim\chi^2_{n-1}.$$

因此 $\sigma^2$ 的边缘信仰分布是一个逆 $\chi^2$ 分布. 若令

$$\widetilde{P}\left(c\leqslant\frac{\sigma^2}{Q^2}\leqslant d\right)=P\left(\frac{1}{d}\leqslant e_2\leqslant\frac{1}{c}\right)=1-\alpha,$$

由于 $Q^2/\sigma^2=e_2\sim\chi^2_{n-1}$, 故可取 $1/d=\chi^2_{n-1}(1-\alpha/2)$, $1/c=\chi^2_{n-1}(\alpha/2)$, 即有

$$\widetilde{P}\left(\frac{Q^2}{\chi^2_{n-1}(\alpha/2)}\leqslant\sigma^2\leqslant\frac{Q^2}{\chi^2_{n-1}(1-\alpha/2)}\right)=1-\alpha.$$

因此 $\sigma^2$ 的信仰系数为 $1-\alpha$ 的信仰区间为

$$\left[\frac{Q^2}{\chi^2_{n-1}(\alpha/2)},\ \frac{Q^2}{\chi^2_{n-1}(1-\alpha/2)}\right].$$

这与 4.2 节中求得的 $\sigma^2$ 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 Neyan 的理论中,$\theta$ 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 $\theta$ 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 $\boldsymbol{X} = (X_1, \cdots, X_m)$ 是从正态总体 $N(\mu_1, \sigma_1^2)$ 中抽取的简单随机样本,$\boldsymbol{Y} = (Y_1, \cdots, Y_n)$ 是从正态总体 $N(\mu_2, \sigma_2^2)$ 中抽取的简单随机样本, 其中 $-\infty \leqslant \mu_1, \mu_2 \leqslant +\infty, \sigma_1^2,\sigma_2^2 > 0$ 为未知参数, 且合样本 $\boldsymbol{X}, \boldsymbol{Y}$ 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,Scheffè 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 Scheffè 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 Scheffè 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 $\overline{X}$,$\overline{Y}$,$S_1^2$ 和 $S_2^2$ 分别表示样本 $X$ 和 $Y$ 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 $\overline{X} \sim N(\mu_1, \sigma_1^2/m)$,$\overline{Y} \sim N(\mu_2, \sigma_2^2/n)$,$Q_1^2/\sigma_1^2 \sim \chi_{m-1}^2$,$Q_2^2/\sigma_2^2 \sim \chi_{n-1}^2$; 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 $e_2$,$\xi_1$ 和 $\xi_2$ 相互独立, 则等价地有

$$\begin{cases} \overline{X} = \mu_1 + \frac{\sigma_1}{\sqrt{m}} \cdot e_1, \\ Q_1^2 = \sigma_1^2 \cdot \xi_1, \\ \overline{Y} = \mu_2 + \frac{\sigma_2}{\sqrt{n}} \cdot e_2, \\ Q_2^2 = \sigma_2^2 \cdot \xi_2. \end{cases} \tag{4.4.8}$$

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

$$\begin{cases}\mu_1 = \overline{X} - \frac{Q_1}{\sqrt{m}} \cdot \frac{e_1}{\sqrt{\xi_1}},\\ \sigma_1^2 = Q_1^2/\xi_1,\\ \mu_2 = \overline{Y} - \frac{Q_2}{\sqrt{n}} \cdot \frac{e_2}{\sqrt{\xi_2}},\\ \sigma_2^2 = Q_2^2/\xi_2.\end{cases} \tag{4.4.9}$$

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

$$\begin{cases}\mu_1 = \overline{X} - \frac{S_1}{\sqrt{m}}t_1 = \overline{X} - S_1^* t_1,\\ \mu_2 = \overline{Y} - \frac{S_2}{\sqrt{n}}t_2 = \overline{Y} - S_2^* t_2,\end{cases} \tag{4.4.10}$$

其中 $t_1 = \dfrac{e_1}{\sqrt{\xi_1/(m-1)}} \sim t_{m-1}$,$t_2 = \dfrac{e_2}{\sqrt{\xi_2/(n-1)}} \sim t_{n-1}$,$S_1^* = S_1/\sqrt{m}$,$S_2^* = S_2/\sqrt{n}$. 记 $Z = \overline{Y} - \overline{X}$,$\theta = \mu_2 - \mu_1$, 将式(4.4.10)中的两式相减, 得到

$$Z - \theta = S_2^* t_2 - S_1^* t_1, \quad \theta = Z - (S_2^* t_2 - S_1^* t_1). \tag{4.4.11}$$

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 $\theta$ 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 $\varphi$. 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 $\theta$ 的信仰分布(4.4.11)可得

$$\widetilde{P}\left(|\theta - Z| \leqslant \gamma\, y_{m,n,\varphi,\alpha}\right) = P\left(|W| \leqslant y_{m,n,\varphi,\alpha}\right) = 1 - \alpha.$$

此处 $\widetilde{P}$ 为信仰概率,$P$ 是通常意义下的概率. 上式的等价的形式为

$$\widetilde{P}\left(Z - \gamma y_{m,n,\varphi,\alpha} < \theta < Z + \gamma y_{m,n,\varphi,\alpha}\right) = 1 - \alpha. \tag{4.4.12}$$

因此,$\theta = \mu_2 - \mu_1$ 的信仰系数为 $1 - \alpha$ 的信仰区间为

$$\left[Z - \gamma y_{m,n,\varphi,\alpha}, Z + \gamma y_{m,n,\varphi,\alpha}\right].$$

注意在有了样本并给定 $\alpha$ 后,$Z$,$\gamma$,$y_{m,n,\varphi,\alpha}$ 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 $\rho = \sigma_1/\sigma_2$, 他就 $m = 12, n = 6$ 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 容忍区间和容忍限

### 4.5.1 问题的提法及定义

设有分布族 $\{F_\theta, \theta\in\Theta\}$,$\Theta$ 是参数空间,$\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从分布族中某总体抽取的简单样本. 本节不是考虑参数 $\theta$ 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 $T_1=T_1(\boldsymbol{X})$ 和 $T_2=T_2(\boldsymbol{X})$, 使得对给定的 $0<\beta<1$,

$$P_\theta^*\{X\in[T_1(\boldsymbol{X}),T_2(\boldsymbol{X})]\}\geqslant 1-\beta, \tag{4.5.1}$$

其中 $P_\theta^*$ 应理解为在给定 $\boldsymbol{X}$ 的条件下关于 r.v.$X$ 的分布计算的条件概率. 请看以下的两个例子.

**Example 4.5.1.** 设某轴承厂有一部自动化机器, 生产直径为 0.25cm 的轴承, 允许误差 0.001cm. 生产中要求 99% 的产品达到以上的规定, 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 $X_1,\cdots,X_n$, 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 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 确定两个统计量 $T_1=T_1(\boldsymbol{X})$ 和 $T_2=T_2(\boldsymbol{X})$, 使得 $P_\theta^*(X\in[T_1,T_2])=F_\theta(T_2)-F_\theta(T_1)\geqslant 0.99$, 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.

**Example 4.5.2.** 钢厂生产某种钢材, 要求其钢材的强度不少于 $\xi_0$(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 $X_1,\cdots,X_n$, 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 $X=(X_1,\cdots,X_n)$ 确定一个统计量 $T_L=T_L(\boldsymbol{X})$, 使得 $P_\theta^*(X\in[T_L,\infty))=1-F_\theta(T_L)\geqslant 0.99$, 若 $T_L\geqslant\xi_0$, 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 $X=(X_1,\cdots,X_n)$ 为从此分布族中抽取的简单样本, 要找到两个统计量 $T_1=T_1(\boldsymbol{X})$ 和 $T_2=T_2(\boldsymbol{X})$,

使得对 $0<\beta<1$ 有

$$P_\theta^*\left(T_1 \leqslant X \leqslant T_2\right)=F_\theta(T_2)-F_\theta(T_1) \geqslant 1-\beta, \tag{4.5.2}$$

或找一个统计量 $T_L=T_L(\boldsymbol{X})$, 使得对 $0<\beta<1$ 有

$$P_\theta^*\left(T_L \leqslant X<+\infty\right)=1-F_\theta(T_L) \geqslant 1-\beta, \tag{4.5.3}$$

或找一个统计量 $T_U=T_U(\boldsymbol{X})$, 对 $0<\beta<1$ 有

$$P_\theta^*\left(-\infty<X \leqslant T_U\right)=F_\theta(T_U) \geqslant 1-\beta. \tag{4.5.4}$$

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 “$F(T_2)-F(T_1) \geqslant 1-\beta$” 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 “$F(T_2)-F(T_1) \geqslant 1-\beta$” 这个事件至少以概率 $1-r$ 成立, 即

$$P_\theta\left\{F(T_2)-F(T_1) \geqslant 1-\beta\right\} \geqslant 1-r.$$

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.

**Definition 4.5.1.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从总体 $X \sim \{F_\theta,\theta \in \Theta\}$ 中抽取的简单样本. 又设 $T_1=T_1(\boldsymbol{X})$ 和 $T_2=T_2(\boldsymbol{X})$ 是两个统计量, 且 $T_1 \leqslant T_2$, 若对任意给定的 $\beta,\gamma$(通常取较小的数, 如 $\beta=0.05,\gamma=0.01$),$0<\beta,\gamma<1$ 有

$$\begin{aligned}&P_\theta\left\{P_\theta^*(T_1 \leqslant X \leqslant T_2) \geqslant 1-\beta\right\}\\=&P_\theta\left\{F_\theta(T_2)-F_\theta(T_1) \geqslant 1-\beta\right\} \geqslant 1-\gamma, \quad 一切\theta \in \Theta,\end{aligned} \tag{4.5.5}$$

则称 $[T_1,T_2]$ 是 $F_\theta$ 的一个水平为 $(1-\beta,1-\gamma)$ 的**容忍区间** (tolerance interval). 设 $T_L=T_L(\boldsymbol{X})$ 和 $T_U=T_U(\boldsymbol{X})$ 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

$$\begin{aligned}P_\theta\left\{P_\theta^*(T_L \leqslant X) \geqslant 1-\beta\right\}=&P_\theta\left\{1-F_\theta(T_L) \geqslant 1-\beta\right\}\\=&P_\theta\left\{F_\theta(T_L) \leqslant \beta\right\} \geqslant 1-\gamma,\end{aligned} \tag{4.5.6}$$

$$P_\theta\left\{P_\theta^*(X \leqslant T_U) \geqslant 1-\beta\right\}=P_\theta\left\{F_\theta(T_U) \geqslant 1-\beta\right\} \geqslant 1-\gamma, \tag{4.5.7}$$

则称 $T_L$ 和 $T_U$ 分别是 $F_\theta$ 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 $P_\theta$ 的含义是不同的, 里面的 $P_\theta^*$ 的意义如式 (4.5.1)中所述, 外面的 $P_\theta$ 是按样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 的联合分布来计算的.

容忍区间和容忍限之间有下列关系.

**Lemma 4.5.1.** 若 $T_2 = T_2(\boldsymbol{X})$ 和 $T_1 = T_1(\boldsymbol{X})$ 分别是分布 $F_\theta$ 的水平为 $(1-\beta/2, 1-\gamma/2)$ 的容忍上下限, 且总有 $T_2 \geqslant T_1$, 则 $[T_1, T_2]$ 为 $F_\theta$ 的水平为 $(1-\beta, 1-\gamma)$ 的容忍区间.

**证:** 令 $A$ 表示事件 "$F(T_1) \leqslant \beta/2$";$B$ 表示事件 "$F(T_2) \geqslant 1-\beta/2''$;$C$ 表示事件 "$F(T_2) - F(T_1) \geqslant 1-\beta$", 则由定义4.5.1可知

$$P_\theta(A) \geqslant 1-\gamma/2, \quad P_\theta(B) \geqslant 1-\gamma/2. \tag{4.5.8}$$

希望证明 $P_\theta(C) \geqslant 1-\gamma$. 由以上定义可知, 若 $A, B$ 同时成立, 则必有 $F(T_2) - F(T_1) \geqslant 1-\beta$, 即 $C$ 成立, 因此 $AB \subset C$, 故有 $P_\theta(C) \geqslant P_\theta(AB)$, 由此可得

$$\begin{aligned}
P_\theta(C) =& P_\theta\left\{F_\theta(T_2) - F_\theta(T_1) \geqslant 1-\beta\right\} \geqslant P(AB) \\
=& P(A) + P(B) - P(A \cup B) \\
\geqslant& (1-\gamma/2) + (1-\gamma/2) - 1 = 1-\gamma.
\end{aligned}$$

引理证毕.

### 4.5.2 正态总体的容忍区间和容忍限

设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 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

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2.$$

将基于充分统计量 $(\overline{X}, S^2)$ 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 $\mu$ 和 $\sigma^2$ 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 $\mu$ 和 $\sigma^2$ 未知, 知道 $\overline{X}$ 和 $S^2$ 分别是 $\mu$ 和 $\sigma^2$ 的良好估计, 因此将上述容忍上限中的 $\mu$ 和 $\sigma$ 用 $\overline{X}$ 和 $S$ 代替得到 $\overline{X} + Su_\beta$. 由于估计带来的随机性, 水平 $(1-\beta, 1-\gamma)$ 的容忍上限不见得正好是 $\overline{X} + Su_\beta$, 而可能要将系数 $u_\beta$ 修改为某个 $\lambda$,$\lambda$ 既与 $\beta$ 有关, 也与 $\gamma$ 有关 (注意 $u_\beta$ 与 $\gamma$ 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 $\lambda$ 使 $\overline{X} + \lambda S$ 为水平 $(1-\beta, 1-\gamma)$ 的容忍上限. 按定义, 对给定的 $\beta$ 和 $\gamma$,$0 < \beta, \gamma < 1$, 要确定 $\lambda$, 使得

$$P_\theta\left\{P_\theta^*\left(X \leqslant \overline{X} + \lambda S\right) \geqslant 1-\beta\right\} \geqslant 1-\lambda.$$

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

$$\begin{aligned}
&P_\theta\left\{P_\theta^*\left(\frac{X-\mu}{\sigma} \leqslant \frac{\overline{X} - \mu + \lambda S}{\sigma}\right) \geqslant 1-\beta\right\} \\
=& P_\theta\left\{\Phi\left(\frac{\overline{X} - \mu + \lambda S}{\sigma}\right) \geqslant 1-\beta\right\} \\
=& P_\theta\left\{\frac{\overline{X} - \mu + \lambda S}{\sigma} \geqslant \varphi^{-1}(1-\beta) = u_\beta\right\}.
\end{aligned} \tag{4.5.9}$$

令 $Z=\sqrt{n}(\overline{X}-\mu)/\sigma$,$S_*=S/\sigma$, 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

$$\begin{aligned}&P_\theta\left\{P_\theta^*\left(X\leqslant\overline{X}+\lambda S\right)\geqslant1-\beta\right\}\geqslant1-\gamma\\ \Leftrightarrow&P_\theta\left\{\frac{\overline{X}-\mu+\lambda S}{\sigma}\geqslant u_\beta\right\}\geqslant1-\gamma\\ \Leftrightarrow&P_\theta\left\{\frac{Z-\sqrt{n}u_\beta}{S_*}\geqslant-\sqrt{n}\lambda\right\}\geqslant1-\gamma.\end{aligned}\tag{4.5.10}$$

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 $\overline{X}+\lambda S$, 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 $\overline{X}-\lambda S$,$\lambda$ 同上. 对常见的 $n,\lambda,\gamma$ 已编制了 $\lambda(n,\beta,\gamma)$ 值的表, 见附录 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 $\left[\overline{X}-\lambda S,\overline{X}+\lambda S\right]$, 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.

**Example 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, 测得其抗拉强度 (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

$$\overline{X}=10632.4,\quad S^2=6738.77,\quad S=82.09.$$

查附表 6 得 $\lambda=2.91$, 因此得容许下限 $T_L=\overline{X}-\lambda S=10632.4-2.91\times82.09=10393.52$. 因此这批镍合金线抗拉强度不低于 10393.52kg/mm$^2$.

**Example 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

$$228.6,232.7,238.8,317.2,315.8,275.1,$$
$$222.2,236.7,224.7,251.2,210.4,270.7.$$

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

$$\overline{X}=252.0,\quad S^2=1263.4,\quad S=35.5.$$

查附表 7 得 $\lambda(12,0.95,0.99)=3.87$, 由此算得

$$T_1=\overline{X}-\lambda S=252.0-3.87\times 35.5=114.6,$$
$$T_2=\overline{X}+\lambda S=252.0+3.87\times 35.5=389.4.$$

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.

**Lemma 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 $t=\inf\{x:F(x)\geqslant y\}$, 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 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从总体 $X\sim F(x)$ 中抽取的简单随机样本,$X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 为其次序统计量，根据引理4.5.2可知，若记 $U_i=F(X_i)$,$i=1,2,\cdots,n$, 则 $U_1,\cdots,U_n$i.i.d$\sim U(0,1)$.$U_{(1)}\leqslant U_{(2)}\leqslant\cdots\leqslant U_{(n)}$ 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 $U_i=F(X_i)$,$i=1,2,\cdots,n$;$V_{ij}=U_{(j)}-U_{(i)}$,$1\leqslant i<j\leqslant n$, 则 $V_{ij}$ 的密度 $g_{ij}$ 已由式(2.3.13)给出. 若以 $\left[X_{(i)},X_{(j)}\right]$ 作为容忍区间，按定义有

$$P\left\{P^*(X_{(i)}\leqslant X\leqslant X_{(j)})\geqslant 1-\beta\right\}\geqslant 1-\gamma. \tag{4.5.11}$$

此处 $P^*$ 的意义与式(4.5.1)中 $P^*_\theta$ 的意义相同. 上式左边为

$$P\left\{F(X_{(j)})-F(X_{(i)})\geqslant 1-\beta\right\}=P\left(V_{ij}\geqslant 1-\beta\right)=\int_{1-\beta}^{1}g_{ij}(v)\mathrm{d}v. \tag{4.5.12}$$

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 $\mathrm{Be}(j-i,n-j+i+1)$.Beta 分布的参数不必是整数，只要大于 0 就行. 当 $>0,q>0$ 时,$\mathrm{Be}(p,q)$ 表示一分布，其分布函数为

$$I_{p,q}(x)=\frac{1}{B(p,q)}\int_0^x t^{p-1}(1-t)^{q-1}\mathrm{d}t, \tag{4.5.13}$$

其中 $B(p,q)=\int_0^1 t^{p-1}(1-t)^{q-1}\mathrm{d}t$, 称为 Beta 积分. 它与 Gamma 函数 $\Gamma(x)$ 有关系 $B(p,q)=\Gamma(p)\Gamma(q)/\Gamma(p+q)$, 这个公式在微积分中已经给出.

式(4.5.13)当 $0<x<1$ 时, 称为**不完全 Beta 积分**,Pearson 曾给他造了表. 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

$$P\left\{P^*\left(X\leqslant X_{(i)}\right)\geqslant 1-\beta\right\}\geqslant 1-\gamma,$$

此式左边是 $P\{F(X_{(i)})\geqslant 1-\beta\}=P(U_{(i)}\geqslant 1-\beta)$. 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

$$P\left\{F(X_{(i)})\geqslant 1-\beta\right\}=P\left(U_{(i)}\geqslant 1-\beta\right)=\int_{1-\beta}^1 f_i(v)\mathrm{d}v. \tag{4.5.14}$$

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

$$P\left\{F(X_{(j)})\leqslant\beta\right\}=P\left(U_{(j)}\leqslant\beta\right)=\int_0^\beta f_j(v)\mathrm{d}v\geqslant 1-\gamma, \tag{4.5.15}$$

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 函数表求得.

**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

$$P\left\{F(X_{(n)})\geqslant 1-\beta\right\}=P(U_{(n)}\geqslant 1-\beta)\geqslant 1-\gamma,$$

即

$$\int_{1-\beta}^1 ny^{n-1}\mathrm{d}y\geqslant 1-\gamma\Leftrightarrow 1-(1-\beta)^n\geqslant 1-\gamma.$$

因此有

$$n\geqslant\frac{\ln\gamma}{\ln(1-\beta)}.$$

对给定的 $\beta,\gamma$, 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

$$P\left\{F(X_{(1)})\leqslant\beta\right\}=P\left(U_{(1)}\leqslant\beta\right)=\int_0^\beta n(1-y)^{n-1}\mathrm{d}y\geqslant 1-\gamma,$$

同样解出 $n\geqslant\ln\gamma/\ln(1-\beta)$.

对通常用的 $\beta,\gamma$, 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

$$P\left\{F(X_{(n)})-F(X_{(1)})\geqslant 1-\beta\right\}=P\left(U_{(n)}-U_{(1)}\geqslant 1-\beta\right)\geqslant 1-\gamma. \tag{4.5.16}$$

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

$$\begin{aligned}\iint\limits_{y_2-y_1\geqslant 1-\beta}p(y_1,y_2)\mathrm{d}y_1\mathrm{d}y_2=&\int_0^\beta\int_{y_1+(1-\beta)}^1 n(n-1)(y_2-y_1)^{n-2}\mathrm{d}y_2\mathrm{d}y_1\\ \geqslant&1-\gamma,\end{aligned}$$

解之可得

$$n(1-\beta)^{n-1}-(n-1)(1-\beta)^n\leqslant\gamma.$$

对给定的 $\beta$ 和 $\gamma$(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 $n=130$.

# Chapter 5

# 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. 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 (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)$\theta$, 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

$$H_0 : a = a_0 \leftrightarrow H_1 : a \neq a_0 \text{或} H_0 : a \leqslant a_0 \leftrightarrow H_1 : a > a_0.$$

(2) 非参数型假设检验: 如果总体分布形式未知, 此时就需要一种与总体分布族的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.

## 5.1 假设检验的若干基本概念

### 5.1.1 检验问题的提法

为了说明检验问题的提法, 考察下面的例子.

**Example 5.1.1.** 某工厂生产的一大铺产品, 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, 问这批产品是否可出厂?

关于这个问题, 在面前存在两种可能性:

$$\text{甲} : 0 < p \leqslant 0.01, \qquad \text{乙} : 0.01 < p < 1.$$

要通过从这批产品中抽样来决定甲, 乙两种可能性中哪一个成立.

由于种种原因, 这个问题常常以下述方式提出: 引进一个“假设”

$$H_0 : 0 < p \leqslant 0.01,$$

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**或是**备择假设** (alternative hypothesis).

目的是要通过样本决定接受 $H_0$ 该是拒绝 $H_0$. 可以形象地把问题写成

$$H_0 : 0 < p \leqslant 0.01 \leftrightarrow H_1 : 0.01 < p < 1.$$

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 $\Theta$ 为参数空间.$X_1,\cdots,X_n$ 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 $\theta$ 是否属于参数空间 $\Theta$ 的某个非空真子集 $\Theta_0$. 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

$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1, \tag{5.1.1}$$

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 $\Theta_0$ 或 $\Theta_1$ 只包含参数空间 $\Theta$ 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 $N(a,\sigma^2)$,$\sigma^2$ 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 $H_0' : a \leqslant a_0 \leftrightarrow H : a > a_0$, 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

**Example 5.1.2.** 设 $\boldsymbol{X} = (X_1,\cdots,X_n)$ 为从正态总体 $N(a,\sigma^2)$ 中抽取的随机样本, 其中 $\sigma^2$ 已知. 考虑检验问题

$$H_0 : a = a_0 \leftrightarrow H_1 : a \neq a_0, \tag{5.1.2}$$

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 $\overline{X} = \sum\limits_{i=1}^{n} X_i/n$ 是 $a$ 的一个优良估计. 若 $\left|\overline{X} - a_0\right|$ 较大, 就倾向于否定 $H_0$; 反之, 如果 $\left|\overline{X} - a_0\right|$ 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 $X_1,\cdots,X_n$ 算出样本均值 $\overline{X}$, 当 $\left|\overline{X} - a_0\right| > A$ 时就否定 $H_0$; 当 $|\overline{X} - a_0| \leqslant A$ 时就接受 $H_0$. 称

$$D = \left\{X = (X_1,\cdots,X_n) : |\overline{X} - a_0| > A\right\} \tag{5.1.3}$$

为**否定域**, 或**拒绝域** (reject region), 即否定域是由样本空间 $\mathscr{X}$ 中一切使 $|\overline{X} - a_0| > A$ 的那些样本 $X = (X_1,\cdots,X_n)$ 构成. 有了否定域, 等价于将样本空间 $\mathscr{X}$ 分成不相交的两部

分 $D$ 和 $\overline{D}=\mathscr{X}-D$. 一旦有了样本 $\boldsymbol{X}$, 当 $\boldsymbol{X}\in D$ 时, 就否定 $H_0$; 当 $X\in\overline{D}$ 时, 就接受 $H_0$. 称 $\overline{D}$ 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:

$$T:\begin{cases}\text{当}|\overline{X}-a_0|>A\text{时, 拒绝}H_0,\\ \text{当}|\overline{X}-a_0|\leqslant A\text{时, 接受}H_0.\end{cases}$$

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** $\varphi(\boldsymbol{x})$ 的概念,$\varphi(x)$ 与检验 $T$ 是一一对应的. 在例5.1.2中,

$$\varphi(\boldsymbol{x})=\begin{cases}1, & \text{当 } |\overline{x}-a_0|>A,\\ 0, & \text{当 } |\overline{x}-a_0|\leqslant A.\end{cases}\tag{5.1.4}$$

有如下定义.

**Definition 5.1.1.** 由式(5.1.4)给出的**检验函数** $\varphi(\boldsymbol{x})$ 是定义在**样本空间** $\mathscr{X}$ 上, 取值于 $[0,1]$ 上的函数. 它表示当有了样本 $\boldsymbol{X}$ 后, 否定 $H_0$ 的概率.

由定义可见, 若 $\varphi(\boldsymbol{x})=1$, 则以概率为 1 否定 $H_0$, 若 $\varphi(\boldsymbol{x})=0$, 则以概率为 0 否定 $H_0$(即以概率为 1 接受 $H_0$). 若 $\varphi(\boldsymbol{x})$ 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:$D=\{\boldsymbol{X}=(X_1,\cdots,X_n):\varphi(\boldsymbol{X})=1\}$.

若对某些样本 $\boldsymbol{X}$, 有 $0<\varphi(\boldsymbol{x})<1$, 则称 $\varphi(\boldsymbol{x})$ 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 $0$,$i=1,\cdots,n$. 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为样本, 令 $T(\boldsymbol{X})=\sum\limits_{i=1}^{100}X_i$, 当 $T(\boldsymbol{X})<c$ 时认为这批产品合格, 接受 $H_0$; 当 $T(\boldsymbol{X})>c$ 时, 认为不合格, 拒绝 $H_0$. 当 $T(\boldsymbol{X})=c$ 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 $T(\boldsymbol{X})=c$ 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 $T(\boldsymbol{X})=c$, 双方都有 $1/2$ 的可能, 做出对自己不利的决定, 双方都觉得合理, 可以接受. 如果取 $r=1/3$, 试验可以通过装有两个白球和一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

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})>c,\\ r, & T(\boldsymbol{x})=c,\\ 0, & T(\boldsymbol{x})<c.\end{cases}\tag{5.1.5}$$

在例5.1.2中要确定检验, 必须定出式(5.1.3)或(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 $\overline{X}$. 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 $T(\boldsymbol{X})=\sum\limits_{i=1}^{100} X_i$. 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 假设检验和功效函数

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 $\overline{D}$, 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:

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})>3,\\ 0, & T(\boldsymbol{x})\leqslant 3.\end{cases}$$

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 $T(\boldsymbol{x})=5$, 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是次品率 $p=0.03>0.01$, 由于样本的随机性, 抽样结果显示 $T(\boldsymbol{x})=1$, 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.

**Definition 5.1.2.** 设 $\varphi(\boldsymbol{x})$ 是 $H_0:\theta\in\Theta_0 \leftrightarrow H_1:\theta\in\Theta_1$ 的一个检验函数, 则

$$\beta_\varphi(\theta)=P_\theta\left\{\text{用检验}\varphi\text{否定了}H_0\right\}=E_\theta[\varphi(\boldsymbol{X})],\quad \theta\in\Theta$$

称为 $\varphi$ 的**功效函数** (power function), 也称为效函数或势函数.

若 $\varphi(\boldsymbol{x})$ 为非随机化检验, 否定域为 $D$, 则

$$\beta_\varphi(\theta)=P_\theta\left\{\boldsymbol{X}=(X_1,\cdots,X_n)\in D\right\},\quad \theta\in\Theta.$$

因此功效函数表示当参数为 $\theta$ 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

$T(\boldsymbol{X}) = \sum\limits_{i=1}^{n} X_i \sim b(n,\theta), 0<\theta<1$ 可知检验的功效函数为

$$\begin{aligned}\beta_\varphi(\theta) =& E_\theta[\varphi(\boldsymbol{X})] = P\left(\sum_{i=1}^{100} X_i > c\right) + rP\left(\sum_{i=1}^{100} X_i = c\right)\\ =& \sum_{k=c+1}^{100}\binom{100}{k}\theta^k(1-\theta)^{100-k} + r\binom{100}{c}\theta^c(1-\theta)^{100-c}.\end{aligned}$$

以下讨论中假定 $\varphi(\boldsymbol{x})$ 皆为非随机化的检验函数, 除非特别申明, 不认为 $\varphi(\boldsymbol{x})$ 为随机化检验函数.

知道了检验 $\varphi(\boldsymbol{x})$ 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

$$\beta_\varphi^*(\theta) = \begin{cases}0, & \theta\in\Theta_0,\\ 1-\beta_\varphi(\theta), & \theta\in\Theta_1.\end{cases}$$

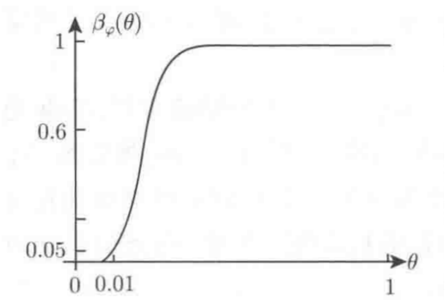



图 5.1.1: 功效函数

如在例5.1.1中,$\Theta=(0,1)$, 则检验问题 $H_0: 0<\theta\leqslant 0.01 \leftrightarrow H_1: 0.01<\theta<1$ 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 $\Theta_0$ 中应尽可能地小, 在 $\Theta_1$ 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 $\alpha(0<\alpha<1$, 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

$$S_\alpha = \{\varphi: \beta_\varphi(\theta)\leqslant\alpha, \theta\in\Theta_0\},$$

$S_\alpha$ 表示由所有犯第一类错误的概率都不超过 $\alpha$ 的检验函数构成的类. 只考虑 $S_\alpha$ 中的检验, 在 $S_\alpha$ 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.

**Definition 5.1.3.** 设 $\varphi$ 是式(5.1.1)的一个检验, 而 $0 \leqslant \alpha \leqslant 1$. 如果 $\varphi$ 犯第一类错误的概率总不超过 $\alpha$(或等价地说,$\varphi$ 满足 $\beta_\varphi(\theta) \leqslant \alpha$, 一切 $\theta \in \Theta_0$), 则称 $\alpha$ 是检验 $\varphi$ 的一个水平, 而 $\varphi$ 称为**显著性水平**为 $\alpha$ 的检验, 简称水平为 $\alpha$ 的检验.

按这一定义, 检验的水平不唯一. 若 $\alpha$ 为检验 $\varphi$ 的水平, 而 $\alpha < \alpha' < 1$, 则 $\alpha'$ 也是检验 $\varphi$ 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

$$\text{检验 } \varphi \text{ 的真实水平} = \sup\{\beta_\varphi(\theta), \theta \in \Theta_0\}. \tag{5.1.6}$$

至于水平的选择, 习惯上把 $\alpha$ 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 $\alpha$ 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 $\alpha$ 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 $\alpha$ 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, 确定检验统计量 $T(\boldsymbol{X})$, 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 $\boldsymbol{X}$ 算出检验统计量 $T(\boldsymbol{X})$ 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

从略.

## 5.2 正态总体的假设检验

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 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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)$H_0':\mu\leqslant\mu_0 \leftrightarrow H_1':\mu>\mu_0$,

(3)$H_0'':\mu\geqslant\mu_0 \leftrightarrow H_1'':\mu<\mu_0$,

其中 $\mu_0$ 和检验水平 $\alpha$ 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 $\sigma^2$ 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 $\overline{X} = \sum\limits_{i=1}^{n} X_i/n$ 是 $\mu$ 的无偏估计且具有良好性质. 直观上看 $|\overline{X} - \mu_0|$ 越大,$H_0$ 越不像成立—也就是说, 成立的可能性越低. 因此, 检验的否定域可取如下形式:$\{\boldsymbol{X} = (X_1, \cdots, X_n) : |\overline{X} - \mu_0| > A\}$,$A$ 待定. 当 $\sigma^2$ 已知时, 由定理2.2.3可知, 在 $H_0$ 成立的条件下 $\overline{X} \sim N(\mu_0, \sigma^2/n)$, 将 $\overline{X}$ 标准化得到

$$U = \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} \sim N(0, 1). \tag{5.2.1}$$

因此取 $U = \sqrt{n}(\overline{X} - \mu_0)/\sigma$ 作为检验统计量, 则否定域的等价形式可取为

$$\{(X_1, \cdots, X_n) : |U| > c\}, c\text{待定}.$$

由检验水平为 $\alpha$, 可知

$$P(|U| > c|H_0) = P\left(\left|\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma}\right| > c\middle|H_0\right) = \alpha.$$

由于 $U|\mu = \mu_0 \sim N(0, 1)$, 故得 $c = u_{\alpha/2}$. 因此由否定域

$$D_1 = \left\{(X_1, \cdots, X_n) : |U| > u_{\alpha/2}\right\} \tag{5.2.2}$$

确定的检验为检验问题 (1) 的水平为 $\alpha$ 的检验.

(ii) 当方差 $\sigma^2$ 已知时, 考虑检验问题 (2) 的水平为 $\alpha$ 的检验, 即

$$H_0' : \mu \leqslant \mu_0 \leftrightarrow H_1' : \mu > \mu_0.$$

从直观上看检验问题 (2) 的否定域为 $\{(X_1, \cdots, X_n) : \sqrt{n}(\overline{X} - \mu_0)/\sigma > c\}$, 其中 $c$ 待定. 此时仍取 $U = \sqrt{n}(\overline{X} - \mu_0)/\sigma$ 作为检验统计量. 令

$$P\big(U > c|H_0'\big) = P\left(\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} > c\middle|\mu = \mu_0\right) = \alpha, \tag{5.2.3}$$

由于 $U|\mu = \mu_0 \sim N(0, 1)$, 故取 $c = u_\alpha$. 得到否定域

$$D_2^* = \left\{(X_1, \cdots, X_n) : U = \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} > u_\alpha\right\}. \tag{5.2.4}$$

由上式给出的否定域是否为检验问题 (2) 的检验水平为 $\alpha$ 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 $\alpha$. 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 $\alpha$? 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 $\mu$ 的非降函数, 则它必为检验问题 (2) 的检验水平为 $\alpha$ 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 $\mu$ 的非降性可知 $\beta_{D_2^*}(\mu) \leqslant \beta_{D_2^*}(\mu_0) = \alpha$, 对一切 $\mu \leqslant \mu_0$. 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 $\alpha$ 的否定域.

下面证明功效函数 $\beta_{D_2^*}(\mu)$ 关于 $\mu$ 的非降性. 记 $\Phi(\cdot)$ 为 $N(0,1)$ 的分布函数, 则

$$\begin{aligned}\beta_{D_2^*}(\mu) =& P_\mu\left(\frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma} > u_\alpha\right)\\ =& P_\mu\left(\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} > u_\alpha + \frac{\sqrt{n}(\mu_0-\mu)}{\sigma}\right)\\ =& 1-\Phi\left(u_\alpha + \frac{\sqrt{n}(\mu_0-\mu)}{\sigma}\right).\end{aligned} \tag{5.2.5}$$

显然由(5.2.5)确定的功效函数 $\beta_{D_2^*}(\mu)$ 是 $\mu$ 的非降函数, 这就证明了所要的结论. 因此由

$$D_2 = D_2^* = \{(X_1, \cdots, X_n) : U > u_\alpha\} \tag{5.2.6}$$

确定的检验是检验问题 (2) 的水平为 $\alpha$ 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 $\alpha$ 的检验的否定域

$$D_3 = \{(X_1, \cdots, X_n) : U < -u_\alpha\}$$

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 $\sigma^2$ 未知且假定 $\mu=\mu_0$ 时, 由推论2.4.2可知

$$T = \frac{\sqrt{n}(\overline{X}-\mu_0)}{S} \sim t_{n-1}. \tag{5.2.7}$$

因此取 $T=\sqrt{n}(\overline{X}-\mu_0)/S$ 作为检验统计量, 用完全与方差 $\sigma^2$ 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 $\alpha$ 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 |
|---|---|---|---|---|
| $\sigma^2$ 已知 | $\mu=\mu_0$ | $\mu\neq\mu_0$ | $U=\sqrt{n}(\overline{X}-\mu_0)/\sigma$<br>$U\vert\mu=\mu_0\sim N(0,1)$ | $\vert U\vert > u_{\alpha/2}$ |
| | $\mu\leqslant\mu_0$ | $\mu>\mu_0$ | | $U>u_\alpha$ |
| | $\mu\geqslant\mu_0$ | $\mu<\mu_0$ | | $U<-u_\alpha$ |
| $\sigma^2$ 未知 | $\mu=\mu_0$ | $\mu\neq\mu_0$ | $T=\sqrt{n}(\overline{X}-\mu_0)/S$<br>$T\vert\mu=\mu_0\sim t_{n-1}$ | $\vert T\vert > t_{n-1}(\alpha/2)$ |
| | $\mu\leqslant\mu_0$ | $\mu>\mu_0$ | | $T>t_{n-1}(\alpha)$ |
| | $\mu\geqslant\mu_0$ | $\mu<\mu_0$ | | $T<-t_{n-1}(\alpha)$ |

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.

**Example 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 $\sigma^2$ 已知，故否定域由表5.2.1中第一行给出，即

$$D = \left\{ (X_1, \cdots, X_n) : \left| \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} \right| > u_{\alpha/2} \right\},$$

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 $\overline{X} = 502$. 因此

$$|U| = \left| \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} \right| = \left| \frac{\sqrt{10}(502 - 500)}{6.5} \right| = 0.973 < 1.96,$$

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 $H_0$.

**Example 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: $\mathrm{kg/cm^2}$) 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 $32.50\mathrm{kg/cm^2}$(取检验水平 $\alpha = 0.05$)?

**解:** 检验问题为

$$H_0 : \mu \geqslant 32.50 \leftrightarrow H_1 : \mu < 32.50.$$

本题中 $\sigma^2$ 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

$$D = \left\{ (X_1, \cdots, X_n) : \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} < -t_{n-1}(\alpha) \right\},$$

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 $\overline{X} = 31.13$,$S = 1.123$，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 $32.50\mathrm{kg/cm^2}$.

### 5.2.2 单个正态总体方差的检验

设 $X_1, \cdots, X_n$ 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:

$(4) H_0 : \sigma^2 = \sigma_0^2 \leftrightarrow H_1 : \sigma^2 \neq \sigma_0^2$,

$(5) H_0' : \sigma^2 \leqslant \sigma_0^2 \leftrightarrow H_1' : \sigma^2 > \sigma_0^2$,

$(6) H_0'' : \sigma^2 \geqslant \sigma_0^2 \leftrightarrow H_1'' : \sigma^2 < \sigma_0^2$,

其中 $\sigma_0^2$ 和检验水平 $\alpha$ 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 $\mu$ 未知时 $S^2 = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$ 是 $\sigma^2$ 的一个无偏估计, 且具有良好性质. 直观上看 $S^2/\sigma_0^2$ 太小或者 $S^2/\sigma_0^2$ 太大时,$H_0$ 不像成立. 因此检验的否定域可取如下形式:$\{(X_1, \cdots, X_n) : S^2/\sigma_0^2 < A_1 或 S^2/\sigma_0^2 > A_2\}$,$A_1$,$A_2$ 待定. 在给定 $\sigma^2 = \sigma_0^2$ 的条件下, 由定理2.2.3可知

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2. \tag{5.2.8}$$

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

$$D = \left\{(X_1, \cdots, X_n) : \frac{(n-1)S^2}{\sigma_0^2} < c_1 \text{或} \frac{(n-1)S^2}{\sigma_0^2} > c_2\right\}.$$

记 $\theta = (\mu, \sigma^2)$, 为了确定 $c_1$,$c_2$, 令

$$\alpha = P_\theta\left(\frac{(n-1)S^2}{\sigma_0^2} < c_1 \text{或} \frac{(n-1)S^2}{\sigma_0^2} > c_2 \middle| H_0\right),$$

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

$$P_\theta\left(\frac{(n-1)S^2}{\sigma_0^2} < c_1 \middle| H_0\right) = \alpha/2, \quad P_\theta\left(\frac{(n-1)S^2}{\sigma_0^2} > c_2 \middle| H_0\right) = \alpha/2.$$

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 $\alpha$ 的接受域为

$$\overline{D}_4 = \left\{(X_1, \cdots, X_n) : \chi_{n-1}^2\left(1 - \frac{\alpha}{2}\right) \leqslant \frac{(n-1)S^2}{\sigma_0^2} \leqslant \chi_{n-1}^2\left(\frac{\alpha}{2}\right)\right\}.$$

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 $\alpha$ 的否定域如下:

$$D_5 = \left\{(X_1, \cdots, X_n) : \frac{(n-1)S^2}{\sigma_0^2} > \chi_{n-1}^2(\alpha)\right\},$$

$$D_6 = \left\{(X_1, \cdots, X_n) : \frac{(n-1)S^2}{\sigma_0^2} < \chi_{n-1}^2(1-\alpha)\right\}.$$

**2. 当均值 $\mu$ 已知时方差 $\sigma^2$ 的检验方法**

简述如下: 当 $\mu$ 已知时,$\sigma^2$ 的一个具有良好性质的无偏估计是 $S_\mu^2=\sum\limits_{i=1}^{n}(X_i-\mu)^2/n$. 当 $\sigma^2=\sigma_0^2$ 时, 由推论2.4.1可知

$$\frac{nS_\mu^2}{\sigma_0^2}=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma_0}\right)^2\sim\chi_n^2. \tag{5.2.9}$$

因此取检验统计量为 $\chi_\mu^2=nS_\mu^2/\sigma_0^2$, 用它代替 $\chi^2=(n-1)S^2/\sigma_0^2$. 采用完全类似于前面 $\mu$ 未知情形的讨论方法, 可得检验问题 (4)-(6) 的水平为 $\alpha$ 的否定域, 仅注意在否定域中将确定临界值的 $\chi^2$ 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 |
|---|---|---|---|---|
| $\mu$ 已知 | $\sigma^2=\sigma_0^2$ | $\sigma^2\neq\sigma_0^2$ | $X_\mu^2=nS_\mu^2/\sigma_0^2$ | $nS_\mu^2/\sigma_0^2<\chi_n^2(1-\alpha/2)$ 或$nS_\mu^2/\sigma_0^2>\chi_n^2(\alpha/2)$ |
| | $\sigma^2\leqslant\sigma_0^2$ | $\sigma^2>\sigma_0^2$ | $X_\mu^2\|\sigma_0^2\sim\chi_n^2$ | $nS_\mu^2/\sigma_0^2>\chi_n^2(\alpha)$ |
| | $\sigma^2\geqslant\sigma_0^2$ | $\sigma^2<\sigma_0^2$ | | $nS_\mu^2/\sigma_0^2<\chi_n^2(1-\alpha)$ |
| $\mu$ 未知 | $\sigma^2=\sigma_0^2$ | $\sigma^2\neq\sigma_0^2$ | $\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}$ | $(n-1)S^2/\sigma_0^2<\chi_{n-1}^2(1-\alpha/2)$ 或$(n-1)S^2/\sigma_0^2>\chi_{n-1}^2(\alpha/2)$ |
| | $\sigma^2\leqslant\sigma_0^2$ | $\sigma^2>\sigma_0^2$ | $\chi^2\|\sigma_0^2\sim\chi_{n-1}^2$ | $(n-1)S^2/\sigma_0^2>\chi_{n-1}^2(\alpha)$ |
| | $\sigma^2\geqslant\sigma_0^2$ | $\sigma^2<\sigma_0^2$ | | $(n-1)S^2/\sigma_0^2<\chi_{n-1}^2(1-\alpha)$ |

这种基于检验统计量服从一定自由度的 $\chi^2$ 分布的检验方法称为 $\chi^2$ **检验**.

**Example 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

$$H_0:\sigma^2=1.44\leftrightarrow H_1:\sigma^2\neq1.44.$$

检验的接受域为

$$\overline{D}=\left\{(X_1,\cdots,X_n):\chi_{n-1}^2(1-\alpha/2)<\frac{(n-1)S^2}{\sigma_0^2}<\chi_{n-1}^2(\alpha/2)\right\},$$

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

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}=\frac{15\times4.41}{1.44}=45.94>27.488.$$

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 $\boldsymbol{X}=(X_1,\cdots,X_m)$ 为自正态总体 $N(\mu_1,\sigma_1^2)$ 中抽取的简单随机样本,$\boldsymbol{Y}=(Y_1,\cdots,Y_n)$ 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 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 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)$H_0':\mu_2-\mu_1\leqslant\mu_0 \leftrightarrow H_1':\mu_2-\mu_1>\mu_0$,

(9)$H_0'':\mu_2-\mu_1\geqslant\mu_0 \leftrightarrow H_1'':\mu_2-\mu_1<\mu_0$,

其中 $\mu_0$ 和检验水平 $\alpha$ 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 $\sigma_1^2$ 和 $\sigma_2^2$ 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 $\overline{Y}-\overline{X}$ 是 $\mu_2-\mu_1$ 的一个良好估计, 直观上看 $|\overline{Y}-\overline{X}-\mu_0|$ 越大,$H_0$ 越不像成立, 故否定域可取如下形式 $\big\{(\boldsymbol{X},\boldsymbol{Y}):|\overline{Y}-\overline{X}-\mu_0|>A\big\}$, $A$ 待定. 由于 $\overline{Y}-\overline{X}\sim N(\mu_2-\mu_1,\sigma_1^2/m+\sigma_2^2/n)$, 故当 $\mu_2-\mu_1=\mu_0$ 时有

$$U=\frac{\overline{Y}-\overline{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\sim N(0,1). \tag{5.2.10}$$

因此, 取检验统计量为 $U=(\overline{Y}-\overline{X}-\mu_0)/\sqrt{\sigma_1^2/m+\sigma_2^2/n}$, 则否定域等价形式为
$\Big\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):|U|>c\Big\}$,$c$ 待定.

记 $\theta=\mu_2-\mu_1$, 为确定 $c$, 令

$$\alpha=P_\theta(|U|>c|H_0)=P\left(\left|\frac{\overline{Y}-\overline{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\right|>c\middle|H_0\right)=2-2\Phi(c),$$

由此可确定临界值 $c=u_{\alpha/2}$. 因此检验问题 (7) 的水平为 $\alpha$ 的检验否定域为

$$D_7=\Big\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):|U|>u_{\alpha/2}\Big\},$$

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 $\alpha$ 的检验的否定域为

$$D_8=\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):U>u_\alpha\},$$
$$D_9=\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):U<-u_\alpha\}.$$

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

$$U=\frac{\overline{Y}-\overline{X}-\mu_0}{\sqrt{\sigma^2/m+\sigma^2/n}}=\frac{\overline{Y}-\overline{X}-\mu_0}{\sigma}\sqrt{\frac{mn}{m+n}}. \tag{5.2.11}$$

而在 $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知时, 上述表达式中的 $\sigma^2$ 常用

$$\begin{aligned} S_w^2=&\frac{1}{n+m-2}\left[(m-1)S_1^2+(n-1)S_2^2\right]\\ =&\frac{1}{n+m-2}\left[\sum_{i=1}^m(X_i-\overline{X})^2+\sum_{j=1}^n(Y_j-\overline{Y})^2\right]\end{aligned} \tag{5.2.12}$$

来估计, 其中

$$S_1^2=\frac{1}{m-1}\sum_{i=1}^m(X_i-\overline{X})^2,\quad S_2^2=\frac{1}{n-1}\sum_{j=1}^n(Y_j-\overline{Y})^2 \tag{5.2.13}$$

分别为样本 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 的样本方差. 当 $\mu_2-\mu_1=\mu_0$ 时, 将式(5.2.11)中的 $\sigma$ 用(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

$$T_w=\frac{\overline{Y}-\overline{X}-\mu_0}{S_w}\sqrt{\frac{mn}{m+n}}\sim t_{n+m-2}. \tag{5.2.14}$$

因此取 $T_w$ 作为检验统计量.

与 5.2.1 节中样本 $t$ 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 $\alpha$ 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 |
|---|---|---|---|---|
| $\sigma_1^2$ $\sigma_2^2$ 已知 | $\mu_2-\mu_1=\mu_0$ | $\mu_2-\mu_1\neq\mu_0$ | $U=\frac{\overline{Y}-\overline{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}$ | $U>u_{\alpha/2}$ |
| | $\mu_2-\mu_1\leqslant\mu_0$ | $\mu_2-\mu_1>\mu_0$ | | $U>u_\alpha$ |
| | $\mu_2-\mu_1\geqslant\mu_0$ | $\mu_2-\mu_1<\mu_0$ | $U\|\mu_0\sim N(0,1)$ | $U<-u_\alpha$ |
| $\sigma_1^2$ $\sigma_2^2$ 未知 | $\mu_2-\mu_1=\mu_0$ | $\mu_2-\mu_1\neq\mu_0$ | $T_w=\frac{\overline{Y}-\overline{X}-\mu_0}{S_w}\sqrt{\frac{mn}{m+n}}$ | $\|T\|>t_{m+n-2}\left(\frac{\alpha}{2}\right)$ |
| | $\mu_2-\mu_1\leqslant\mu_0$ | $\mu_2-\mu_1>\mu_0$ | $T_w\|\mu_0\sim t_{m+n-2}$ | $T_w>t_{m+n-2}(\alpha)$ |
| | $\mu_2-\mu_1\geqslant\mu_0$ | $\mu_2-\mu_1<\mu_0$ | $S_w^2=\frac{(m-1)S_1^2+(n-1)S_2^2}{n+m-2}$ | $T<-t_{m+n-2}(\alpha)$ |

这种基于检验统计量服从 $t_{m+n-2}$ 分布的检验方法, 称为**两样本 $t$ 检验**.

**Example 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

$$\begin{aligned}&男: \overline{X}=465.13万/\mathrm{mm}^3, S_1=54.80万/\mathrm{mm}^3,\\ &女: \overline{Y}=422.16万/\mathrm{mm}^3, S_2=49.20万/\mathrm{mm}^3.\end{aligned}$$

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 $X_1,\cdots,X_m$i.i.d.$\sim N(\mu_1,\sigma^2)$;$Y_1,\cdots,Y_n$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

$$D=\left\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):\left|\frac{\overline{Y}-\overline{X}}{S_w}\sqrt{\frac{mn}{m+n}}\right|>t_{n+m-2}\left(\frac{\alpha}{2}\right)\right\},$$

其中

$$\begin{aligned}&m=156,n=74,\overline{X}=465.13,\overline{Y}=422.16,\quad S_1=54.80,S_2=49.20,\\ &S_w^2=\frac{1}{n+m-2}\left[(m-1)S_1^2+(n-1)S_2^2\right]=2816.6,\quad S_w=53.07,\end{aligned}$$

查表得 $t_{228}(0.005)=2.576$. 由

$$|T|=\left|\frac{\overline{Y}-\overline{X}}{S_w}\sqrt{\frac{mn}{m+n}}\right|=\left|\frac{422.16-465.13}{53.07}\sqrt{\frac{156\times 74}{156+74}}\right|=5.74>2.576,$$

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**

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 $X_1,X_2,\cdots,X_n$ 和服用此安眠药后每晚睡眠时间 $Y_1,Y_2,\cdots,Y_n$, 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,$X_1,X_2,\cdots,X_n$ 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.$Y_1,Y_2,\cdots,Y_n$ 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 $Z_1,\cdots,Z_n$ 为自 $N(\mu,\sigma^2)$ 中抽取的简单随机样本,$\mu$ 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 $Z_1, \cdots, Z_n$ 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

$$D = \{(Z_1, Z_2, \cdots, Z_n) : |T_Z| > t_{n-1}(\alpha/2)\},$$

此处 $\alpha$ 为检验水平,$T_Z = \sqrt{n}\overline{Z}/S_Z$ 为检验统计量, 其中 $\overline{Z}$ 和 $S_Z^2$ 分别为 $Z_1, Z_2, \cdots, Z_n$ 的样本均值和样本方差.

**Example 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

$$u(\text{单位}:\%) : 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00.$$

$$v(\text{单位}:\%) : 0.10, 0.21, 0.52, 0.32, 0.78, 0.59, 0.68, 0.77, 0.89.$$

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 $X_1, X_2, \cdots, X_9$, 和 $Y_1, Y_2, \cdots, Y_9$. 由于这 9 件样品金属含量不同, 所以 $X_1, X_2, \cdots, X_9$ 不能看成来自同一总体.$Y_1, Y_2, \cdots, Y_9$ 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

$$Z_i = Y_i - X_i, \quad i = 1, 2, \cdots, 9.$$

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 $Z_1, \cdots, Z_n$ 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

$$\{(Z_1, \cdots, Z_n) : |T_z| > t_{n-1}(\alpha/2)\}$$

其中 $n = 9$. 由题中数据算得

$$\overline{Z} = \frac{1}{9}\sum_{i=1}^{9} Z_i = 0.06, \quad S_z^2 = \frac{1}{8}\sum_{i=1}^{9}\left(Z_i - \overline{Z}\right)^2 = 0.01505, \quad S_z = 0.12268,$$

查表得 $t_{n-1}(\alpha/2) = t_8(0.005) = 3.3554$. 由于

$$|T_Z| = \left|\frac{\sqrt{n}\overline{Z}}{S_Z}\right| = \frac{3 \times 0.06}{0.12268} = 1.47 < 3.354,$$

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 $\boldsymbol{X}=(X_1,\cdots,X_m)$ 是从正态总体 $N(\mu_1,\sigma_1^2)$ 中抽取的简单随机样本,$\boldsymbol{Y}=(Y_1,\cdots,Y_n)$ 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 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 独立. 讨论下列三类假设检验问题:

(10) $H_0:\dfrac{\sigma_2^2}{\sigma_1^2}=1\leftrightarrow H_1:\dfrac{\sigma_2^2}{\sigma_1^2}\neq 1$,

(11) $H_0':\dfrac{\sigma_2^2}{\sigma_1^2}\leqslant 1\leftrightarrow H_1':\dfrac{\sigma_2^2}{\sigma_1^2}>1$,

(12) $H_0'':\dfrac{\sigma_2^2}{\sigma_1^2}\geqslant 1\leftrightarrow H_1'':\dfrac{\sigma_2^2}{\sigma_1^2}<1$,

检验水平 $\alpha$ 给定.

记 $\overline{X}$ 和 $S_1^2$ 为 $X_1,\cdots,X_m$ 的样本均值和样本方差;$\overline{Y}$ 和 $S_2^2$ 为 $Y_1,\cdots,Y_n$ 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 $\mu_1$ 和 $\mu_2$ 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 $\sigma_1^2$ 和 $\sigma_2^2$ 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

$$\left\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):\frac{S_2^2}{S_1^2}<c_1\text{或}\frac{S_2^2}{S_1^2}>c_2\right\},c_1,c_2\text{待定}.$$

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

$$P_\theta\left(\frac{S_2^2}{S_1^2}<c_1\text{或}\frac{S_2^2}{S_1^2}>c_2\middle|H_0\right)=\alpha.$$

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

$$P_\theta\left(\frac{S_2^2}{S_1^2}<c_1\middle|H_0\right)=\frac{\alpha}{2},\quad P_\theta\left(\frac{S_2^2}{S_1^2}>c_2\middle|H_0\right)=\frac{\alpha}{2}.$$

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 $\alpha$ 的接受域为

$$\overline{D}_{10}=\left\{(\boldsymbol{X},\boldsymbol{Y}):F_{n-1,m-1}(1-\alpha/2)\leqslant\frac{S_2^2}{S_1^2}\leqslant F_{n-1,m-1}(\alpha/2)\right\}.$$

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 $\alpha$ 的否定域如下:

$$D_{11}=\left\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):\frac{S_2^2}{S_1^2}>F_{n-1,m-1}(\alpha)\right\},$$
$$D_{12}=\left\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):\frac{S_2^2}{S_1^2}<F_{n-1,m-1}(1-\alpha)\right\}.$$

这里要注意一点的是: 当 $\alpha$ 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

$$F_{n,m}(1-\alpha)=\frac{1}{F_{m,n}(\alpha)}, \tag{5.2.16}$$

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 $\mu_1$ 和 $\mu_2$ 已知时方差比的检验方法**

简述如下: 当 $\mu_1$ 和 $\mu_2$ 已知时,$\sigma_1^2$ 和 $\sigma_2^2$ 具有良好性质的无偏估计分别是

$$S_{1*}^2=\frac{1}{m}\sum_{i=1}^{m}(X_i-\mu_1)^2, S_{2*}^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\mu_2)^2.$$

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

$$F_*=\frac{S_{2*}^2}{S_{1*}^2}\sim F_{n,m}. \tag{5.2.17}$$

因此, 取检验统计量 $F_*$ 代替 $F=S_2^2/S_1^2$, 完全类似于 $\mu_1$ 和 $\mu_2$ 未知情形的讨论, 可得到检验问题 (10)-(12) 的水平为 $\alpha$ 的否定域, 只要在意在否定域中, 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 |
|---|---|---|---|---|
| $\mu_1$ $\mu_2$ 已知 | $\sigma_2^2=\sigma_1^2$ | $\sigma_2^2\neq\sigma_1^2$ | $F_*=S_{2*}^2/S_{1*}^2$ <br> $F_*\vert_{\sigma_2^2=\sigma_1^2}\sim F_{n,m}$ <br> $S_{1*}^2=\frac{1}{m}\sum_{i=1}^{m}(X_i-\mu_1)^2$ <br> $S_{2*}^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\mu_2)^2$ | $F_*<F_{n,m}(1-\alpha/2)$ 或$F_*>F_{n,m}(\alpha/2)$ |
| | $\sigma_2^2\leqslant\sigma_1^2$ | $\sigma_2^2>\sigma_1^2$ | | $F_*>F_{n,m}(\alpha)$ |
| | $\sigma_2^2\geqslant\sigma_1^2$ | $\sigma_2^2<\sigma_1^2$ | | $F_*<F_{n,m}(1-\alpha)$ |
| $\mu_1$ $\mu_2$ 未知 | $\sigma_2^2=\sigma_1^2$ | $\sigma_2^2\neq\sigma_1^2$ | $F=S_2^2/S_1^2$ <br> $F\vert_{\sigma_2^2=\sigma_1^2}\sim F_{n-1,m-1}$ <br> $S_1^2=\frac{1}{m-1}\sum_{i=1}^{m}(X_i-\mu_1)^2$ <br> $S_2^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\mu_2)^2$ | $F<F_{n-1,m-1}(1-\alpha/2)$ 或$F>F_{n-1,m-1}(\alpha/2)$ |
| | $\sigma_2^2\leqslant\sigma_1^2$ | $\sigma_2^2>\sigma_1^2$ | | $F>F_{n-1,m-1}(\alpha)$ |
| | $\sigma_2^2\geqslant\sigma_1^2$ | $\sigma_2^2<\sigma_1^2$ | | $F<F_{n-1,m-1}(1-\alpha)$ |

**Example 5.2.6.** 测得两批样本大小皆为 6 的电子器材电阻的均值 $\overline{X} = 0.14, \overline{Y} = 0.139$, 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, 问这两批电子器材的电阻是否相同?$(\alpha = 0.05)$

**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) $H_0 : \sigma_1^2 = \sigma_2^2 \leftrightarrow H_1 : \sigma_1^2 \neq \sigma_2^2, \alpha = 0.05$.

由表5.2.4可知, 此检验的接受域是

$$\left\{(X, Y) : F_{m-1,n-1}(1-\alpha/2) \leqslant \frac{S_1^2}{S_2^2} \leqslant F_{m-1,n-1}(\alpha/2)\right\},$$

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) $H_0' : \mu_1 = \mu_2 \leftrightarrow H_1' : \mu_1 \neq \mu_2$, $\alpha = 0.05$.

由表5.2.3可知此检验的否定域为

$$\{(X, Y) : |T_w| > t_{n+m-2}(\alpha/2)\},$$

此处 $m = n = 6$, $\overline{X} = 0.14$, $\overline{Y} = 0.139$, $S_1 = 0.0026$, $S_2 = 0.0024$, 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

$$|T| = \sqrt{\frac{nm}{n+m}}\left|\frac{\overline{Y} - \overline{X}}{S_w}\right| = \sqrt{3} \times \left|\frac{0.14 - 0.139}{0.0025}\right| = 0.6928 < 2.228,$$

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 $X_1,\cdots,X_m$ 是从正态总体 $N(\mu_1,\sigma_1^2)$ 中抽取的简单随机样本, $Y_1,\cdots,Y_n$ 是正态总体 $N(\mu_2,\sigma_2^2)$ 中抽取的简单随机样本, 且样本 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 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) $H_0':\mu_2-\mu_1\leqslant\mu_0 \leftrightarrow H_1':\mu_2-\mu_1>\mu_0$,

(c) $H_0'':\mu_2-\mu_1\geqslant\mu_0 \leftrightarrow H_1'':\mu_2-\mu_1<\mu_0$,

其中 $\mu_0$ 和检验水平 $\alpha$ 给定.

上述检验问题的几个特例情形:①$\sigma_1^2$ 和 $\sigma_2^2$ 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 $\sigma_1^2$ 和 $\sigma_2^2$ 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 $\sigma_1^2$ 和 $\sigma_2^2$ 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 $Y-X\sim N(\mu_2-\mu_1,\sigma_1^2/m+\sigma_2^2/n)$, 将其标准化有

$$\frac{\overline{Y}-\overline{X}-(\mu_2-\mu_1)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\sim N(0,1).$$

由于 $S_1^2$ 和 $S_2^2$ 分别为 $\sigma_1^2$ 和 $\sigma_2^2$ 的合相估计, 用 $S_1^2$ 和 $S_2^2$ 代替 $\sigma_1^2$ 和 $\sigma_2^2$, 由引理2.5.1可知, 当 $m,n\to\infty$ 时有

$$\frac{\overline{Y}-\overline{X}-(\mu_2-\mu_1)}{\sqrt{S_1^2/m+S_2^2/n}}=\frac{\overline{Y}-\overline{X}-(\mu_2-\mu_1)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\cdot\frac{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}{\sqrt{S_1^2/m+S_2^2/n}}\xrightarrow{\mathscr{L}}N(0,1).$$

特别当 $\mu_2-\mu_1=\mu_0$ 且 $m,n\to\infty$ 时, 有

$$U^*=\frac{\overline{Y}-\overline{X}-\mu_0}{\sqrt{S_1^2/m+S_2^2/n}}\xrightarrow{\mathscr{L}}N(0,1).\tag{5.2.18}$$

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 $\alpha$, 而是近似为 $\alpha$.

$$\begin{aligned}D_a&=\{(X,Y):|U^*|>u_{\alpha/2}\};\\ D_b&=\{(X,Y):U^*>u_\alpha\};\\ D_c&=\{(X,Y):U^*<-u_\alpha\}.\end{aligned}\tag{5.2.19}$$

*(2) 当 $\sigma_1^2$ 和 $\sigma_2^2$ 未知, 但是 $m$,$n$ 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

$$\frac{\overline{Y}-\overline{X}-(\mu_2-\mu_1)}{\sqrt{S_1^2/m+S_2^2/n}}$$

近似服从自由度为 $r$ 的 $t$ **分布**, 其中

$$r=S_*^4\bigg/\left[\frac{S_1^4}{m^2(m-1)}+\frac{S_2^4}{n^2(n-1)}\right],\quad S_*^2=\frac{S_1^2}{m}+\frac{S_2^2}{n},$$

当 $\mu_2-\mu_1=\mu_0$ 时, 近似地有

$$T_*=\frac{\overline{Y}-\overline{X}-\mu_0}{\sqrt{S_1^2/m+S_2^2/n}}\sim t_r, \tag{5.2.20}$$

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) 的检验水平近似为 $\alpha$ 的检验的否定域

$$\begin{aligned}
&D_a=\{(\boldsymbol{X},\boldsymbol{Y}):|T_*|>t_r(\alpha/2)\};\\
&D_b=\{(\boldsymbol{X},\boldsymbol{Y}):T_*>t_r(\alpha)\};\\
&D_b=\{(\boldsymbol{X},\boldsymbol{Y}):T_*<-t_r(\alpha)\};
\end{aligned}\tag{5.2.21}$$

**Example 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 $S_1=54.80$ 万/mm$^3$, $S_2=39.20$ 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

$$\{(\boldsymbol{X},\boldsymbol{Y}):|U^*|>u_{\alpha/2}\},$$

其中检验统计量 $U^*=(\overline{Y}-\overline{X})/\sqrt{S_1^2/m+S_2^2/n}$,$m=176$,$n=74$,$\overline{X}=465.13$,$\overline{Y}=422.16$,$S_1=54.80$,$S_2=39.20$, 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.

***Example 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

$$\text{甲矿石}:\overline{X}=\frac{1}{10}\sum_{i=1}^{10}X_i=16.01,S_1^2=\frac{1}{9}\sum_{i=1}^{10}(X_i-\overline{X})^2=10.80,$$

$$\text{乙矿石}:\overline{Y}=\frac{1}{5}\sum_{j=1}^{5}Y_i=18.98,S_2^2=\frac{1}{4}\sum_{i=1}^{5}(Y_i-\overline{Y})^2=0.27.$$

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 $X_1, \cdots, X_m$ i.i.d. $\sim N(\mu_1, \sigma_1^2)$, $Y_1, \cdots, Y_n$ 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

$$H_0 : \mu_1 \geqslant \mu_2 \leftrightarrow H_1 : \mu_1 < \mu_2.$$

检验的否定域

$$D = \left\{ (\boldsymbol{X}, \boldsymbol{Y}) : T_* = \frac{\overline{X} - \overline{Y}}{\sqrt{S_1^2/m + S_2^2/n}} < -t_r(\alpha) \right\}.$$

首先计算 $t$ 分布的自由度

$$r = \frac{\left(S_1^2/m + S_2^2/n\right)^2}{S_1^4/m^2(m-1) + S_2^4/n^2(n-1)} = 9.88 \approx 10,$$

查表得 $t_{10}(0.01) = 2.764$, 由数据得

$$T_* = \frac{\overline{X} - \overline{Y}}{\sqrt{S_1^2/m + S_2^2/n}} = \frac{-2.97}{1.065} = -2.79 < -2.764,$$

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 $X_1, \cdots, X_n$ i.i.d. $\sim b(1, p)$, 显见 $T = \sum\limits_{i=1}^n X_i \sim b(n, p)$, 考虑下列检验问题:

$$H_0 : p = p_0 \leftrightarrow H_1 : p \neq p_0, \tag{5.2.22}$$

其中 $p_0$ 和检验水平 $\alpha$ 给定.

由独立同分布场合的中心极限定理可知 $(T - np)/\sqrt{np(1-p)} \xrightarrow{\mathscr{L}} N(0, 1)$, 当 $n \to \infty$ 时. 故当 $H_0$ 成立, 即 $p = p_0$ 时有

$$U = \frac{T - np_0}{\sqrt{np_0(1 - p_0)}} \xrightarrow{\mathscr{L}} N(0, 1), \quad \text{当} n \to \infty.$$

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 $\alpha$ 的否定域为

$$D_1 = \{(X_1, \cdots, X_n) : |U| > u_{\alpha/2}\}.$$

类似可知 $p$ 的两个单边检验问题及水平近似为 $\alpha$ 的否定域如下:

$$H_0' : p \leqslant p_0 \leftrightarrow H_1' : p > p_0, \quad D_2 = \{(X_1, \cdots, X_n) : U > u_\alpha\}.$$
$$H_0'' : p \geqslant p_0 \leftrightarrow H_1'' : p < p_0, \quad D_3 = \{(X_1, \cdots, X_n) : U < -u_\alpha\}.$$

**3.Poisson 分布参数的大样本检验**

设 $X_1,\cdots,X_n$ 为自 Poisson 总体 $P(\lambda)$ 中抽取的随机样本, 考虑检验问题

$$H_0:\lambda=\lambda_0\leftrightarrow H_1:\lambda\neq\lambda_0, \tag{5.2.23}$$

其中 $\lambda_0$ 和检验水平 $\alpha$ 给定.

由于 $T=\sum\limits_{i=1}^{n}X_i$ 服从参数为 $n\lambda$ 的 Poisson 分布 $P(n\lambda)$. 由中心极限定理可知 $\frac{T-n\lambda}{\sqrt{n\lambda}}\xrightarrow{\mathscr{L}} N(0,1)$, 当 $n\to\infty$. 因此当 $H_0$ 成立, 即 $\lambda=\lambda_0$ 时有

$$U_0=\frac{T-n\lambda_0}{\sqrt{n\lambda_0}}\xrightarrow{\mathscr{L}} N(0,1),\quad 当 n\to\infty.$$

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 $\alpha$ 的否定域为

$$D_1=\{(X_1,\cdots,X_n):|U_0|>u_{\alpha/2}\}.$$

类似可知 $\lambda$ 的两个单边检验问题及水平近似为 $\alpha$ 的否定域如下:

$$H_0':\lambda\leqslant\lambda_0\leftrightarrow H_1':\lambda>\lambda_0,\quad D_2=\{(X_1,\cdots,X_n):U_0>u_\alpha\}.$$
$$H_0'':\lambda\geqslant\lambda_0\leftrightarrow H_1'':\lambda<\lambda_0,\quad D_2=\{(X_1,\cdots,X_n):U_0<-u_\alpha\}.$$

**4. 两样本检验问题**

设 $X_1,\cdots,X_m$ i.i.d. $\sim b(1,p_1)$, $Y_1,\cdots,Y_n$ i.i.d. $\sim b(1,p_2)$, 且样本 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 独立. 求检验问题

$$H_0:p_2-p_1=0\leftrightarrow H_1:p_2-p_1\neq 0, \tag{5.2.24}$$

检验水平 $\alpha$ 给定.

记 $\overline{X}$ 和 $\overline{Y}$ 分别为两组样本的均值. 由中心极限定理可知

$$\frac{\overline{Y}-\overline{X}-(p_2-p_1)}{\sqrt{p_1(1-p_1)/m+p_2(1-p_2)/n}}\xrightarrow{\mathscr{L}} N(0,1),\quad 当 n,m\to\infty.$$

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 $m,n\to\infty$ 时, 有

$$\widetilde{U}=\frac{\overline{Y}-\overline{X}}{\sqrt{\hat{p}(1-\hat{p})}}\sqrt{\frac{mn}{m+n}}=\frac{\overline{Y}-\overline{X}}{\sqrt{p(1-p)}}\sqrt{\frac{mn}{m+n}}\cdot\frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}(1-\hat{p})}}\xrightarrow{\mathscr{L}} N(0,1).$$

因此取 $\widetilde{U}$ 为检验统计量, 当 $m,n$ 都较大时, $\widetilde{U}$ 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 $\alpha$ 的否定域为

$$D_1=\left\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):|\widetilde{U}|>u_{\alpha/2}\right\}.$$

类似可知两个单边检验问题及水平近似为 $\alpha$ 的否定域如下:

$$H_0':p_2\leqslant p_1\leftrightarrow H_1':p_2>p_1,\quad D_2=\left\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):\widetilde{U}>u_\alpha\right\}$$
$$H_0':p_2\geqslant p_1\leftrightarrow H_1':p_2<p_1,\quad D_3=\left\{(X_1,\cdots,X_m;Y_1,\cdots,Y_n):\widetilde{U}<-u_\alpha\right\}.$$

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

### 5.3.1 似然比检验的定义

设有分布族 $\{f(x,\theta),\theta\in\Theta\}$,$\Theta$ 为参数空间. 令 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为自上述分布族中抽取的简单随机样本,$f(\boldsymbol{x},\theta)$ 为样本的概率函数. 要考虑检验问题(5.1.1). 在有了样本 $\boldsymbol{x}$ 后将 $f(\boldsymbol{x},\theta)$ 视为 $\theta$ 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 $f(\boldsymbol{x},\theta_1)<f(\boldsymbol{x},\theta_2)$, 则认为真参数为 $\theta_2$ 的"似然性"较其为 $\theta_1$ 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:

$$L_{\Theta_0}(x)=\sup_{\theta\in\Theta_0}f(x,\theta),$$
$$L_{\Theta_1}(x)=\sup_{\theta\in\Theta_1}f(x,\theta).$$

考虑其比值 $L_{\Theta_1}(\boldsymbol{x})/L_{\Theta_0}(\boldsymbol{x})$, 若此比值较大, 则说明真参数在 $\Theta_1$ 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 $\lambda(\boldsymbol{x})=L_\Theta(\boldsymbol{x})/L_{\Theta_0}(\boldsymbol{x})$, 其中 $L_\Theta(\boldsymbol{x})=\sup\limits_{\theta\in\Theta}f(\boldsymbol{x},\theta)$. 由于 $\lambda(\boldsymbol{x})$ 与 $L_{\Theta_1}(\boldsymbol{x})/L_{\Theta_0}(\boldsymbol{x})$ 同增或同减, 可用 $\lambda(\boldsymbol{x})$ 代替比值 $L_{\Theta_1}(\boldsymbol{x})/L_{\Theta_0}(\boldsymbol{x})$, 这样做的好处是 $L_\Theta(\boldsymbol{x})=\sup_{\theta\in\Theta}f(\boldsymbol{x},\theta)$ 的计算比 $L_{\Theta_1}(\boldsymbol{x})$ 要容易. 因此得到如下定义

**Definition 5.3.1.** 设样本 $\boldsymbol{X}$ 有概率函数 $f(\boldsymbol{x},\theta)$,$\theta\in\Theta$, 而 $\Theta_0$ 为参数空间 $\Theta$ 的真子集, 考虑检验问题(5.1.1), 则统计量

$$\lambda(x)=\frac{\sup\limits_{\theta\in\Theta}f(\boldsymbol{x},\theta)}{\sup\limits_{\theta\in\Theta_0}f(\boldsymbol{x},\theta)}\tag{5.3.1}$$

称为关于该检验问题的似然比. 而由下述定义的检验函数

$$\varphi(\boldsymbol{x})=\begin{cases}1, & \lambda(\boldsymbol{x})>c,\\ r, & \lambda(\boldsymbol{x})=c,\\ 0, & \lambda(\boldsymbol{x})<c,\end{cases}\tag{5.3.2}$$

其中 $c$,$r(0\leqslant r\leqslant 1)$ 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

$$\varphi(\boldsymbol{x})=\begin{cases}1, & \lambda(\boldsymbol{x})>c,\\ 0, & \lambda(\boldsymbol{x})\leqslant c.\end{cases}$$

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 $\alpha$.

根据上面所说, 找似然比检验有以下步骤:

(1) 求似然函数 $f(\boldsymbol{x},\theta)$, 并明确参数空间 $\Theta$ 和 $\Theta_0$ 是什么.

(2) 算出 $L_\Theta(\boldsymbol{x})=\sup\limits_{\theta\in\Theta} f(\boldsymbol{x},\theta)$ 和 $L_{\Theta_0}(\boldsymbol{x})=\sup\limits_{\theta\in\Theta_0} f(\boldsymbol{x},\theta)$.

(3) 求出 $\lambda(\boldsymbol{x})$ 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 $\alpha$.

其中最关键的是第 3 步. 一般 $\lambda(\boldsymbol{x})$ 的表达式复杂, 求其分布不易. 但若 $\lambda(\boldsymbol{x})=g(T(\boldsymbol{x}))$ 为 $T(\boldsymbol{x})$ 的单调上升 (或下降) 函数, 则检验式(5.3.2)显然等价于

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})>c',\\ r, & T(\boldsymbol{x})=c',\\ 0, & T(\boldsymbol{x})<c'.\end{cases}$$

因此代替求 $\lambda(\boldsymbol{X})$ 的分布, 只要求出 $T(\boldsymbol{X})$ 的分布即可 (若 $\lambda(\boldsymbol{x})$ 为 $T(\boldsymbol{x})$ 的单调下降函数, 则将 $\varphi(\boldsymbol{x})$ 中的不等式反向).

如果 $\lambda(\boldsymbol{X})$ 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

**Example 5.3.1.** 设 $X=(X_1,\cdots,X_n)$ 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 $\alpha$ 的似然比检验:

$$H_0:\mu=\mu_0\leftrightarrow H_1:\mu\neq\mu_0. \tag{5.3.3}$$

**解:** 记 $\boldsymbol{\theta}=(\mu,\sigma^2)$, 则 $\boldsymbol{\theta}$ 的似然函数为

$$f(\boldsymbol{x},\boldsymbol{\theta})=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2\right\}, \tag{5.3.4}$$

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 $\Theta$ 的子集为

$$\Theta_0=\left\{\boldsymbol{\theta}=(\mu,\sigma^2):\mu=\mu_0,\sigma^2>0\right\}.$$

在 $\Theta$ 上,$\mu$ 和 $\sigma^2$ 的极大似然估计 (MLE) 为

$$\hat{\mu}=\overline{X},\quad \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2;$$

在 $\Theta_0$ 上,$\sigma^2$ 的 MLE 为

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2.$$

故有

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}f(\boldsymbol{x},\boldsymbol{\theta})&=f(\boldsymbol{x},\hat{\mu},\hat{\sigma}^2)=\left(\frac{2\pi e}{n}\right)^{-n/2}\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^{-n/2},\\ \sup_{\boldsymbol{\theta}\in\Theta_0}f(\boldsymbol{x},\boldsymbol{\theta})&=f(\boldsymbol{x},\mu_0,\hat{\sigma}^2)=\left(\frac{2\pi e}{n}\right)^{-n/2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2\right)^{-n/2}.\end{aligned}\tag{5.3.5}$$

从而有

$$\begin{aligned}\lambda(\boldsymbol{x})&=\left[\sum_{i=1}^{n}(x_i-\overline{x})^2\Big/\sum_{i=1}^{n}(x_i-\mu_0)^2\right]^{-n/2}\\&=\left[1+n(\overline{x}-\mu_0)^2\Big/\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^{n/2}\\&=\left(1+\frac{1}{n-1}T^2\right)^{n/2},\end{aligned}$$

其中 $T=\sqrt{n}(\overline{X}-\mu_0)/S$, 而 $S^2=\sum\limits_{i=1}^{n}(X_i-\overline{X})^2/(n-1)$. 由于 $\lambda(\boldsymbol{x})$ 为 $|T|$ 的严格增函数, 故检验的否定域 $D=\{\boldsymbol{X}=(X_1,\cdots,X_n):\lambda(\boldsymbol{X})>c'\}=\{\boldsymbol{X}:|T|>c\}$. 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

$$\varphi(\boldsymbol{x})=\begin{cases}1, & |T|>t_{n-1}(\alpha/2),\\ 0, & |T|\leqslant t_{n-1}(\alpha/2)\end{cases}$$

是检验问题(5.3.3)的一个水平为 $\alpha$ 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.

**Example 5.3.2.** 问题与例5.3.1相同, 求下列检验:

$$H_0:\mu\leqslant\mu_0\leftrightarrow H_1:\mu>\mu_0\tag{5.3.6}$$

的水平为 $\alpha$ 的似然比检验.

**解:** 此时似然函数 $f(\boldsymbol{x},\theta)$ 和 $\Theta$ 与例5.3.1中相同, 但

$$\Theta_0=\left\{\theta=(\mu,\sigma^2):\mu\leqslant\mu_0,\sigma^2>0\right\}.$$

因此,$L_{\Theta}(\boldsymbol{x})=\sup\limits_{\theta\in\Theta} f(\boldsymbol{x},\mu,\sigma^2)$ 与例5.3.1中完全相同, 但要注意到

$$\begin{aligned} L_{\Theta_0}(\boldsymbol{x}) &= \sup_{\theta\in\Theta_0}(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\ &= \sup_{\theta\in\Theta_0}(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2-\frac{n(\overline{x}-\mu)^2}{2\sigma^2}\right\}. \end{aligned}$$

记 $g(\mu)=\exp\{-n(\overline{x}-\mu)^2/(2\sigma^2)\}$. 当 $\sigma^2$ 固定,$\mu\leqslant\overline{x}$ 时,$g'(\mu)\geqslant 0$, 故 $g(\mu)$ 关于 $\mu$ 单调增; 当 $\mu\geqslant\overline{x}$ 时,$g'(\mu)\leqslant 0$, 故 $g(\mu)$ 关于 $\mu$ 单调降. 因此,

(1) 当 $\overline{x}>\mu_0$ 时, 若 $H_0$ 成立,$g(\mu)$ 在 $\mu=\mu_0$ 处达到最大, 故有

$$\min_{\mu\leqslant\mu_0}\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}(x_i-\mu_0)^2.$$

(2) 当 $\overline{x}\leqslant\mu_0$ 时, 若 $H_0$ 成立,$g(\mu)$ 在 $\mu=\overline{x}$ 处达到最大, 故有

$$\min_{\mu\leqslant\mu_0}\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}(x_i-\overline{x})^2,$$

因此有

$$L_{\Theta_0}(\boldsymbol{x})=\begin{cases} L_{\Theta}(\boldsymbol{x}), & \overline{x}\leqslant\mu_0, \\ \left(\frac{2\pi e}{n}\right)^{-n/2}\left(\sum\limits_{i=1}^{n}(x_i-\mu_0)^2\right)^{-n/2}, & \overline{x}>\mu_0. \end{cases}$$

故

$$\lambda(\boldsymbol{x})=\begin{cases} 1, & \overline{x}\leqslant\mu_0, \\ \left[\sum\limits_{i=1}^{n}(x_i-\mu_0)^2\Big/\sum\limits_{i=1}^{n}(x_i-\overline{x})^2\right]^{\frac{n}{2}}=\left(1+\frac{1}{n-1}T^2\right)^{\frac{n}{2}}, & \overline{x}>\mu_0. \end{cases}$$

此处 $T=\sqrt{n}(\overline{X}-\mu_0)/S$, 而 $S^2=\sum\limits_{i=1}^{n}(X_i-\overline{X})^2/(n-1)$. 由于 $\lambda(\boldsymbol{x})$ 为 $T$ 的严格增函数, 因此似然比检验的否定域为

$$D=\{\boldsymbol{X}=(X_1,\cdots,X_n):\lambda(\boldsymbol{X})>c'\}=\{\boldsymbol{X}:T>c\}.$$

由于检验水平 $\alpha$ 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 $\mu$ 的单调增函数, 故有

$$\beta_\varphi(\mu)\leqslant\beta_\varphi(\mu_0)=\alpha,\quad \mu\leqslant\mu_0.$$

因此

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})>t_{n-1}(\alpha),\\ 0, & T(\boldsymbol{x})\leqslant t_{n-1}(\alpha)\end{cases}$$

为检验问题(5.3.6)的水平为 $\alpha$ 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

$$H_0:\mu\geqslant\mu_0\leftrightarrow H_1:\mu<\mu_0 \tag{5.3.7}$$

的水平为 $\alpha$ 的似然比检验, 这留给读者作为练习.

**Example 5.3.3.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\alpha$ 的似然比检验:

$$H_0:\sigma^2=\sigma_0^2\leftrightarrow H_1:\sigma^2\neq\sigma_0^2, \tag{5.3.8}$$

其中 $\sigma_0^2$ 和 $\alpha$ 给定.

**解:** 此时似然函数仍为式(5.3.4), 参数空间 $\Theta$ 如例5.3.1, 而

$$\Theta_0=\{\boldsymbol{\theta}=(\mu,\sigma^2):-\infty<\mu<+\infty,\sigma^2=\sigma_0^2\}, \tag{5.3.9}$$

$L_{\Theta_0}(\boldsymbol{x})$ 也与例 5.3.1 相同, 但

$$\begin{aligned}L_{\Theta_0}(\boldsymbol{x})=&\sup_{\mu}\left\{(2\pi\sigma_0^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}\right\}\\ =&(2\pi\sigma_0^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right\},\end{aligned}$$

因此有

$$\lambda(\boldsymbol{x})=\left(\frac{e}{n}\right)^{-n/2}\left[\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^{-n/2}\exp\left\{\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right\}.$$

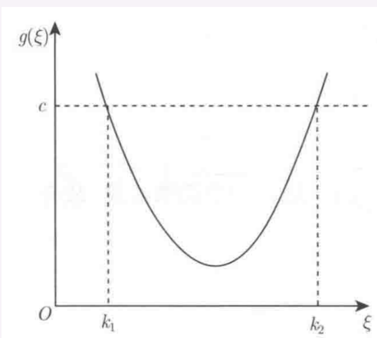



**图 5.3.1**

令 $\xi=\frac{1}{\sigma_0^2}\sum\limits_{i=1}^{n}(x_i-\overline{x})^2$,$g(\xi)=\xi^{-\frac{n}{2}}e^{\xi/2}$, 则在 $\xi>0$ 时 $g(\xi)$ 关于 $\xi$ 先降后升, 且当 $\xi\to0$ 和 $\xi\to\infty$ 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 $\overline{D}$=$\{\boldsymbol{X}:g(\xi)\leqslant c\}$=$\{\boldsymbol{X}:k_1\leqslant\sum\limits_{i=1}^{n}(X_i-\overline{X})^2/\sigma_0^2\leqslant k_2\}$.

由于在 $H_0$ 成立时,$\sum\limits_{i=1}^{n}(X_i-\overline{X})^2/\sigma_0^2\sim\chi_{n-1}^2$, 因而 $k_1$ 和 $k_2$ 为下列方程组的解:

$$\begin{cases}g(k_1)=g(k_2),\\ P(k_1\leqslant\xi\leqslant k_2\mid H_0)=1-\alpha\end{cases}\Leftrightarrow\begin{cases}k_1^{n/2}e^{-k_1/2}=k_2^{n/2}e^{-k_2/2},\\ P(\xi<k_1|H_0)+P(\xi>k_2|H_0)=\alpha.\end{cases} \tag{5.3.10}$$

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

$$k_1 = \chi^2_{n-1}(1-\alpha/2), \quad k_2 = \chi^2_{n-1}(\alpha/2).$$

因此检验问题(5.3.8)的水平为 $\alpha$ 的似然比检验是

$$\varphi(\boldsymbol{x}) = \begin{cases} 0, & \chi^2_{n-1}(1-\alpha/2) < \frac{1}{\sigma_0^2}\sum\limits_{i=1}^{n}(x_i-\overline{x})^2 < \chi^2_{n-1}(\alpha/2), \\ 1, & \text{其他}. \end{cases}$$

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 $H_0 : \sigma^2 \leqslant \sigma_0^2 \leftrightarrow H_1 : \sigma^2 > \sigma_0^2$ 的水平为 $\alpha$ 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.

**Example 5.3.4.** 设 $X = (X_1, \cdots, X_n)$ 为自均匀分布总体 $U(0,\theta)$ 中抽取的随机样本, 求

$$H_0 : \theta \leqslant \theta_0 \leftrightarrow H_1 : \theta > \theta_0 \tag{5.3.11}$$

水平为 $\alpha$ 的似然比检验, 其中 $\alpha$ 和 $\theta_0$ 给定.

**解:** 此时似然函数为

$$f(\boldsymbol{x},\theta) = \begin{cases} \theta^{-n}, & 0 < x_{(n)} < \theta, \\ 0, & \text{其他}. \end{cases}$$

参数空间 $\Theta = (0,\infty)$,$\Theta_0 = (0,\theta_0]$. 由于 $X_{(n)} = \max\{X_1, \cdots, X_n\}$ 为 $\theta$ 的 MLE, 故有

$$L_\Theta(\boldsymbol{x}) = \sup_{\theta\in\Theta} f(\boldsymbol{x},\theta) = (x_{(n)})^{-n}$$

和

$$L_{\Theta_0}(\boldsymbol{x}) = \sup_{\theta\in\Theta_0} f(\boldsymbol{x},\theta) = \begin{cases} L_\Theta(\boldsymbol{x}), & 0 < x_{(n)} \leqslant \theta_0, \\ 0, & x_{(n)} > \theta_0. \end{cases}$$

因此有

$$\lambda(\boldsymbol{x}) = \begin{cases} 1, & 0 < x_{(n)} \leqslant \theta_0, \\ \infty, & x_{(n)} > \theta_0. \end{cases}$$

由于 $\lambda(\boldsymbol{x})$ 为 $T(\boldsymbol{x}) = x_{(n)}$ 的非降函数, 故检验的否定域为

$$D = \boldsymbol{X} = (X_1, \cdots, X_n) : X_{(n)} > c.$$

**Note 5.3.1.** 为使检验水平等于 $\alpha$, 将集合 $G=\{x=(x_1,\cdots,x_n):\lambda(x)=1\}$ 分为两部分 $G_1=\{\boldsymbol{x}:c<x_{(n)}\leqslant\theta_0\}, G_2=G-G_1$, 则 $G_1\cup\{\boldsymbol{x}:\lambda(\boldsymbol{x})=\infty\}=\{\boldsymbol{x}:x_{(n)}>\theta_0\}$ 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

$$\alpha=P(X_{(n)}>c|\theta=\theta_0)=\int_c^{\theta_0}\frac{nt^{n-1}}{\theta_0^n}\mathrm{d}t=1-\left(\frac{c}{\theta_0}\right)^n,$$

解出 $c=\theta_0\sqrt[n]{1-\alpha}$, 故否定域为

$$D=\{\boldsymbol{X}:X_{(n)}>\theta_0\sqrt[n]{1-\alpha}\}.$$

检验的功效函数为

$$\begin{aligned}\beta_\varphi(\theta)=&P_\theta(X_{(n)}>\theta_0\sqrt[n]{1-\alpha})=\int_{\theta_0\sqrt[n]{1-\alpha}}^{\theta}\frac{nt^{n-1}}{\theta^n}\mathrm{d}t\\=&\frac{1}{\theta^n}\left[\theta^n-\theta_0^n(1-\alpha)\right]=1-(1-\alpha)\left(\frac{\theta_0}{\theta}\right)^n.\end{aligned}$$

它是 $\theta$ 的单调增函数, 故有

$$\beta_\varphi(\theta)\leqslant\beta_\varphi(\theta_0),\quad \theta\leqslant\theta_0,$$

因此以 $D$ 为否定域的检验水平为 $\alpha$. 因此

$$\varphi(\boldsymbol{x})=\begin{cases}1, & X_{(n)}>\theta_0\sqrt[n]{1-\alpha},\\0, & X_{(n)}\leqslant\theta_0\sqrt[n]{1-\alpha}\end{cases}$$

为检验问题(5.3.11)的水平为 $\alpha$ 的似然比检验.

**Example 5.3.5.** 设样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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

$$H_0:\lambda=\lambda_0\leftrightarrow H_1:\lambda\neq\lambda_0 \tag{5.3.12}$$

的检验水平为 $\alpha$ 的似然比检验, 此处 $\lambda_0$ 和 $\alpha$ 给定.

**解:** 参数 $\lambda$ 的似然函数为

$$L(\lambda,x)=\begin{cases}\lambda^{-n}\exp\{-n\overline{x}/\lambda\}, & \text{当}\, x_1,\cdots,x_n>0,\\0, & \text{其他},\end{cases}$$

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 $\lambda$

的极大似然估计为 $\hat{\lambda} = \overline{X}$, 故在 $\Theta$ 和 $\Theta_0$ 上似然函数的最大值分别为

$$L_{\Theta}(\boldsymbol{x}) = \sup_{\lambda \in \Theta} L(\lambda, x) = e^{-n/\overline{x}^n}, L_{\Theta_0}(\boldsymbol{x}) = L(\lambda_0, x) = \lambda_0^{-n} \exp\{-n\overline{x}/\lambda_0\}.$$

记 $T = \sum_{i=1}^{n} x_i = n\overline{x}$, 则似然比为

$$\lambda(T) = \frac{L_{\Theta}(\boldsymbol{x})}{L_{\Theta_0}(\boldsymbol{x})} = \frac{n^n \lambda_0^n}{e^n T^n} \exp\{T/\lambda_0\} = c \cdot g(T).$$

此处 $c = (n\lambda_0/e)^n$,$g(T) = T^{-n}e^{T/\lambda_0}$. 显见当 $T \to \infty$ 和 $T \to 0$ 时 $g(T) \to \infty$,$g(t)$ 的形状与图5.3.1类似. 因此有

$$\varphi(\boldsymbol{x}) = \begin{cases} 1, & \text{当}\lambda(T) > c \\ 0, & \text{当}\lambda(T) \leqslant c \end{cases} = \begin{cases} 1, & \text{当}T < k_1 \text{ 或}T > k_2, \\ 0, & \text{其他}. \end{cases}$$

由推论2.4.5可知, 当 $H_0$ 成立时, $2T/\lambda_0 \sim \chi^2_{2n}$. 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

$$k_1 = \frac{\lambda_0}{2}\chi^2_{2n}\left(1 - \frac{\alpha}{2}\right), \quad k_2 = \frac{\lambda_0}{2}\chi^2_{2n}\left(\frac{\alpha}{2}\right).$$

因此检验问题(5.3.12)的检验水平为 $\alpha$ 的似然比检验为

$$\varphi(x) = \begin{cases} 1, & \text{当}\lambda(T) < \frac{\lambda_0}{2}\chi^2_{2n}\left(1 - \frac{\alpha}{2}\right) \text{或}\lambda > \frac{\lambda_0}{2}\chi^2_{2n}\left(\frac{\alpha}{2}\right), \\ 0, & \text{其他}. \end{cases}$$

### *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 $\lambda(\boldsymbol{X})$ 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 $X_1, \cdots, X_n$ 是简单随机样本, 则当 $n \to \infty$ 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(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 $\Theta$ 的维数要高于零假设成立时的 $\Theta_0$ 的维数, 如样本 $X_1, \cdots, X_n$i.i.d.$\sim N(\mu, \sigma^2)$,$H_0: \mu = \mu_0 \to H_1: \mu \neq \mu_0$, 则 $\Theta = \{\theta = (\mu, \sigma^2): -\infty < \mu < +\infty, \sigma^2 > 0\}$ 是 $\mathbb{R}^2$ 中的上半平面,$\Theta$ 的维数是 2; 而 $\Theta_0 = \{\theta = (\mu, \sigma^2): \mu = \mu_0, \sigma^2 > 0\}$, 它是 $\Theta$ 中的一条直线, 其维数为 1. 因此此例中 $\Theta$ 维数高于 $\Theta_0$ 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

**Theorem 5.3.1.** 设 $\Theta$ 的维数为 $k$,$\Theta_0$ 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 $n\to\infty$ 时有

$$2\log\lambda(\boldsymbol{X})\xrightarrow{\mathscr{L}}\chi_t^2.$$

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 $\Theta_0$ 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 $\Theta_0$ 中何处,$2\log\lambda(\boldsymbol{X})$ 的极限分布总是自由度为 $t$ 的 $\chi^2$ 分布.

**Example 5.3.6.** 设样本 $\boldsymbol{X}_i=(X_{i1},\cdots,X_{in_i})$ 为从正态总体 $N(\mu_i,\sigma_i^2)$,$i=1,\cdots,m$ 中抽取的简单样本, 且全部样本独立. 要检验假设

$$H_0:\sigma_1^2=\cdots=\sigma_m^2\leftrightarrow H_1:\sigma_1^2,\cdots,\sigma_m^2\text{不完全相同}.$$

**解:** 记 $\boldsymbol{\theta}=(\mu_1,\cdots,\mu_m;\sigma_1^2,\cdots,\sigma_m^2)$, 易见 $\boldsymbol{\theta}$ 的似然函数为

$$\begin{aligned}L(\boldsymbol{\theta};\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m)&=\prod_{i=1}^m\left[(2\pi\sigma_i^2)^{-\frac{n_i}{2}}\exp\left\{-\frac{1}{2\sigma_i^2}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2\right\}\right]\\&=\prod_{i=1}^m\left[(2\pi\sigma_i^2)^{-\frac{n_i}{2}}\exp\left\{-\frac{1}{2\sigma_i^2}\sum_{j=1}^{n_i}\left[(x_{ij}-\overline{x}_i)^2+(\overline{x}_i-\mu)^2\right]\right\}\right]\end{aligned}$$

参数空间为 $\Theta=\{\boldsymbol{\theta}=(\mu_1,\cdots,\mu_m;\sigma_1^2,\cdots,\sigma_m^2):\mu_i\in\mathbb{R}^1,\sigma_i^2>0\}$, 其维数为 $k=2m$. 零假设对应的 $\Theta$ 的子空间为

$$\Theta_0=\left\{\boldsymbol{\theta}=(\mu_1,\cdots,\mu_m;\sigma_1^2,\cdots,\sigma_m^2):\mu_i\in\mathbb{R}^1,\sigma_1^2=\cdots=\sigma_m^2>0\right\},$$

其维数为 $s=m+1$. 记

$$\overline{X}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}X_{ij},S_i^2=\frac{1}{n_i}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2,S^2=\frac{1}{n}\sum_{i=1}^m n_iS_i^2.\tag{5.3.13}$$

此处 $n=\sum\limits_{i=1}^m n_i$. 易见,$\mu_i$ 和 $\sigma_i^2$ 的 MLE 分别为 $\hat{\mu}_i=\overline{X}_i$,$\hat{\sigma}_i^2=S_i^2$,$i=1,\cdots,m$. 故

$$L_\Theta(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m)=\sup_{\theta\in\Theta}L(\boldsymbol{\theta};\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m)=(2\pi e)^{-\frac{n}{2}}\prod_{i=1}^m S_i^{-n_i}.$$

当 $\sigma_1^2=\cdots=\sigma_m^2=\sigma^2$ 时,$\sigma^2$ 的 MLE 为 $\hat{\sigma}^2=S^2$, 故

$$L_{\Theta_0}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m)=\sup_{\theta\in\Theta_0}L(\boldsymbol{\theta};\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m)=(2\pi e)^{-\frac{n}{2}}S^{-n}.$$

因此不难算出

$$\lambda(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m)=\frac{L_\Theta(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m)}{L_{\Theta_0}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m)}=S^m\Big/\prod_{i=1}^m S_i^{n_i}.$$

令

$$Y_n \equiv 2\log\lambda(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m) = n\log S^2 - \sum_{i=1}^{n} n_i \log s_i^2.$$

由定理5.3.1可知, 当 $H_0$ 成立且当 $\min\{n_1,n_2,\cdots,n_m\}\to\infty$ 时, 有

$$Y_n \xrightarrow{\mathscr{L}} \chi_t^2 = \chi_{m-1}^2,$$

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 $\alpha$ 的否定域

$$D=\{(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_m): Y_n > \chi_{m-1}^2(\alpha)\}.$$

## *5.4 一致最优检验和无偏检验

### 5.4.1 引言与定义

设有分布族 $\{f(x,\theta),\theta\in\Theta\}$, 其中 $\theta$ 为参数空间. 样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从上述分布族抽取的简单样本, 如 5.1 节所述, 参数 $\theta$ 的假设检验问题可以表示成如下的一般形式:

$$H_0:\theta\in\Theta_0 \leftrightarrow H_1:\theta\in\Theta_1, \tag{5.4.1}$$

其中 $\Theta_0$ 为参数空间 $\Theta$ 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.

**Definition 5.4.1.** 设有检验问题(5.4.1), 令 $0<\alpha<1$, 记 $\Phi_\alpha$ 为式(5.4.1)的一切水平为 $\alpha$ 的检验的集合. 若 $\varphi\in\Phi_\alpha$, 且对任何检验 $\varphi_1\in\Phi_\alpha$, 有

$$\beta_\varphi(\theta)\geqslant\beta_{\varphi_1}(\theta),\quad \theta\in\Theta_1, \tag{5.4.2}$$

则称 $\varphi$ 为式(5.4.1)的一个水平为 $\alpha$ 的**一致最优检验** (uniformly most powerful test, UMPT). 当 $\varphi$ 为水平 $\alpha$ 的 UMPT 时, 它在限制第一类错误概率不超过 $\alpha$ 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 $\Theta_1$ 包含不止一个点, 当在其中取两个不同点 $\theta_1$ 和 $\theta_2$ 时, 为使 $\beta_\varphi(\theta_1)$ 达到最大的那种检验 $\varphi$, 不见得同时也能使 $\beta_\varphi(\theta_2)$ 达到最大. 在 $\Theta_0$ 和 $\Theta_1$ 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

**Theorem 5.4.1** (NP 基本引理)**.** 设样本 $\boldsymbol{X}$ 的分布有概率函数 $f(\boldsymbol{x},\theta)$, 参数 $\theta$ 只有两个可能的值 $\theta_0$ 和 $\theta_1$, 考虑下列检验问题:

$$H_0:\theta=\theta_0 \leftrightarrow H_1:\theta=\theta_1, \tag{5.4.3}$$

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 $\varphi(x)$ 及非负常数 $c$ 和 $0\leqslant r\leqslant 1$, 满足条件

(i)

$$\varphi(\boldsymbol{x})=\begin{cases}1, & f(\boldsymbol{x},\theta_1)/f(\boldsymbol{x},\theta_0)>c,\\ r, & f(\boldsymbol{x},\theta_1)/f(\boldsymbol{x},\theta_0)=c,\\ 0, & f(\boldsymbol{x},\theta_1)/f(\boldsymbol{x},\theta_0)<c.\end{cases} \tag{5.4.4}$$

(ii)

$$E_{\theta_0}[\varphi(\boldsymbol{X})]=\alpha. \tag{5.4.5}$$

(2) 一致最优性. 任何满足式(5.4.4)和式(5.4.5)的检验 $\varphi(\boldsymbol{x})$ 是检验问题(5.4.3)的 UMPT.

**Note 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

$$\varphi(x)=\begin{cases}1, & f(\boldsymbol{x},\theta_1)/f(\boldsymbol{x},\theta_0)>c,\\ 0, & f(\boldsymbol{x},\theta_1)/f(\boldsymbol{x},\theta_0)\leqslant c,\end{cases}$$

其中 $c$ 由 $E_{\theta_0}[\varphi(\boldsymbol{X})]=P\left(f(\boldsymbol{X},\theta_1)/f(\boldsymbol{X},\theta_0)>c\mid H_0\right)=\alpha$ 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 $\boldsymbol{X},\theta_1$ 和 $\theta_0$ 的"似然度"分别为 $f(\boldsymbol{x},\theta_1)$ 和 $f(\boldsymbol{x},\theta_0)$. 比值 $f(\boldsymbol{x},\theta_1)/f(\boldsymbol{x},\theta_0)$ 越大, 就反映在得到样本 $\boldsymbol{X}$ 时,$\theta$ 越像 $\theta_1$ 而非 $\theta_0$, 这样的样本 $\boldsymbol{X}$ 就越倾向于否定"$H_0:\theta=\theta_0$"的假设.

证明略. 不过其中有一重要关系式[1]:

$$\begin{aligned}&\int_{\mathscr{X}}\varphi(\boldsymbol{x})f(\boldsymbol{x},\theta_1)\mathrm{d}\boldsymbol{x}-\int_{\mathscr{X}}\varphi_1(\boldsymbol{x})f(\boldsymbol{x},\theta_1)\mathrm{d}\boldsymbol{x}\\ \geqslant& c\left[\int_{\mathscr{X}}\varphi(\boldsymbol{x})f(\boldsymbol{x},\theta_0)\mathrm{d}\boldsymbol{x}-\int_{\mathscr{X}}\varphi_1(\boldsymbol{x})f(\boldsymbol{x},\theta_0)\mathrm{d}\boldsymbol{x}\right].\end{aligned} \tag{5.4.6}$$

[1] 为了防止编号错乱进行的占坑行为, 不用在意

**Example 5.4.1.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为自正态总体 $N(\mu,1)$ 中抽取的随机样本, 其中 $\mu$ 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 $\alpha$ 的 UMPT, 其中 $\mu_1$ 和 $\alpha$ 给定.

**解:** 由 NP 引理, 先求 $f_0(x)$ 和 $f_1(x)$ 的表达式

$$f_0(\boldsymbol{x})=(2\pi)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^n x_i^2\right\},$$
$$f_1(\boldsymbol{x})=(2\pi)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i-\mu_1)^2\right\}.$$

似然比可表示为

$$\lambda(\boldsymbol{x})=\frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})}=\exp\left\{-\frac{1}{2}n\mu_1^2+n\mu_1\overline{x}\right\}.$$

显然当 $\mu_1>0$ 时,$\lambda(\boldsymbol{x})$ 为 $\overline{x}$ 的严格增函数, 故 UMPT 的否定域为

$$D=\{\boldsymbol{X}:\lambda(\boldsymbol{X})>c'\}=\{\boldsymbol{X}:\sqrt{n}\overline{X}>c\},$$

当 $H_0$ 成立时,$U=\sqrt{n}\overline{X}\sim N(0,1)$, 故由 NP 引理可知

$$E_0[\varphi(\boldsymbol{X})]=P(\sqrt{n}\overline{X}>c|H_0)=\alpha,$$

显然 $c=u_\alpha$. 因此检验水平为 $\alpha$ 的 UMPT 的检验函数为

$$\varphi(\boldsymbol{x})=\begin{cases}1, & \overline{x}>u_\alpha/\sqrt{n},\\ 0, & \overline{x}\leqslant u_\alpha/\sqrt{n}.\end{cases}$$

可见 $\varphi(\boldsymbol{x})$ 与 $\mu_1$ 无关, 因此上述检验函数 $\varphi(\boldsymbol{x})$ 也是检验问题

$$H_0:\mu=0\leftrightarrow H_1':\mu>0$$

的水平为 $\alpha$ 的 UMPT.

**Note 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 $\alpha$ 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 $\alpha$ 的 UMPT, 具体的推导留给读者作为练习.

**Example 5.4.2.** 设 $X=(X_1,\cdots,X_n)$ 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

$$H_0: p=p_0 \leftrightarrow H_1: p=p_1 \ (p_1>p_0)$$

的水平为 $\alpha$ 的 UMPT, 其中 $p_0, p_1$ 和 $\alpha$ 给定.

**解:** 由 NP 引理, 先求 $f_0$ 和 $f_1$ 的表达式

$$\begin{aligned} f_0(\boldsymbol{x}) =& p(\boldsymbol{x},p_0)=p_0^{\sum\limits_{i=1}^n x_i}(1-p_0)^{n-\sum\limits_{i=1}^n x_i}, \\ f_1(\boldsymbol{x}) =& p(\boldsymbol{x},p_1)=p_1^{\sum\limits_{i=1}^n x_i}(1-p_1)^{n-\sum\limits_{i=1}^n x_i}. \end{aligned}$$

记 $T(\boldsymbol{x})=\sum\limits_{i=1}^n x_i$, 似然比

$$\lambda(\boldsymbol{x})=\frac{p(\boldsymbol{x},p_1)}{p(\boldsymbol{x},p_0)}=\left(\frac{1-p_1}{1-p_0}\right)^n\left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right]^{T(x)}$$

由于 $p_1>p_0$,$1-p_0>1-p_1$, 所以 $\frac{p_1(1-p_0)}{p_0(1-p_1)}>1$, 故 $\lambda(\boldsymbol{x})$ 关于 $T(\boldsymbol{x})$ 严格单调增. 由于 r.v.$T(\boldsymbol{X})$ 服从离散型分布, 故需要随机化. 由 NP 引理可知检验函数为

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})>c,\\ r, & T(\boldsymbol{x})=c,\\ 0, & T(\boldsymbol{x})<c.\end{cases}$$

当 $H_0$ 成立时 $T(\boldsymbol{X})=\sum\limits_{i=1}^n X_i$ 服从二项分布 $b(n,p_0)$, 当 $\alpha$ 给定时,$c$ 由下列不等式确定:

$$\alpha_1=\sum_{k=c+1}^n\binom{n}{k}p_0^k(1-p_0)^{n-k}\leqslant\alpha\leqslant\sum_{k=c}^n\binom{n}{k}p_0^k(1-p_0)^{n-k}.$$

取

$$r=\frac{\alpha-\alpha_1}{\binom{n}{c}p_0^c(1-p_0)^{n-c}},$$

则必有

$$E_{p_0}[\varphi(\boldsymbol{X})]=P_{p_0}(T(\boldsymbol{X})>c)+r\cdot P_{p_0}(T(\boldsymbol{X})=c)=\alpha.$$

因此 $\varphi(\boldsymbol{x})$ 为水平为 $\alpha$ 的 UMPT.

由于上述检验函数 $\varphi(\boldsymbol{x})$ 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 $\alpha$ 的 UMPT.

**Note 5.4.3.** 关于随机化检验问题. 本例中当出现 $T(\boldsymbol{x})=\sum\limits_{i=1}^{n}x_i=c$ 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 $\left(\text{如本例中} \sum\limits_{i=1}^{n}x_i=c\right)$ 需随机化时再做一次试验. 试验结果为 $A$ 或 $\overline{A}$, 成功概率为 $P(A)=r$. 若 $A$ 发生, 则拒绝 $H_0$; 否则接受 $H_0$.

**Example 5.4.3.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\alpha$ 的 UMPT.

**解:** 服从均匀分布的样本 $\boldsymbol{X}$ 的密度函数和似然比分别为

$$\begin{aligned} f(\boldsymbol{x},\theta)=&\frac{1}{\theta^n}I_{(0,\theta)}(x_{(n)}),\\ \lambda(\boldsymbol{x})=&\frac{f(\boldsymbol{x},\theta_1)}{f(\boldsymbol{x},\theta_0)}=\begin{cases}\left(\frac{\theta_0}{\theta_1}\right)^n, & 0<x_{(n)}<\theta_0,\\ \infty, & \theta_0<x_{(n)}<\infty.\end{cases}\end{aligned}$$

此处定义了 $0/0=\infty$, 因 $\lambda(\boldsymbol{x})$ 关于 $T(\boldsymbol{x})=x_{(n)}$ 非降, 故由 NP 引理, 可知水平为 $\alpha$ 的 UMPT 函数有形式

$$\varphi(\boldsymbol{x})=\begin{cases}1, & x_{(n)}>c\\ 0, & x_{(n)}\leqslant c.\end{cases}$$

$T=X_{(n)}$ 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,$T(\boldsymbol{X})$ 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

$$E_{\theta_0}[\varphi(\boldsymbol{X})]=\int_0^{\infty}\varphi(t)g_{\theta_0}(t)\mathrm{d}t=\int_c^{\theta_0}\frac{nt^{n-1}}{\theta_0^n}\mathrm{d}t=1-\frac{c^n}{\theta_0^n}=\alpha,$$

故得 $c=\theta_0\sqrt[n]{1-\alpha}$, 因此

$$\varphi(\boldsymbol{x})=\begin{cases}1, & x_{(n)}>\theta_0\sqrt[n]{1-\alpha},\\ 0, & x_{(n)}\leqslant\theta_0\sqrt[n]{1-\alpha},\end{cases}$$

为一个水平为 $\alpha$ 的 UMPT.

由于此检验 $\varphi(\boldsymbol{x})$ 与 $\theta_1$ 无关, 故它也是

$$H_0:\theta=\theta_0\leftrightarrow H_1':\theta>\theta_0$$

的水平为 $\alpha$ 的 UMPT.

**Note 5.4.4.** 由上面三个例子可见 UMPT 检验函数 $\varphi(\boldsymbol{x})$ 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,$\boldsymbol{X} = (X_1, \cdots, X_n)$ 为自总体 $X$ 中抽取的随机样本,$T = T(\boldsymbol{X})$ 为 $\theta$ 的充分统计量, 则由式(5.4.4)和式(5.4.5)定义的检验函数 $\varphi(\boldsymbol{x})$ 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

设样本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 的分布族为下列指数族:

$$f(x,\theta) = c(\theta)\exp\{Q(\theta)T(\boldsymbol{x})\}h(x), \tag{5.4.7}$$

其中 $c(\theta) > 0$ 和 $Q(\theta)$ 为 $\theta$ 的函数,$T(\boldsymbol{x})$ 和 $h(\boldsymbol{x})$ 是样本 $\boldsymbol{x}$ 的函数.

对如下单边检验问题:

$$H_0 : \theta \leqslant \theta_0 \leftrightarrow H_1 : \theta > \theta_0, \tag{5.4.8}$$

有下列重要结论.

**Theorem 5.4.2.** 设样本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 的分布为指数族(5.4.7), 参数空间 $\Theta$ 为整个实数集 $\mathbb{R} = (-\infty, +\infty)$ 上的一有限或无限区间,$\theta_0$ 为 $\Theta$ 的一个内点且 $Q(\theta)$ 为 $\theta$ 的严格增函数, 则检验问题(5.4.8)的水平为 $\alpha$ 的 UMPT 存在 $(0 < \alpha < 1)$, 且有形式

$$\varphi(\boldsymbol{x}) = \begin{cases} 1, & T(\boldsymbol{x}) > c, \\ r, & T(\boldsymbol{x}) = c, \\ 0, & T(\boldsymbol{x}) < c, \end{cases} \tag{5.4.9}$$

其中 $c$ 和 $r(0 \leqslant r \leqslant 1)$ 满足条件

$$E_{\theta_0}[\varphi(\boldsymbol{X})] = P_{\theta_0}(T(\boldsymbol{X}) > c) + r \cdot P_{\theta_0}(T(\boldsymbol{X}) = c) = \alpha. \tag{5.4.10}$$

证明需要任取 $\theta_1 > \theta_0$, 首先考虑检验问题

$$H_0' : \theta = \theta_0 \leftrightarrow H_1' : \theta = \theta_1 \tag{5.4.11}$$

余略.

**Note 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 $\alpha$ 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 $\theta$ 的严格增函数” 为 “$Q(\theta)$ 为 $\theta$ 的严格降函数”, 其余不变, 则检验问题(5.4.8)的水平为 $\alpha$ 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

$$H_0:\theta\geqslant\theta_0\leftrightarrow H_1:\theta<\theta_0, \tag{5.4.12}$$

关于这一检验问题的水平为 $\alpha$ 的 UMPT 有下列定理.

**Theorem 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 $\alpha$ 的 UMPT 存在, 且有形式

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})<c,\\ r, & T(\boldsymbol{x})=c,\\ 0, & T(\boldsymbol{x})>c,\end{cases} \tag{5.4.13}$$

其中 $c$ 和 $r(0\leqslant r\leqslant 1)$ 满足条件

$$E_{\theta_0}[\varphi(\boldsymbol{X})]=P_{\theta_0}(T(\boldsymbol{X})<c)+r\cdot P_{\theta_0}(T(\boldsymbol{X})=c)=\alpha, \tag{5.4.14}$$

此定理的证明方法与定理5.4.2类似, 从略.

**Note 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 $\alpha$ 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 $\theta$ 的严格增函数” 为 “$Q(\theta)$ 为 $\theta$ 的严格降函数”, 其余条件不变, 则检验问题(5.4.12)的水平为 $\alpha$ 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.

**Example 5.4.4.** 问题与例5.4.1相同, 即设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从正态总体 $N(\theta,1)$ 中抽取的简单样本, 求检验问题 $H_0:\theta\leqslant\theta_0\leftrightarrow H_1:\theta>\theta_0$ 的 UMPT, 此处 $\theta_0$ 和检验水平 $\alpha$ 给定.

**解:** 正态分布为指数族分布, 样本密度为

$$\begin{aligned}f(x_1,\cdots x_n,\theta)=&(2\pi)^{-\frac{n}{2}}\exp\{-n\theta^2/2\}\exp\{n\theta\overline{x}\}\exp\left\{-\sum_{i=1}^n x_i^2/2\right\}\\=&c(\theta)\exp\{Q(\theta)T(\boldsymbol{x})\}h(\boldsymbol{x}),\end{aligned}$$

其中 $c(\theta)=(2\pi)^{n/2}\exp\{-n\theta^2/2\}$,$h(\boldsymbol{x})=\exp\left\{-\sum\limits_{i=1}^n x_i^2/2\right\}$,$T(\boldsymbol{x})=\overline{x}$,$Q(\theta)=n\theta$ 为 $\theta$ 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 $\alpha$ 的 UMPT 由下式给出:

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})>c,\\ 0, & T(\boldsymbol{x})\leqslant c.\end{cases}$$

由于 $T(\boldsymbol{X})=\overline{X}\sim N(\theta,1/n)$, 故 $\sqrt{n}(\overline{X}-\theta)\sim N(0,1)$, 令

$$\alpha=E_{\theta_0}[\varphi(\boldsymbol{X})]=P_{\theta_0}(T(\boldsymbol{X})>c)=P_{\theta_0}(\sqrt{n}(\overline{X}-\theta_0)>\sqrt{n}(c-\theta_0)),$$

可知 $\sqrt{n}(c-\theta_0)=u_\alpha$, 即 $c=\theta_0+u_\alpha/\sqrt{n}$. 因 $T(\boldsymbol{x})=\overline{x}$, 故水平为 $\alpha$ 的 UMPT 为

$$\varphi(\boldsymbol{x})=\begin{cases}1, & \overline{x}>\theta_0+u_\alpha/\sqrt{n},\\ 0, & \overline{x}\leqslant\theta_0+u_\alpha/\sqrt{n}.\end{cases}$$

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.

**Example 5.4.5.** 从一大批产品中抽取 $n$ 个检查其结果, 得样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$, 其中若第 $i$ 个产品为废品, 则 $X_i=1$; 否则 $X_i=0$,$i=0,1,\cdots,n$. 求

$$H_0: p\leqslant p_0 \leftrightarrow H_1: p>p_0$$

的水平为 $\alpha$ 的 UMPT, 其中 $p_0$ 和 $\alpha$ 给定.

**解:** 令 $T=T(\boldsymbol{X})=\sum\limits_{i=1}^n X_i$ 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 $c(p)=(1-p)^n$,$t=\sum\limits_{i=1}^n x_i$ 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

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})>c,\\ r, & T(\boldsymbol{x})=c,\\ 0, & T(\boldsymbol{x})<c,\end{cases}$$

其中 $c$ 由下列不等式决定:

$$\alpha_1=\sum_{i=c+1}^n\binom{n}{i}p_0^i(1-p_0)^{n-i}\leqslant\alpha\leqslant\sum_{i=c}^n\binom{n}{i}p_0^i(1-p_0)^{n-i}.$$

取 $r$ 为

$$r = \frac{\alpha - \alpha_1}{\binom{n}{t} p_0^c (1-p_0)^{n-c}}$$

则必有

$$E_{p_0}[\varphi(\boldsymbol{X})] = P_{p_0}(T(\boldsymbol{X}) > c) + r \cdot P_{p_0}(T(\boldsymbol{X}) = c) = \alpha,$$

因此上述检验 $\varphi(\boldsymbol{x})$ 为水平为 $\alpha$ 的 UMPT. 这是对例5.4.2的补充.

**Example 5.4.6.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 为自 Poisson 总体 $P(\lambda)$ 中抽取的随机样本,$\lambda > 0$ 为未知参数. 求

$$H_0 : \lambda \leqslant \lambda_0 \leftrightarrow H_1 : \lambda > \lambda_0$$

的水平为 $\alpha$ 的 UMPT, 其中 $\lambda_0$ 和 $\alpha$ 给定.

**解:** Poisson 分布为指数族分布. 样本 $\boldsymbol{X}$ 的密度函数为

$$f(\boldsymbol{x}, \lambda) = \frac{\lambda^{T(\boldsymbol{x})} e^{-n\lambda}}{x_1! \cdots x_n!} = c(\lambda) \exp\{Q(\lambda) T(\boldsymbol{x})\} h(\boldsymbol{x}),$$

其中 $c(\lambda) = e^{-n\lambda}$,$T(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$,$h(\boldsymbol{x}) = 1/(x_1! \cdots x_n!)$,$Q(\lambda) = \log \lambda$ 为 $\lambda$ 的严格增函数. 由定理5.4.2可知

$$\varphi(\boldsymbol{x}) = \begin{cases} 1, & T(\boldsymbol{x}) > c, \\ r, & T(\boldsymbol{x}) = c, \\ 0, & T(\boldsymbol{x}) < c, \end{cases}$$

其中 $c$ 由下列不等式确定 (注意检验统计量 $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i \sim P(n\lambda)$):

$$\alpha_1 = \sum_{k=c+1}^{\infty} \frac{(n\lambda_0)^k e^{-n\lambda_0}}{k!} \leqslant \alpha \leqslant \sum_{k=c}^{\infty} \frac{(n\lambda_0)^k e^{-n\lambda_0}}{k!}.$$

取 $r$ 为

$$r = \frac{(\alpha - \alpha_1) c!}{(n\lambda_0)^c e^{-n\lambda_0}},$$

则必有

$$E_{\lambda_0}[\varphi(\boldsymbol{X})] = P_{\lambda_0}(T(\boldsymbol{X}) > c) + r \cdot P_{\lambda_0}(T(\boldsymbol{X}) = c) = \alpha,$$

故上述检验 $\varphi(\boldsymbol{x})$ 为水平为 $\alpha$ 的 UMPT.

**Example 5.4.7.** 设 $X = (X_1, \cdots, X_n)$ 为自指数分布总体 $\mathrm{Exp}(\lambda)$ 中抽取的随机样本,$\lambda > 0$ 为未知参数. 求

$$H_0 : \lambda \leqslant \lambda_0 \leftrightarrow H_1 : \lambda > \lambda_0$$

的水平为 $\alpha$ 的 UMPT, 其中 $\lambda_0$ 和 $\alpha$ 给定.

**解:** 指数分布属于指数族. 样本 $\boldsymbol{X}$ 的密度函数为

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

其中 $c(\lambda)=\lambda^n$,$h(\boldsymbol{x})=I_{[x_i>0,i=1,2,\cdots,n]}$,$T(\boldsymbol{x})=\sum\limits_{i=1}^n x_i$,$Q(\lambda)=-\lambda$ 为 $\lambda$ 的单调降函数, 故由注(5.4.5)可知

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})<c,\\ 0, & T(\boldsymbol{x})\geqslant c.\end{cases}$$

由推论2.4.5可知 $2\lambda T(X)\sim\chi^2_{2n}$, 故有

$$\alpha=P_{\lambda_0}(T(\boldsymbol{X})<c)=P_{\lambda_0}(2\lambda_0T(\boldsymbol{X})<2\lambda_0c),$$

因此 $2\lambda_0c=\chi^2_{2n}(1-\alpha)$, 即 $c=\chi^2_{2n}(1-\alpha)/(2\lambda_0)$. 因此

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})<\frac{1}{2\lambda_0}\chi^2_{2n}(1-\alpha),\\ 0, & T(\boldsymbol{x})\geqslant\frac{1}{2\lambda_0}\chi^2_{2n}(1-\alpha)\end{cases}$$

为水平为 $\alpha$ 的 UMPT.

**Example 5.4.8.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为自正态总体 $N(0,\sigma^2)$ 中抽取的随机样本,$\sigma^2$ 为未知参数. 求

$$H_0:\sigma^2\geqslant\sigma_0^2\leftrightarrow H_1:\sigma^2<\sigma_0^2$$

的水平为 $\alpha$ 的 UMPT, 其中 $\sigma_0^2$ 和 $\alpha$ 给定.

**解:** 正态分布 $N(0,\sigma^2)$ 为指数族, 样本 $X$ 密度函数为

$$f(\boldsymbol{x},\sigma^2)=\prod_{i=1}^n f(x_i,\sigma^2)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\}$$

其中 $c(\sigma)=(2\pi\sigma^2)^{-n/2}$,$h(\boldsymbol{x})\equiv 1$,$T(\boldsymbol{x})=\sum\limits_{i=1}^n x_i^2$,$Q(\sigma^2)=-1/(2\sigma^2)$ 为 $\sigma^2$ 的严格单调增函数, 由定理5.4.3可知

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})<c,\\ 0, & T(\boldsymbol{x})\geqslant c.\end{cases}$$

由于 $\sum\limits_{i=1}^n X_i^2/\sigma^2\sim\chi^2_n$, 令

$$\alpha=E_{\sigma_0^2}[\varphi(\boldsymbol{X})]=P_{\sigma_0^2}\left(\sum_{i=1}^n X_i^2<c\right)=P_{\sigma_0^2}\left(\sum_{i=1}^n X_i^2/\sigma_0^2<\frac{c}{\sigma_0^2}\right),$$

故有 $c/\sigma_0^2=\chi_n^2(1-\alpha)$, 即 $c=\sigma_0^2\chi_n^2(1-\alpha)$. 因此

$$\varphi(\boldsymbol{x})=\begin{cases}1, & T(\boldsymbol{x})<\sigma_0^2\ \chi_n^2(1-\alpha),\\ 0, & T(\boldsymbol{x})\geqslant\sigma_0^2\ \chi_n^2(1-\alpha)\end{cases}$$

为水平为 $\alpha$ 的 UMPT.

### *5.4.4 无偏检验

**Definition 5.4.2.** 设 $\varphi$ 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 $\beta_\varphi(\theta_0)\leqslant\alpha$, 对 $\forall\theta_1\in\Theta_1$ 有 $\beta_\varphi(\theta_1)\geqslant\alpha$, 则称 $\varphi$ 为水平为 $\alpha$ 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 $\varphi$ 为 $H_0\to H_1$ 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

$$\mathscr{U}_\alpha=\{\varphi:\varphi\ \text{为检验问题(5.4.1)的水平}\ \alpha\ \text{的无偏检验}\},$$

即 $\mathscr{U}_\alpha$ 为一切水平为 $\alpha$ 的无偏检验的类.

**Definition 5.4.3.** 若 $\varphi\in\mathscr{U}_\alpha$ 且对任何 $\varphi_1\in\mathscr{U}_\alpha$, 有

$$\beta_\varphi(\theta)\geqslant\beta_{\varphi_1}(\theta),\quad \theta\in\Theta_1,$$

则称 $\varphi$ 是式(5.4.1)的一个水平为 $\alpha$ 的**一致最优无偏检验** (uniformly most powerful unbiased test,UMPUT).

**Note 5.4.7.** 由上述定义可知任一 UMPT 必为 UMPUT. 说明如下: 记 UMPT$\varphi$ 的功效函数为 $\beta_\varphi(\theta)$, 由 UMPT 的定义可知 $\beta_\varphi(\theta)\leqslant\alpha$, 对一切 $\theta\in\Theta_0$, 又显见 $\varphi^*\equiv\alpha$ 是水平为 $\alpha$ 的检验, 由 UMPT 定义可知 $\beta_\varphi(\theta)\geqslant\beta_{\varphi^*}(\theta)\equiv\alpha$, 对一切 $\theta\in\Theta_1$. 可见有

$$\beta_\varphi(\theta_1)\geqslant\alpha\geqslant\beta_\varphi(\theta_0),\quad \forall\theta_0\in\Theta_0,\theta_1\in\Theta_1,$$

故检验 $\varphi$ 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 $\alpha$ 的 UMPT 存在, 因而也是 UMPUT 的.

(1) $H_0 : \theta \leqslant \theta_0 \leftrightarrow H_1 : \theta > \theta_0$,

(2) $H_0 : \theta \geqslant \theta_0 \leftrightarrow H_1 : \theta < \theta_0$.

还可进一步证明下列两类单参数指数族的水平为 $\alpha$ 的 UMPUT 是存在的:

(3) $H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$,

(4) $H_0 : \theta_1 \leqslant \theta \leqslant \theta_2 \leftrightarrow H_1 : \theta < \theta_1$ 或 $\theta > \theta_2$.

## 5.5 假设检验和区间估计

### 5.5.1 如何由假设检验得到置信区间

设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从总体 $\{f(x, \theta), \theta \in \Theta\}$ 中抽取的简单样本, 目的是求参数 $\theta$ 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 $\alpha$ 的接受域 $\overline{D}$, 则有

$$P(\overline{D}|H_0) = 1 - \alpha, \tag{5.5.1}$$

解由 $\overline{D}$ 确定的不等式, 得到如下不等式:$\hat{\theta}_1(\boldsymbol{X}) \leqslant \theta_0 \leqslant \hat{\theta}_2(\boldsymbol{X})$, 由于式(5.5.1)是在条件 “$H_0 : \theta = \theta_0$” 下成立, 改 $\theta_0$ 为 $\theta$ 得 $\hat{\theta}_1(\boldsymbol{X}) \leqslant \theta \leqslant \hat{\theta}_2(\boldsymbol{X})$, 则 $[\hat{\theta}_1(\boldsymbol{X}), \hat{\theta}_2(\boldsymbol{X})]$ 即为所求的置信系数为 $1-\alpha$ 的置信区间.

若要求 $\theta$ 的置信上, 下限, 就需要考虑单边检验 $H_0'' : \theta \geqslant \theta_0 \leftrightarrow H_1'' : \theta < \theta_0$ 或 $H_0' : \theta \leqslant \theta_0 \leftrightarrow H_1' : \theta > \theta_0$ 的检验问题. 下面通过例子来说明.

**Example 5.5.1.** 设 $X_1, \cdots, X_n$ 为自正态总体 $N(\mu, \sigma^2)$ 抽取的随机样本.$\mu$,$\sigma^2$ 皆未知, 要分别求 $\mu$ 和 $\sigma^2$ 的置信系数为 $1-\alpha$ 的置信区间和置信上下限.

**解:** 先考虑 $\mu$ 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 $\alpha$ 的检验的接受域 $\overline{D} = \{(X_1, \cdots, X_n) : |T| \leqslant t_{n-1}(\alpha/2)\}$, 其中 $T = \sqrt{n}(\overline{X} - \mu_0)/S$. 记 $\theta = (\mu, \sigma^2)$, 故有

$$P_\theta \left( \left| \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} \right| \leqslant t_{n-1}(\alpha/2) \middle| H_0 \right) = 1 - \alpha. \tag{5.5.2}$$

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 $\mu_0$ 用 $\mu$ 代替是等价的. 解式(5.5.2)括号中的不等式得

$$\overline{X} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2) \leqslant \mu \leqslant \overline{X} + \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2),$$

因此

$$\left[ \overline{X} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2), \overline{X} + \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2) \right]$$

即为 $\mu$ 的置信系数为 $1-\alpha$ 的置信区间.

若要求 $\mu$ 的置信下限, 则考虑检验问题

$$H_0': \mu \leqslant \mu_0 \leftrightarrow H_1': \mu > \mu_0.$$

在 5.2 节中已给出水平为 $\alpha$ 的接受域 $\overline{D} = \{(X_1, \cdots, X_n) : T \leqslant t_{n-1}(\alpha)\}$, 故有

$$P_\theta\left(\frac{\sqrt{n}(\overline{X}-\mu_0)}{S} \leqslant t_{n-1}(\alpha) \middle| \mu = \mu_0\right) = 1-\alpha.$$

解括号中的不等式得

$$\overline{X} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha) \leqslant \mu_0,$$

再改 $\mu_0$ 为 $\mu$ 得到 $\overline{X} - St_{n-1}(\alpha)/\sqrt{n} \leq \mu < \infty$. 因此 $\mu$ 的置信系数为 $1-\alpha$ 的置信下限为 $\overline{X} - St_{n-1}(\alpha)/\sqrt{n}$. 同理, 可求 $\mu$ 的置信系数为 $1-\alpha$ 的置信上限为 $\overline{X} + St_{n-1}(\alpha)/\sqrt{n}$.

关于正态总体方差 $\sigma^2$ 的置信区间和置信上下限留作练习.

**Example 5.5.2.** 设 $X_1, \cdots, X_m$ 和 $Y_1, \cdots, Y_n$ 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 $X_1, \cdots, X_m$ 和 $Y_1, \cdots, Y_n$ 独立. 令 $\mu = \mu_2 - \mu_1$, 求 $\mu$ 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 $\overline{D} = \{(\boldsymbol{X}, \boldsymbol{Y}) : |T_w| \leqslant t_{n+m-2}(\alpha/2)\}$. 检验统计量为

$$T_w = \frac{\overline{Y} - \overline{X} - \mu_0}{S_w}\sqrt{\frac{mn}{m+n}}$$

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

$$P_\theta\left(|T_w| \leqslant t_{m+n-2}(\alpha/2) | H_0\right) = 1-\alpha, \tag{5.5.3}$$

改 $\mu_0$ 为 $\mu$, 解式(5.5.3)括号中的不等式得到

$$\overline{Y} - \overline{X} - S_w t_{n+m-2}\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{m}+\frac{1}{n}} \leqslant \mu \leqslant \overline{Y} - \overline{X} + S_w t_{n+m-2}\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{m}+\frac{1}{n}}.$$

因此 $\mu = \mu_2 - \mu_1$ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left[\overline{Y} - \overline{X} - S_w t_{n+m-2}\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{m}+\frac{1}{n}}, \overline{Y} - \overline{X} + S_w t_{n+m-2}\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{m}+\frac{1}{n}}\right].$$

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

$$\overline{Y}-\overline{X}-S_w t_{n+m-2}(\alpha)\sqrt{1/m+1/n}$$

和

$$\overline{Y}-\overline{X}+S_w t_{n+m-2}(\alpha)\sqrt{1/m+1/n}.$$

这里假定了两总体有相同的方差 $\sigma^2$. 若去掉这一假设, 假定两总体的方差分别为 $\sigma_1^2$ 和 $\sigma_2^2$, 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 $\theta$ 的置信水平为 $1-\alpha$ 的区间估计 $\left[\hat{\theta}_1,\hat{\theta}_2\right]$, 对给定的 $\theta_0$ 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 $\alpha$ 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 $H_0':\theta\geqslant\theta_0\leftrightarrow H_1':\theta<\theta_0$ 和 $H_0'':\theta\leqslant\theta_0\leftrightarrow H_1'':\theta>\theta_0$ 的水平为 $\alpha$ 的检验.

### *5.5.3 一致最精确置信集和一致最优检验

#### 1. 引言和定义

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.

**Definition 5.5.1.** 称 $\hat{\theta}_U^*(\boldsymbol{X})$,$\hat{\theta}_L^*(\boldsymbol{X})$ 和 $[\hat{\theta}_1^*(\boldsymbol{X})$,$\hat{\theta}_2^*(\boldsymbol{X})]$ 分别是 $\theta$ 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 $\hat{\theta}_U(\boldsymbol{X})$, 以及任何 $\theta<\theta'$, 有

$$P_\theta(\hat{\theta}_U^*(\boldsymbol{X})\geqslant\theta')\leqslant P_\theta(\hat{\theta}_U(\boldsymbol{X})\geqslant\theta');$$

(2) 对任何置信水平为 $1-\alpha$ 的置信下限 $\hat{\theta}_L(\boldsymbol{X})$, 以及任何 $\theta>\theta'$, 有

$$P_\theta(\hat{\theta}_L^*(\boldsymbol{X})\leqslant\theta')\leqslant P_\theta(\hat{\theta}_U(\boldsymbol{X})\leqslant\theta');$$

(3) 对任何置信水平为 $1-\alpha$ 的置信区间 $[\hat{\theta}_1(\boldsymbol{X}), \hat{\theta}_2(\boldsymbol{X})]$, 以及任何 $\theta \neq \theta'$, 有

$$P_\theta(\hat{\theta}_1^*(\boldsymbol{X}) \leqslant \theta' \leqslant \hat{\theta}_2^*(\boldsymbol{X})) \leqslant P_\theta(\hat{\theta}_1(\boldsymbol{X}) \leqslant \theta' \leqslant \hat{\theta}_2(\boldsymbol{X})).$$

**Note 5.5.1.** 在上述定义中 $\theta$ 为真参数,$\theta'$ 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 $\theta'$ 的概率是最小的.

**2. 如何求 UMA 置信集**

**Theorem 5.5.1.** 由本节 5.5.1 小节中介绍过的由假设 $\theta=\theta_0, \theta \geqslant \theta_0, \theta \leqslant \theta_0$ 去构造置信水平为 $1-\alpha$ 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 $\alpha$ 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.

证明略.

**Example 5.5.3.** 设 $X_1, \cdots, X_n$i.i.d.$\sim N(0,1)$, 求 $\theta$ 的置信水平 $1-\alpha$ 的 UMA 置信上下限.

**解:** 显然检验问题 $H_0: \theta \leqslant \theta_0 \leftrightarrow H_1: \theta > \theta_0$ 的水平为 $\alpha$ 的 UMPT 的接受域为 $\{\sqrt{n}(\overline{X}-\theta_0) \leqslant u_\alpha\} \iff \{\overline{X}-u_\alpha/\sqrt{n} \leqslant \theta_0\}$. 由定理5.5.1可知 $\overline{X}-u_\alpha/\sqrt{n}$ 就是 $\theta$ 的置信水平 $1-\alpha$ 的 UMA 置信下限, 同理可知 $\overline{X}+u_\alpha/\sqrt{n}$ 是 $\theta$ 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 $\mu$ 值与 0 相差有多大. 但如果被告知 $\mu$ 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,$\mu$ 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 $\mu$"事实上" 可以认为是 0, 而不只是接受 "$\mu=0$" 了. 若在后一场合, 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 $\mu$ 的可能范围很大, 实际上只能说对 $\mu$"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:$\mu$ 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 $\mu$ 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 $\mu$ 的值与 0 相距至少是 30, 不仅要否定 $\mu=0$, 从实际上看 $\mu$ 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

设 $X_1,\cdots,X_{16}$ 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

$$\left\{(X_1,\cdots,X_{16}):\left|\sqrt{16}\overline{X}\right|>1.96\right\}=\left\{(X_1,\cdots,X_n):\left|\overline{X}\right|>0.49\right\}.$$

设对一组样本 $X_1,\cdots,X_{16}$, 有 $\overline{X}=0.48$, 则根据拒绝域, 应接受 $H_0:\mu=0$; 又设有另一组样本 $X_1,\cdots,X_{16}$ 算得 $\overline{X}=0.12$, 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 $X_1,\cdots,X_n$, 算出 $\overline{X}$ 的具体值记为 $\overline{x}_0$, 则这组样本的 $p$ 值定义为

$$p=P\left(|\overline{X}|>|\overline{x}_0|\,|H_0\right)=P\left(\sqrt{n}\,|\overline{X}|>\sqrt{n}\,|\overline{x}_0|\,|H_0\right)=P\left(|U|>\sqrt{n}\,|\overline{x}_0|\right),$$

此处 $U=\sqrt{2}|\overline{X}|$, 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 $\overline{X}$ 的具体值为 $\overline{x}_0$, 它与假设 $\mu=0$ 的偏差 $|\overline{x}_0|$, 问: 达到 $|\overline{x}_0|$ 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 $\mu$ 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 $p\geqslant\alpha$, 但离 $\alpha$ 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 $\overline{x}_0=0.48$ 和 $\overline{x}_0=0.12$ 两种情形的 $p$ 值分别为

$$p=P(|\overline{X}|\geqslant 0.48\,|H_0)=P(|U|\geqslant 1.92)=0.0548;$$
$$p=P(|\overline{X}|\geqslant 0.12\,|H_0)=P(|U|\geqslant 0.48)=0.6312.$$

前一情况离水平 $\alpha$ 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

$$p=P(|T|>|t_0||H_0);\tag{5.5.4}$$

对单边检验, 若原假设为 $H_0:\theta\leqslant\theta_0$, 其否定域为 $T>c$, 则 $p$ 值为

$$p=P(T>t_0|\theta=\theta_0); \tag{5.5.5}$$

对单边检验问题, 若原假设为 $H_0:\theta\geqslant\theta_0$, 否定域为 $T<c$, 则 $p$ 值为

$$p=P(T<t_0|\theta=\theta_0). \tag{5.5.6}$$

将一样本问题中的 $U$ 检验,$t$ 检验和 $\chi^2$ 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 | $\sigma^2$ 已知 | $\mu=\mu_0$ | $\mu\neq\mu_0$ | $U=\dfrac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma}\sim N(0,1)$ | $p=P\left(\lvert U\rvert>\dfrac{\sqrt{n}\lvert\overline{x}_0-\mu_0\rvert}{\sigma}\right)$ |
| | 同上 | $\mu\leqslant\mu_0$ | $\mu>\mu_0$ | 同上 | $p=P\left(U>\dfrac{\sqrt{n}\lvert\overline{x}_0-\mu_0\rvert}{\sigma}\right)$ |
| | $\sigma^2$ 未知 | $\mu=\mu_0$ | $\mu\neq\mu_0$ | $T=\dfrac{\sqrt{n}(\overline{X}-\mu_0)}{S}\sim t_{n-1}$ | $p=P\left(\lvert T\rvert>\dfrac{\sqrt{n}\lvert\overline{x}_0-\mu_0\rvert}{s_0}\right)$ |
| | 同上 | $\mu\leqslant\mu_0$ | $\mu>\mu_0$ | 同上 | $p=P\left(T>\dfrac{\sqrt{n}\lvert\overline{x}_0-\mu_0\rvert}{s_0}\right)$ |
| | $\mu$ 未知 | $\sigma^2\leqslant\sigma_0^2$ | $\sigma^2>\sigma_0^2$ | $\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^2_{n-1}$ | $p=P\left(\chi^2>\dfrac{(n-1)s_0^2}{\sigma_0^2}\right)$ |
| 两样本 | $\sigma_1^2=\sigma_2^2$ 未知 | $\mu_1=\mu_2$ | $\mu_1\neq\mu_2$ | $T=\dfrac{\overline{Y}-\overline{X}}{S_w\sqrt{\frac{m+n}{mn}}}\sim t_{n+m-2}$ | $p=P\left(\lvert T\rvert>\dfrac{\lvert\overline{y}_0-\overline{x}_0\rvert}{S_{w_0}\sqrt{\frac{m+n}{mn}}}\right)$ |
| | 同上 | $\mu_1\leqslant\mu_2$ | $\mu_1>\mu_2$ | 同上 | $p=P\left(T>\dfrac{\overline{y}_0-\overline{x}_0}{S_{w_0}\sqrt{\frac{m+n}{mn}}}\right)$ |
| | $\mu_1$,$\mu_2$ 未知 | $\sigma_1^2\leqslant\sigma_2^2$ | $\sigma_1^2>\sigma_2^2$ | $F=S_1^2/S_2^2\sim F_{m-1,n-1}$ | $p=P\left(F>s_{10}^2/s_{20}^2\right)$ |

表5.5.1中一样本问题中 $\overline{x}_0$,$s_0^2$ 分别由 $\overline{X}$,$S^2$ 的具体样本算得的值; 两样本问题中 $\overline{y}_0$,$\overline{x}_0$,$s_{10}^2$,$s_{20}^2$ 和 $s_{w_0}$ 分别由 $\overline{Y}$,$\overline{X}$,$S_1^2$,$S_2^2$ 及 $S_{w_0}$ 的具体样本算得的值. 一样本问题中的 $\overline{X}$,$S^2$ 及两样本问题中的 $\overline{Y}$,$\overline{X}$,$S_1^2$,$S_2^2$ 和 $S_w^2$ 的表达式与 5.2 节中相同.

**Example 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 $\overline{X} = 32.80$ 元,$S = 20.80$ 元. 假定每月长途电话费用服从正态分布 $N(\mu, \sigma^2)$,$\sigma^2$ 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 $n = 25$,$\mu_0 = 30$,$\overline{x}_0 = 32.80$,$s_0 = 20.80$, 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, 即有足够根据认为 $H_0 : \mu = 30$ 成立.

# Chapter 6

# 非参数假设检验

## 6.1 引言

略.

## 6.2 符号检验及符号秩和检验

### 6.2.1 符号检验法

**Example 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 $Z_i = X_i - Y_i$,$i = 1, \cdots, N$. 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

$$S_i = \begin{cases} +, & Z_i > 0, \\ -, & Z_i < 0, \\ 0, & Z_i = 0, \end{cases} \tag{6.2.1}$$

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 $S_1, \cdots, S_N$. 检验问题

$$H_0 : \text{甲乙两种酒一样好} \leftrightarrow H_1 : \text{甲乙两种酒不一样} \tag{6.2.2}$$

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 $S_1,\cdots,S_N$ 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 $X\sim b(n,\theta)$,$0\leqslant\theta\leqslant 1$, 要检验

$$H_0:\ \theta=\frac{1}{2}\leftrightarrow H_1:\ \theta\neq\frac{1}{2}. \tag{6.2.3}$$

一个水平为 $\alpha$ 的检验的否定域为

$$\{X=n_+\geqslant c\text{或}n_+\leqslant d\},$$

其中 $c$ 和 $d$ 的值由下式确定:

$$\sum_{i=c}^{n}\binom{n}{i}\left(\frac{1}{2}\right)^n\leqslant\frac{\alpha}{2},\quad d=n-c,$$

其中 $c$ 的值已制成表, 可查表获得. 可以通过查附表直接获得 $c$ 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 $S_1,\cdots,S_N$ 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

$$p=\sum_{i=0}^{x_0'}\binom{n}{i}\left(\frac{1}{2}\right)^n+\sum_{i=n-x_0'}^{n}\binom{n}{i}\left(\frac{1}{2}\right)^n. \tag{6.2.4}$$

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 $\alpha$, 则当 $p<\alpha$ 时否定 $H_0$, 当 $p\geqslant\alpha$ 时接受 $H_0$.

在例6.2.1中, 令 $N=13,S_1,\cdots,S_{13}$ 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 $\alpha=0.10$, 查表 "符号检验临界值表" 得 $c=10$[1], 故 $d=n-c=2$. 故检验的否定域 $D=\{X=n_+\geqslant 10\text{或}X\leqslant 2\}$. 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 $H_0$.

[1] 表中的概率的定义是单边的, 但是这里的否定域的形式是双边的, 因此在查表时需要将水平减半, 即左侧水平 + 右侧水平 = 总水平 =0.01, 因此这里的概率并没有错误.

**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 $n\to\infty$ 时有

$$U=\frac{X-E(X)}{\sqrt{D(X)}}=\frac{X-n/2}{\sqrt{n/4}}=\frac{2X-n}{\sqrt{n}}\xrightarrow{\mathscr{L}}N(0,1).$$

因此检验问题(6.2.2)水平近似为 $\alpha$ 的检验的否定域是

$$\{X:|U|>u_{\alpha/2}\}\tag{6.2.5}$$

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 $X\sim b(n,\theta)$,$0\leqslant\theta\leqslant 1$, 而

$$H_0':\theta\leqslant\frac{1}{2}\leftrightarrow H_1':\theta>\frac{1}{2}.$$

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 $\alpha$ 的检验的否定域是

$$\{X:U>u_\alpha\}.$$

**Example 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

$$p=P(|U|\geqslant 3\mid H_0)=2(1-0.9987)=0.0026$$

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.

**Example 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 $\xi$ 和 $\eta$. 设 $\xi$ 和 $\eta$ 为连续随机变量, 其分布函数分别为 $F$ 和 $G$. 检验问题是

$$H_0: F=G \leftrightarrow H_1: F\neq G. \tag{6.2.6}$$

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

$$X_i=\mu_i+\xi_i,\quad Y_i=\mu_i+\eta_i,\quad i=1,2,\cdots,n,$$

其中 $\mu_i$ 为第 $i$ 天水中的含氯量,$\xi_i$ 和 $\eta_i$ 分别表示第 $i$ 天化验室 A,B 的测量误差. 显然 $\xi_1,\cdots,\xi_n$ 和 $\eta_1,\cdots,\eta_n$ 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 $\mu_i$ 和 $\mu_j$ 的差异有关. 因此虽然 $X_1,\cdots,X_n$ 相互独立, 但不能假定它们同分布,$Y_1,\cdots,Y_n$ 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 $\mu_i$ 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 $\mu_i$ 的影响消除掉. 令

$$Z_i=X_i-Y_i=\xi_i-\eta_i,\quad i=1,2,\cdots,n. \tag{6.2.7}$$

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 $Z_1,\cdots,Z_n$ 可看成来自总体 $Z$ 的随机样本, 即 $Z_1,\cdots,Z_n$ 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 $Z_1,\cdots,Z_n$ 中取正值的个数,$n_-$ 为 $Z_1,\cdots,Z_n$ 中取负值的个数, 它们都是随机变量. 由于假定了 $\xi$ 和 $\eta$ 是连续随机变量, 故 $Z_1,\cdots,Z_n$ 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 $X=n_+\sim b(n,\theta)$,$0\leqslant\theta\leqslant1$, 检验问题为

$$H_0': \theta=\frac{1}{2} \leftrightarrow H_1': \theta\neq\frac{1}{2},$$

否定域 $D=\{n_+ \geqslant c 或 n_+ \leqslant d\}$.

因此, 在给定显著性水平 $\alpha$ 之后,$c$ 和 $d$ 的值由

$$\sum_{k=c}^{n}\binom{n}{k}\left(\frac{1}{2}\right)^n \leqslant \frac{\alpha}{2}, \quad d=n-c$$

所确定. 也可以通过查表直接获得 $c$ 的值, 再按公式计算得到 $d$ 的值.

在本例中 $n=11$,$\alpha=0.10$, 查表得 $c=9$,$d=n-c=2$. 故水平 $\alpha=0.10$ 的符号检验的否定域为

$$\{n_+ \leqslant 2 或 n_+ \geqslant 9\}.$$

作差值 $z_i=x_i-y_i$, 得

$$\begin{array}{llllll} 0.15, & -0.04, & -0.14, & 0.02, & -0.06, & -0.05, \\ -0.03, & -0.01, & -0.11, & -0.07, & -0.12, & \end{array}$$

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:

$$p=\sum_{i=0}^{2}\binom{11}{i}\left(\frac{1}{2}\right)^1+\sum_{i=9}^{11}\binom{11}{i}\left(\frac{1}{2}\right)^{11}=0.0654<\alpha=0.10,$$

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.

**Example 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 | 0.32 | 1.35 | 1.38 | 1.41 | 1.44 | 1.47 | 1.50 | 0.53 |
| 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 $X_i, i=1,\cdots,100$, 令 $Y_i=X_i-1.40, i=1,\cdots,100$. 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 $X \sim b(100,\theta), 0 \leqslant \theta \leqslant 1$,

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 $\alpha$ 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.

**Example 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.

**Definition 6.2.1.** 设 $X_1, \cdots, X_n$ 为两两不相等的一组样本, 将其按大小排列为 $X_{(1)} < \cdots < X_{(n)}$, 若 $X_i = X_{(R_i)}$, 则称 $X_i$ 在样本 $(X_1, \cdots, X_n)$ 中的秩为 $R_i$.

显然, 若 $X_1, \cdots, X_n$ 为来自连续分布 $F(x)$ 的样本, 则以概率为 1 保证 $X_1, \cdots, X_n$ 中两两互不相等.

**Definition 6.2.2.** 设 $X_1, \cdots, X_n$ 为来自单个总体的样本, 或来自多个总体样本的合样本, 则 $\boldsymbol{R} = (R_1, R_2, \cdots, R_n)$ 称为 $(X_1, \cdots, X_n)$ 的**秩统计量** (rank statistics), 其中 $R_i$ 为 $X_i$ 的秩. 由 $\boldsymbol{R}$ 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

$$\overline{V}_i = \begin{cases} 1, & Z_i > 0, \\ 0, & \text{其他}. \end{cases}$$

$R_i$ 为 $|Z_i|$ 在 $(|Z_1|, \cdots, |Z_n|)$ 中的秩, 则 Wilcoxon **符号秩和** (the sum of Wilcoxon signed rank) 检验统计量定义为

$$W^+ = \sum_{i=1}^{n} \overline{V}_i R_i. \tag{6.2.8}$$

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

$$\left\{W^+ \leqslant d \text{或} W^+ \geqslant c\right\}, \tag{6.2.9}$$

此处 $d$ 和 $c$ 取决于 $n$(本例中 $n = 12$) 及指定的检验水平 $\alpha$, 即当给定 $\alpha$ 时,$c$,$d$ 分别由下列两式决定:

$$P(W^+ \leqslant d|H_0) \leqslant \alpha/2, P(W^+ \geqslant c|H_0) \leqslant \alpha/2.$$

$H_0$ 为真时 $W^+$ 本书不作详细展开. 对某些特定的 $\alpha$ 及不大的 $n$,$c$ 和 $d$ 可以查表求得, 在表中仅可查到 $c$, 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

$$\{W^+ \leqslant 17 \text{ 或} W^+ \geqslant 61\}.$$

Table 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 |

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.

**Note 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

$$E(W^+) = \frac{n(n+1)}{4}, D(W^+) = \frac{1}{24}n(n+1)(2n+1).$$

与 6.3 节的秩和统计量 $W$ 相似, 当 $H_0$ 成立且 $n \to \infty$ 时,$W^+$ 的标准化随机变量

$$W_*^+ = \frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \xrightarrow{\mathscr{L}} N(0,1), \tag{6.2.10}$$

故水平近似为 $\alpha$ 的双侧 $W^+$ 检验的否定域为

$$\left\{|W_*^+| > u_{\alpha/2}\right\}.$$

**Example 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

$$\left\{|W_*^+| = \left|\frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}\right| \geqslant u_{\alpha/2}\right\}.$$

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

### 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 $X_1, \cdots, X_m$ 和 $Y_1, \cdots, Y_n$ 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 $X_1, \cdots, X_m$ 和 $Y_1, \cdots, Y_n$ 独立. 要检验下列假设:

$$H_0: F_1 = F_2 \leftrightarrow H_1: F_1 \neq F_2. \tag{6.3.1}$$

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 $\sigma^2$ 皆未知,$-\infty < a, b < +\infty$,$\sigma^2 > 0$. 这时检验问题转化为

$$H_0' : a = b \leftrightarrow H_1' : a \neq b. \tag{6.3.2}$$

在这个家定下, 总体分布 $F_1$ 和 $F_2$ 只依赖于 3 个未知参数 $a$,$b$ 和 $\sigma^2$, 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 $\theta$ 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

$$H_0 : \theta = 0 \leftrightarrow H_1 : \theta \neq 0. \tag{6.3.3}$$

式(6.3.3)是一个很重要的假设检验问题. 在这一模型中, 假设 $F$ 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 $\theta$, 而它是一个实参数. 其实不然, 因为总体的分布与 $F$ 和 $\theta$ 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

$$H_0^* : \sigma = 1 \leftrightarrow H_1^* : \sigma \neq 1. \tag{6.3.4}$$

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.

**Definition 6.3.1.** 设 $X_1, \cdots, X_m$,$Y_1, \cdots, Y_n$ 这 $n+m$ 个值两两不相同, 把它们按大小排列, 结果为

$$Z_1 < Z_2 < \cdots Z_N, \quad N = m + n. \tag{6.3.5}$$

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 $X_1, \cdots, X_m$, $Y_1, \cdots, Y_n$ 中的秩为 $R_i$. 而 $Y_1, \cdots, Y_n$ 的秩和为

$$W = R_1 + \cdots + R_n \tag{6.3.6}$$

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 $X_1,\cdots,X_m$i.i.d.$\sim F(x)$, $Y_1,\cdots,Y_n$ 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 $Y_1,\cdots,Y_n$ 的秩和 $W$ 由式(6.3.6)给出. 现在这样推理: 每个 $R_i$ 都可取 $1,2,\cdots,N$ 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:

$$\text{当 } W\leqslant d \text{ 或 } W\geqslant c \text{ 时, 否定 } H_0. \tag{6.3.7}$$

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 $(R_1,\cdots,R_n)$ 的联合分布为

$$\begin{aligned}&P\left(R_1=r_1,\cdots,R_n=r_n\right)\\=&\begin{cases}\frac{1}{N(N-1)\cdots(N-n+1)}, & r_1,\cdots,r_n\leqslant N\text{为互不相同的自然数},\\0, & \text{其他}.\end{cases}\end{aligned}$$

由此不难形式地写出 $W$ 的分布. 从而由

$$\alpha=P(W\leqslant d \text{ 或} W\geqslant c\mid H_0)$$

定出 $c$ 和 $d$. 对较小的 $m,n$ 已制成表.

如果假设检验是单边的, 即要检验

$$H_0':\theta\leqslant 0\leftrightarrow H_1':\theta>0 \tag{6.3.8}$$

由于 $W=\sum\limits_{i=1}^{n}R_i$ 是 $Y_1,\cdots,Y_n$ 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,$R_1,\cdots,R_n$ 当 $\theta>0$ 时倾向于取集合 $\{1,2,\cdots,N\}$ 中较大的值 (此处 $N=m+n$), 同样 $\theta<0$, 则 $R_1,\cdots,R_n$ 倾向于取集合 $\{1,2,\cdots,N\}$ 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

$$\text{当 } W\geqslant c \text{ 时, 否定 } H_0';$$

同样, 检验问题

$$H_0'':\theta\geqslant 0\leftrightarrow H_1'':\theta<0 \tag{6.3.9}$$

的检验为

$$\text{当 } W\leqslant d \text{ 时, 否定 } H_0''.$$

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 |
| --- | --- | --- |
| $\theta = 0$ | $\theta \neq 0$ | $W \leqslant d$或$W \geqslant c$ |
| $\theta \leqslant 0$ | $\theta > 0$ | $W \geqslant$ |
| $c\theta \geqslant 0$ | $\theta < 0$ | $W \leqslant d$ |

如何确定临界值 $c$ 和 $d$? 对较小的 $m$,$n(m \leqslant n)$ 已经制成表. 关于此表, 做一下两点说明:

(1) 分别记 $X_1, \cdots, X_m$ 和 $Y_1, \cdots, Y_n$ 在合样本中的秩和为 $W_1$ 和 $W_2$, 则

$$W_1 + W_2 = 1 + 2 + \cdots + (m+n) = \frac{(m+n)(m+n+1)}{2}$$

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 $n \leqslant m$.

(2) 在 $c$ 值表中只给出了秩和检验的临界值 $P(W \geqslant c) \leqslant \alpha$ 中 $c$ 的值, 对 $P(W \leqslant d) \leqslant \alpha$ 的临界值 $d$ 如何利用此表求出? 可以证明 $P(W \leqslant d) = P(W \geqslant n(m+n+1) - d)$, 即若记 $c = n(m+n+1) - d$, 先对给定的 $\alpha, m, n$ 求出 $P(W \geqslant c) \leqslant \alpha$ 的临界值 $c$, 然后 $d$ 由公式 $d = n(m+n+1) - c$ 算出.

**Example 6.3.1.** 某种样报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

$$H_0: \ \theta \geqslant 0 \leftrightarrow H_1: \theta < 0,$$

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 $W \leqslant d$ 时否定原假设. 由前面关于表的使用说明 (2), 先从附表中由 $P(W \geqslant c) \leqslant \alpha$ 查出 $c$, 则 $d = n(m+n+1) - c$.

本题中 $m = n = 5$,$\alpha = 0.05$, 由表, 查出 $c = 36$, 故有

$$d = n(m+n+1) - c = 5 \times 11 - 36 = 19.$$

这表明

$$P(W \leqslant 19) = P(W \geqslant 36) \leqslant 0.05.$$

因此上述检验问题的否定域

$$D = \{(X, Y) : W \leqslant 19\}.$$

现将两组样本观察值按从小到大排成一列表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 |
| $\underline{1}$ | $\underline{2}$ | 3 | $\underline{4}$ | $\underline{5}$ | 6 | $\underline{7}$ | 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 $W = 19 \leqslant d(d = 19)$, 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 $\alpha$ 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

$$\begin{aligned} E(W) &= \frac{n(m+n+1)}{2} = \frac{n(N+1)}{2}, \\ D(W) &= \frac{mn(n+m+1)}{12} = \frac{mn(N+1)}{12}. \end{aligned}$$

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 $m, n \to \infty$ 时有

$$W^* = \frac{W - n(N+1)/2}{\sqrt{mn(N+1)/12}} \xrightarrow{\mathscr{L}} N(0,1).$$

因此可得检验问题(6.3.3)的水平近似为 $\alpha$ 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 $\alpha$ 的否定域, 详见表6.3.3.

表 6.3.3:Wilcoxon 两样本秩和检验 (大样本性质)

| $H_0$ | $H_1$ | 否定域 |
|---|---|---|
| $\theta = 0$ | $\theta \neq 0$ | $\lvert W^* \rvert \geqslant u_{\alpha/2}$ |
| $\theta \leqslant 0$ | $\theta > 0$ | $W^* \geqslant u_\alpha$ |
| $\theta \geqslant 0$ | $\theta < 0$ | $W^* \leqslant -u_\alpha$ |

**Example 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 $X_1, \cdots, X_m$ 和 $Y_1, \cdots, Y_n$ 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

$$D = \{(X, Y) : |W^*| \geqslant u_{0.005} = 2.58\},$$

其中

$$|W^*| = \left|\frac{W - n(n+m+1)}{\sqrt{mn(m+n+1)/12}}\right| = \left|\frac{6540 - 7240}{347.37}\right| = 2.02 < 2.58,$$

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 $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ 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 $X_1, \cdots, X_n$ 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 $X_1, \cdots, X_n$ 去检验假设

$$H_0 : \text{r.v.}X\text{的分布为}F \tag{6.4.1}$$

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 $X_1, \cdots, X_n$ 与理论分布 $F$ 偏差的量 $D = D(X_1, \cdots, X_n; F)$. 如果 $D$ 较大, 如 $D \geqslant c$, 则认为理论分布 $F$ 与数据 $X_1,\cdots,X_n$ 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:

$$p(d_0) = P(D \geqslant d_0|H_0)$$

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 $\alpha$ 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:

$$\text{当 } p(d_0) < \alpha \text{ 时否定 } H_0, \text{当 } p(d_0) \geqslant \alpha \text{ 时接受 } H_0. \tag{6.4.2}$$

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 $\chi^2$ 检验 (称为 Pearson $\chi^2$ 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 $\chi^2$ 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 $a_1, \cdots, a_r$ 的情形**

设 $X_1, \cdots, X_n$ 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 $p_1,p_2,\cdots,p_r$ 已知, 且 $\sum_{i=1}^{r} p_i=1$. 检验问题(6.4.1)写成

$$H_0: P(X=a_i)=p_i, \quad i=1,2,\cdots,r. \tag{6.4.3}$$

设样本 $X_1,\cdots,X_n$ 中等于 $a_i$ 的个数记为 $\nu_i$,$i=1,2,\cdots,r$, 则 $\nu_i$ 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 $\nu_i/n$ 为 $X_1,\cdots,X_n$ 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,

$$K_n=K_n\left(X_1,\cdots,X_n;F\right)=\sum_{i=1}^{r}\frac{\left(\nu_i-np_i\right)^2}{np_i} \tag{6.4.4}$$

的极限分布 (当 $n\to\infty$ 时) 为 $\chi_{r-1}^2$. 因此有如下定理.

**Theorem 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 $n\to\infty$ 时有

$$K_n\xrightarrow{\mathscr{L}}\chi_{r-1}^2,$$

即 $K_n$ 的分布收敛于自由度为 $r-1$ 的 $\chi^2$ 分布.

证明略.

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 $\chi_{r-1}^2$. 于是得到检验问题(6.4.3)的水平近似为 $\alpha$ 的检验

$$\text{当}K_n>\chi_{r-1}^2(\alpha)\text{时否定}H_0,\text{否则就接受}H_0. \tag{6.4.5}$$

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

$$p(k_0)=P(K_n\geqslant k_0|H_0)\approx P(\chi_{r-1}^2\geqslant k_0), \tag{6.4.6}$$

其中 $\chi_{r-1}^2$ 表示自由度为 $r-1$ 的 $\chi^2$ 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 $\alpha$ 给定, 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.

**Example 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 $A_i$, $i = 1, 2, 3, 4$). 按孟德尔理论, 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

$$\begin{aligned} H_0: &P(X=1)=\frac{9}{16}, \quad P(X=2)=\frac{3}{16}, \\ &P(X=3)=\frac{3}{16}, \quad P(X=4)=\frac{1}{16}. \end{aligned}$$

令 $\nu_1 = 315$,$\nu_2 = 108$,$\nu_3 = 101$,$\nu_4 = 32$, 则

$$\begin{aligned} k_0 = &\frac{(315-556\times 9/16)^2}{556\times 9/16} + \frac{(108-556\times 3/16)^2}{556\times 3/16} \\ &+ \frac{(101-556\times 3/16)^2}{556\times 3/16} + \frac{(32-556\times 1/16)^2}{556\times 1/16} \quad = 0.47 \end{aligned}$$

拟合优度为

$$p(k_0) = P(K_n \geqslant k_0 | H_0) \approx P(\chi_3^2 \geqslant 0.47) > 0.90,$$

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 $X_1, \cdots, X_n$ 是从总体 $X$ 中抽取的简单样本, 要检验

$$H_0: \text{r.v.}X\text{的分布为}F, \tag{6.4.7}$$

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 $a_1, \cdots, a_{r-1}$, 满足 $a_0 = -\infty < a_1 < a_2 < \cdots < a_{r-1} < +\infty = a_r$, 将实数轴 $(-\infty, +\infty)$ 分成 $r$ 个子区间

$$\begin{aligned} &I_1 = (-\infty, a_1), I_2 = [a_1, a_2), \cdots, \\ &I_j = [a_{j-1}, a_j), \cdots, I_r = [a_{r-1}, +\infty). \end{aligned} \tag{6.4.8}$$

若 r.v.$X$ 是 $m$ 维的,$m > 1$, 则将 $\mathbb{R}^m$ 分解为 $r$ 个彼此无公共点的区域 $I_1, \cdots, I_r$.

第二步, 计算 $r$ 个事件在 $H_0$ 成立下的概率

$$p_j = P_F(X \in I_j) = F(a_j) - F(a_{j-1}), \quad j = 1, 2, \cdots, r, \tag{6.4.9}$$

此处易见 $F(a_0)=0$,$F(a_r)=1$, 则检验问题(6.4.7)转化为

$$H_0: P(X \in I_j)=p_j, \quad j=1,2,\cdots,r. \tag{6.4.10}$$

第三步, 令 $\nu_j$ 为 $X_1,\cdots,X_n$ 落入式(6.4.8)中的区间 $I_j$ 的观察频数,$j=1,2,\cdots,r$. 计算检验统计量

$$K_n=\sum_{j=1}^{r} \frac{(\nu_j-np_j)^2}{np_j} \tag{6.4.11}$$

的值. 在 $H_0$ 成立时, 当 $n \to \infty$, 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 $\alpha$ 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 $a_1,\ldots,a_{r-1}$ 的选择应使理论频数 $n_{pi}$ 和观察频数 $v_i(i=1,2,\ldots,r)$ 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 $a_1,\ldots,a_r$ 必须不依赖于样本. 就是说不能根据样本 $X_1,\ldots,X_n$ 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 $\chi^2$ 检验: 理论分布带有未知参数的情况

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 $\{F(x;\theta_1,\cdots,\theta_s):(\theta_1,\cdots,\theta_s)\in \Theta\}$, 其中 $\Theta$ 为参数空间. 令 $X_1,\cdots,X_n$ 为自总体 $X$ 中抽取的简单样本, 要检验假设

$$H_0: \text{存在}(\theta_1^0,\cdots,\theta_s^0)\in\Theta, \text{使}X\text{的分布为}F(x;\theta_1^0,\cdots,\theta_s^0). \tag{6.4.12}$$

例如, 有一批数据 $X_1,\cdots,X_n$, 要检验它是否来自 $N(a,\sigma^2)$, 其中 $a$,$\sigma^2$ 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 $I_1,\cdots,I_r$ 如式(6.4.8)所示, 改式(6.4.9)为下列形式:

$$\begin{aligned} p_j(\theta_1,\cdots,\theta_s)=P(X\in I_j)=F(a_j;\theta_1,\cdots,\theta_s)-F(a_{j-1};\theta_1,\cdots,\theta_s),& \\ j=1,2,\cdots,r, a_0=-\infty, a_r=+\infty.& \end{aligned} \tag{6.4.13}$$

因此检验问题(6.4.12)转化为

$$\begin{aligned} H_0: \text{存在}\left(\theta_1^0,\cdots,\theta_s^0\right)\in\Theta, \text{使得}P(X\in I_j)=p_j\left(\theta_1^0,\cdots,\theta_s^0\right),& \\ j=1,2,\cdots,r,& \end{aligned} \tag{6.4.14}$$

其中对 $p_j\left(\theta_1,\cdots,\theta_s\right)$ 作如下假定:

(1) 对任何 $(\theta_1,\cdots,\theta_s)\in\Theta, p_j=p_j\left(\theta_1,\cdots,\theta_s\right)>0, j=1,\cdots,r$ 且 $\sum\limits_{j=1}^{r} p_j=1$;

(2) 对任何 $1\leqslant j\leqslant r$,$p_j\left(\theta_1,\cdots,\theta_s\right)$ 对 $\theta_1,\cdots,\theta_s$ 有一阶连续偏导数.

算出样本 $X_1,\cdots,X_n$ 落入 $I_j$ 的观察频数 $\nu_j$, 求出类似于式(6.4.11)的表达式

$$K_n\left(\theta_1,\cdots,\theta_s\right)=\sum_{j=1}^{r}\frac{\left(\nu_j-np_j\left(\theta_1,\cdots,\theta_s\right)\right)^2}{np_j\left(\theta_1,\cdots,\theta_s\right)}. \tag{6.4.15}$$

由于 $\theta_1,\cdots,\theta_s$ 未知, 故式(6.4.15)中 $K_n(\theta_1,\cdots,\theta_s)$ 还不能作为检验统计量. 因此要按某种估计方法, 将 $\theta_1,\cdots,\theta_s$ 用样本 $X_1,\cdots,X_n$ 估出. 若令 $\hat{\theta}_j=\hat{\theta}_j(X_1,\cdots,X_n)$ 为 $\theta_j$ 的估计, 以之取代式(6.4.15)中的 $\theta_j,j=1,2,\cdots,s$, 则检验问题(6.4.12)可表为 $H_0$:r.v.$X$ 的分布为 $F(x;\hat{\theta}_1,\cdots,\hat{\theta}_s)$, 得检验统计量

$$K_n^*=K_n^*(\hat{\theta}_1,\cdots,\hat{\theta}_s)=\sum_{j=1}^{r}\frac{\left(\nu_j-np_j(\hat{\theta}_1,\cdots,\hat{\theta}_s)\right)^2}{np_j(\hat{\theta}_1,\cdots,\hat{\theta}_s)}. \tag{6.4.16}$$

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 $n\to\infty$ 时,$K_n^*$ 的分布收敛于 $\chi^2_{r-1}$.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 $\theta_j(j=1,\cdots,s)$ 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 $X_1,\cdots,X_n$, 用极大似然方法对 $\theta_1,\cdots,\theta_s$ 作一估计. 如前, 设 $\nu_j$ 为样本 $X_1,\cdots,X_n$ 中落入 $I_j$ 的个数,$j=1,2,\cdots,r$, 则似然函数为

$$L=\frac{n!}{\nu_1!\cdots\nu_r!}\left(p_1\left(\theta_1,\cdots,\theta_s\right)\right)^{\nu_1}\cdots\left(p_r\left(\theta_1,\cdots,\theta_s\right)\right)^{\nu_r},$$

对 $\log L$ 关于 $\theta_i$ 求导, 得方程组

$$\sum_{j=1}^{r}\frac{\nu_j}{p_j\left(\theta_1,\cdots,\theta_s\right)}\cdot\frac{\partial p_j\left(\theta_1,\cdots,\theta_s\right)}{\partial\theta_i}=0,\quad i=1,2,\cdots,s. \tag{6.4.17}$$

解这一方程组得

$$\hat{\theta}_i=\hat{\theta}_i\left(X_1,\cdots,X_n\right),\quad i=1,2,\cdots,s. \tag{6.4.18}$$

按式(6.4.16)算出检验统计量 $K_n^*$.R.A.Fisher 证明了如下定理.

**Theorem 6.4.2** (R.A. Fisher)**.** 设 $\theta_1^0,\cdots,\theta_s^0$ 为参数空间 $\Theta$ 的内点, 使得 $X$ 的分布为 $F(x,\theta_1^0,\cdots,\theta_s^0)$. $(\hat{\theta}_1,\cdots,\hat{\theta}_s)$ 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 $n\to\infty$ 时有

$$K_n^*\xrightarrow{\mathscr{L}}X^2_{r-1-s}$$

证明略.

利用上述定理可得检验问题(6.4.12)的水平近似为 $\alpha$ 的大样本检验如下: 若

$$\text{若}K_n^*>\chi^2_{r-1-s}(\alpha)\text{则否定}H_0,\text{否则接受}H_0. \tag{6.4.19}$$

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

$$p\left(k_0^*\right)=P\left(K_n^*>k_0^*|H_0\right)\approx P\left(\chi^2_{r-1-s}>k_0^*\right) \tag{6.4.20}$$

其中 $\chi^2_{r-1-s}$ 表示自由度为 $r-1-s$ 的 $\chi^2$ 变量. 对给定的检验水平 $\alpha$, 当 $p(k_0^*)<\alpha$ 时否定 $H_0$, 当 $p(k_0^*)\geqslant\alpha$ 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.

**Note 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 $F(x;\theta_1,\cdots,\theta_s)$, 直接由样本 $X_1,\cdots,X_n$(而不是由 $\nu_1,\cdots,\nu_r$) 求出参数 $\theta_1,\cdots,\theta_s$ 的 MLE. 例如,$F$ 为正态分布 $N(a,\sigma^2)$, 参数 $a$ 和 $\sigma^2$ 的 MLE 为 $\overline{X}$ 和 $S_n^2=\frac{\sum\limits_{i=1}^n(X_i-\overline{X})^2}{n}$, 用 $\hat{\theta}_1^*=\overline{X},\hat{\theta}_2^*=S_n^2$ 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 $\theta_1,\cdots,\theta_s$ 的基于样本 $X_1,\cdots,X_n$ 的 MLE,$\hat{\theta}_1^*,\cdots,\hat{\theta}_s^*$, 代替式(6.4.15)中的 $\theta_1,\cdots,\theta_s$, 而不用式(6.4.17)的解 $\hat{\theta}_1,\cdots,\hat{\theta}_s$. 记 $\hat{p}_j^*=\hat{p}_j^*(\hat{\theta}_1^*,\cdots,\hat{\theta}_s^*)$, 则所算出的检验统计量为

$$\widetilde{K}_n^*=\sum_{j=1}^r\frac{(v_j-n\hat{p}_j^*)^2}{n\hat{p}_j^*}.$$

此检验统计量 $\widetilde{K}_n^*$ 不一定有极限分布 $\chi^2_{r-1-s}$. 更确切地说, 可以证明: 当 $n$ 充分大时, 真正的拟合优度值 $p(\tilde{k}_0^*)=P(\widetilde{K}_n^*>\tilde{k}_0^*|H_0)$ 介于 $P(\chi^2_{r-1-s}\geqslant\tilde{k}_0^*)$ 和 $P(\chi^2_{r-s-1}\geqslant\tilde{k}_0^*)$ 之间 (后者较大). 按定理6.4.1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 $(\hat{\theta}_1^*,\cdots,\hat{\theta}_s^*)$ 而不用式(6.4.17)的解 $(\hat{\theta}_1,\cdots,\hat{\theta}_s)$, 相当于把失掉的自由度 $s$ 挽回一部分. 问题在于 $\widetilde{K}_n^*$ 的极限分布不再是 $\chi^2_{r-s-1}$ 分布了.

在下面的例子中, 用极大似然估计 $(\hat{\theta}_1^*,\cdots,\hat{\theta}_s^*)$ 代替式(6.4.17)的解 $(\hat{\theta}_1,\cdots,\hat{\theta}_s)$, 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.

**Example 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

$$H_0:\text{存在}\lambda_0\text{使得}X\sim\text{Poisson 分布}P\left(\lambda_0\right). \tag{6.4.21}$$

(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 (实测) | 263 | 112 | 38 | 30 |

(2) 求未知参数的 MLE.Poisson 分布中参数 $\lambda$ 的 MLE 为

$$\hat{\lambda}^* = \overline{X} = \frac{308}{443} = 0.7,$$

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 $n = 443$,

$$\begin{aligned}
n\hat{p}_0^* &= nP_{\lambda^*}(X=0) = 443 \times \frac{(0.7)^0 e^{-0.7}}{0!} = 443 \times 0.496585 = 219.99, \\
n\hat{p}_1^* &= 443 \times \frac{(0.7) \times e^{-0.7}}{1!} = 153.99, \\
n\hat{p}_2^* &= 443 \times \frac{(0.7)^2 \times e^{-0.7}}{2!} = 53.90, \\
n\hat{p}_3^* &= 443 \times P(3 \leqslant X \leqslant 8) = 443 \cdot \sum_{i=3}^{8} \frac{(0.7)^i e^{-0.7}}{i!} = 15.13.
\end{aligned}$$

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 $r - s - 1 = 2$, 故拟合优度为

$$p(\tilde{k}_0^*) = P\left(\tilde{K}_n^* > \tilde{k}_0^* | H_0\right) \approx P\left(\chi_2^2 > 39.16\right) < 0.005 < \alpha,$$

故否定 $H_0$, 即认为锭子的断头数不能服从 Poisson 分布.

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.

**Example 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 $\nu_i$ | 10 | 26 | 56 | 64 | 30 | 14 |

设 $X_1, \cdots, X_{200}$ 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

$$H_0:\text{存在}\mu_0,\sigma_0^2\text{使得}X\text{服从正态分布}N(\mu_0,\sigma_0^2). \tag{6.4.22}$$

此题不需要重新分组. 取每个区间的中点作为代表值, 则 $\mu$ 和 $\sigma^2$ 的 MLE 为

$$\begin{aligned}\hat{\mu}^* =&\overline{X}=\frac{1}{200}(195\times 10+205\times 26+215\times 56+225\times 64\\ &+235\times 30+245\times 14)=221\left(\mathrm{kg/cm^2}\right),\\ \hat{\sigma}^2 =&\frac{1}{200}\sum_{i=1}^{200}\left(X_i-\overline{X}\right)^2=\frac{1}{200}\left[(195-221)^2\times 10+(205-221)^2\times 26\right.\\ &\left.+\cdots+(245-221)^2\times 14\right]=12.33^2,\end{aligned}$$

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 $\widetilde{K}_n^*$, 先计算 r.v.$X$ 在每个区间中的概率. 区间的个数 $r=6$,

$$\hat{p}_i^*=P\left(a_{i-1}\leqslant X\leqslant a_i\right)=P\left(u_{i-1}\leqslant U\leqslant u_i\right)=\Phi\left(u_i\right)-\Phi\left(u_{i-1}\right),$$

此处 $i=1,\cdots,6,U=(X-221)/12.33$, 显然当 $H_0$ 成立时 $U\sim N(0,1)$;

$$\Phi(u_i)=\int_{-\infty}^{u_i}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t,$$

其中 $u_i=(a_i-221)/12.33,a_0=-\infty,a_6=+\infty$, 故有

$$\begin{aligned}\hat{p}_1^* =&P(-\infty<X<200)=P\left(-\infty<U<\frac{200-221}{12.33}\right)\\ =&P(U<-1.70)=0.045,\\ \hat{p}_2^* =&P(200<X<210)=P(-1.70<U<-0.89)=0.142,\end{aligned}$$

类似可求出 $\hat{p}_3^*,\cdots,\hat{p}_6^*$, 可按表6.4.4计算 $K_n^*$ 的值.

**表 6.4.4:**$\chi^2$ **检验统计量的计算**

| $i$ | 区间 | $\nu_i$ | $\hat{p}_i^*$ | $n\hat{p}_i^*$ | $(\nu_i-n\hat{p}_i^*)^2$ | $\dfrac{(\nu_i-n\hat{p}_i^*)^2}{n\hat{p}_i^*}$ |
|---|---|---|---|---|---|---|
| 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 $\widetilde{K}_n^*$ 的具体值 $\widetilde{k}_0^*=1.33$, 拟合优度为

$$p\left(\widetilde{k}_0^*\right)=P\left(\widetilde{K}_n^*>\widetilde{k}_0^*|H_0\right)\approx P\left(\chi_3^2>1.33\right)>0.70>\alpha.$$

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

略.

## 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—$\chi^2$ 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 / A | 患肺癌 | 未患肺癌 |
|---|---|---|
| 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 $2\times 2$ **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 $A_1,\cdots,A_r$ 和 $B_1,\cdots,B_s$ (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);$A_2$: 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 $(A_{ri},B_{si})$,$i=1,2,\cdots,n$, 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 $X=(X^{(1)},X^{(2)})$,$X^{(1)}$ 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,$X_i=(r_i,s_i)$,$i=1,2,\cdots,n$. 将 $n$ 个观察结果列表成6.5.2.

**表 6.5.2**

| $X^{(1)} \backslash X^{(2)}$ | $1$ | $\cdots$ | $j$ | $\cdots$ | $s$ | $\sum$ |
|---|---|---|---|---|---|---|
| $1$ | $n_{11}$ | $\cdots$ | $n_{1j}$ | $\cdots$ | $n_{1s}$ | $n_{1\cdot}$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ |
| $i$ | $n_{i1}$ | $\cdots$ | $n_{ij}$ | $\cdots$ | $n_{is}$ | $n_{i\cdot}$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ |
| $r$ | $n_{r1}$ | $\cdots$ | $n_{rj}$ | $\cdots$ | $n_{rs}$ | $n_{r\cdot}$ |
| $\sum$ | $n_{\cdot 1}$ | $\cdots$ | $n_{\cdot j}$ | $\cdots$ | $n_{\cdot s}$ | $n$ |

以 $n_{ij}$ 记 $X_1, \cdots, X_n$ 中取 $(i,j)$ 为值的个数. 这表称为 $r \times s$ **列联表**. 上表中 $n_{i\cdot} = \sum\limits_j n_{ij}$,$n_{\cdot j} = \sum\limits_i n_{ij}$, $n = \sum\limits_i n_{ik} = \sum\limits_j n_{jk}$, 要检验

$$H_0 : X^{(1)}\text{和}X^{(2)}\text{独立}. \tag{6.5.1}$$

**2. 检验方法**

若记

$$P\left(X^{(1)} = i, X^{(2)} = j\right) = p_{ij}, \quad i = 1, \cdots, r; \quad j = 1, \cdots, s,$$

满足 $\sum\limits_i \sum\limits_j p_{ij} = 1$. 因此如果 $X^{(1)}$ 和 $X^{(2)}$ 独立, 则对一切 $(i,j)$ 有

$$P\left(X^{(1)} = i, X^{(2)} = j\right) = P\left(X^{(1)} = i\right) P\left(X^{(2)} = j\right) = p_i \cdot p_j,$$

其中 $p_i = \sum\limits_j p_{ij}$,$p_j = \sum\limits_i p_{ij}$ 且有 $\sum\limits_i p_i = 1$,$\sum\limits_j p_j = 1$. 因此检验问题(6.5.1)转化为

$$H_0 : P\left(X^{(1)} = i, X^{(2)} = j\right) = p_i \cdot p_j, \quad i = 1, \cdots, r; \quad j = 1, \cdots, s, \tag{6.5.2}$$

满足限制条件

$$\sum_i p_i = 1, \quad \sum_j p_j = 1. \tag{6.5.3}$$

将 $p_{\cdot i}$ 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 $p_i$ 和 $p_j$ 的 MLE 如下:

$$\hat{p}_{i\cdot}^*=\frac{n_{i\cdot}}{n},\quad i=1,2,\cdots,r;\quad \hat{p}_{\cdot j}^*=\frac{n_{\cdot j}}{n},\quad j=1,2,\cdots,s.$$

算出 $\chi^2$ 统计量的值

$$\begin{aligned}K_n^*=&\sum_{i=1}^r\sum_{j=1}^s\frac{\left(n_{ij}-n\hat{p}_{i\cdot}^*\hat{p}_{\cdot j}^*\right)^2}{n\hat{p}_i^*\hat{p}_j^*}=n\sum_{i=1}^r\sum_{j=1}^s\frac{\left(n_{ij}-n_{i\cdot}n_{\cdot j}/n\right)^2}{n_{i\cdot}n_{\cdot j}}\\=&n\left(\sum_{i=1}^r\sum_{j=1}^s\frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}}-1\right).\end{aligned}\tag{6.5.4}$$

由定理6.4.2可知, 当 $H_0$ 成立且 $n\to\infty$ 时有

$$K_n^*\xrightarrow{\mathscr{L}}\chi^2_{rs-1-(r+s-2)}=\chi^2_{(r-1)(s-1)}.$$

对检验水平 $\alpha$, 查表求出 $\chi^2_{(r-1)(s-1)}(\alpha)$. 得水平近似为 $\alpha$ 的检验如下:

$$\text{若}K_n^*>\chi^2_{(r-1)(s-1)}(\alpha)\text{时, 否定}H_0\text{, 否则就接受}H_0.\tag{6.5.5}$$

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

$$p\left(k_0^*\right)=P\left(K_n^*\geqslant k_0^*\mid H_0\right)\approx P\left(\chi^2_{(r-1)(s-1)}\geqslant k_0^*\right)$$

对给定的检验水平 $\alpha$, 当 $p\left(k_0^*\right)<\alpha$ 时否定 $H_0$, 当 $p\left(k_0^*\right)\geqslant\alpha$ 时接受 $H_0$,$p\left(k_0^*\right)$ 越大, 则接受 $H_0$ 的结论越可靠.

**Note 6.5.1.** 在 $r=s=2$ 这个特例情形, 即 $2\times2$ 列联表情形, 简单的代数计算证明, 式(6.5.4)可简化为

$$K_n^*=\frac{n\left(n_{11}n_{22}-n_{12}n_{21}\right)^2}{n_{1\cdot}n_{2\cdot}n_{\cdot1}n_{\cdot2}}.\tag{6.5.6}$$

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.

**Example 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

$$H_0: 色盲和性别独立.$$

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

$$K_n^*=\frac{1000(442\times 6-514\times 38)^2}{956\times 44\times 480\times 520}=27.14>\chi_1^2(0.01)=6.64,$$

检验的拟合优度为

$$p\left(k_0^*\right)=P\left(K_n^*\geqslant k_0^*\mid H_0\right)\approx P\left(\chi_1^2\geqslant 27.14\right)<0.005.$$

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 $p_i(j), j=1,2,\cdots,s; i=1,2,\cdots,r$."$r$ 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

$$H_0: p_1(j)=p_2(j)=\cdots=p_r(j), \quad j=1,2,\cdots,s. \tag{6.5.7}$$

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,$j=1,2,\cdots,s;i=1,2,\cdots,r$. 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 $X^{(1)},\cdots,X^{(r)}$, 它们可能的取值都相同, 为 $a_1,\cdots,a_s$, 且第 $i$ 个总体 $X^{(i)}$ 取值为 $a_j$ 的概率记为 $p_i(j), j=1,2,\cdots,s; i=1,2,\cdots,r$, 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,$j=1,\cdots,s;i=1,\cdots,r$, 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**

| $X \backslash a$ | 1 | $\cdots$ | $j$ | $\cdots$ | $s$ | $\sum$ |
|---|---|---|---|---|---|---|
| 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

$$p_1(j) = \cdots = p_r(j) = p_j^0, \quad j = 1, 2, \cdots, s$$

而 $p_1^0, \cdots, p_s^0$ 皆已知且满足 $\sum\limits_{j=1}^{s} p_j^0 = 1$, 此时

$$K = K_n = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(n_{ij} - n_{i\cdot}p_j^0\right)^2}{n_{i\cdot}p_j^0},$$

则由定理6.4.1可知, 当 $H_0$ 成立且 $n_{i\cdot} \to \infty (i = 1, \cdots, r)$ 时有

$$K_n \xrightarrow{\mathscr{L}} \chi^2_{(s-1)r}.$$

检验问题(6.5.7)的水平近似为 $\alpha$ 的检验为

$$\text{当} K > \chi^2_{(s-1)r}(\alpha) \text{否定} H_0, \text{否则接受} H_0. \tag{6.5.8}$$

也可计算检验的拟合优度为 $p(k_0) = P(K \geqslant k_0 | H_0) \approx P(\chi^2_{(s-1)r} \geqslant k_0)$, 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

$$p_1(j) = \cdots = p_r(j) = p_j, \quad j = 1, 2, \cdots, s,$$

而 $p_1, \cdots, p_s$ 皆未知. 这时首先对 $p_1, \cdots, p_s$ 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 $\dfrac{\partial \log L}{\partial p_j} = 0, \quad j = 1, 2, \cdots, s$, 类似于例3.3.8可解得

$$\hat{p}_j^* = \frac{n_{\cdot j}}{n}, \quad j = 1, 2, \cdots, s,$$

算出

$$\hat{K}_{ni}^* = \sum_{j=1}^{s} \frac{(n_{ij} - n_{i\cdot}\hat{p}_j^*)^2}{n_i \hat{p}_j^*} = n \sum_{j=1}^{s} \frac{(n_{ij} - n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}}, \quad i = 1, \cdots, r,$$

故总的 $\hat{K}_n^*$ 为

$$\begin{aligned}\hat{K}^* =& \hat{K}_n^* = \sum_{i=1}^{r} \hat{K}_{ni}^* = n \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}} \\ =& n \left( \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} - 1 \right). \end{aligned} \tag{6.5.9}$$

在齐一性问题中, 第 $i$ 个总体中抽取的样本 $n_{i\cdot}$ 是事先给定的, 它没有随机性. 而在独立性问题中 $n_{i\cdot}$ 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, 证明了在齐一性问题中若假设(6.5.7)成立, 则当 $n_{i\cdot} \to \infty (i = 1, 2, \cdots, r)$ 时, 则 $\hat{K}_n^*$ 的极限分布仍为 $\chi^2_{(r-1)(s-1)}$. 故知检验问题(6.5.7)的水平近似为 $\alpha$ 检验为

$$\text{当}\hat{K}^* > \chi^2_{(r-1)(s-1)}(\alpha)\text{时否定}H_0, \text{否则就接受}H_0. \tag{6.5.10}$$

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

$$p(\hat{k}^*) = P(\hat{K}^* \geqslant \hat{k}^* | H_0) \approx P(\chi^2_{(r-1)(s-1)} \geqslant \hat{k}^*)$$

对给定的检验水平 $\alpha$, 当 $p(\hat{k}^*) < \alpha$ 时否定 $H_0$, 当 $p(\hat{k}^*) \geqslant \alpha$ 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,$\chi^2$ 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.

**Example 6.5.2.** 设一工厂分三班生产. 产品分成 $1, 2, \cdots, 6$ 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 $(r-1)(s-1)=2\times5=10$, 查表得 $\chi^2_{10}(0.01)=23.21$. 按式(6.5.9)计算检验统计量

$$\begin{aligned}\hat{K}^* =&261\left[\left(\frac{11^2}{34\times83}+\frac{23^2}{73\times83}+\cdots+\frac{9^2}{42\times83}\right)-1\right]\\ =&26.3>\chi^2_{10}(0.01)=23.21,\end{aligned}$$

检验的拟合优度为

$$p(\hat{k}^*)=P\left(\hat{K}^*\geqslant\hat{k}^*|H_0\right)\approx P\left(\chi^2_{10}\geqslant26.3\right)<0.01.$$

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 $\chi^2$ 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 $\chi^2$ 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 $\chi^2$ 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 $\chi^2$ 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,$X_1,\cdots,X_n$ 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:

$$H_0: F(x)=F_0(x). \tag{6.6.1}$$

首先, 从样本出发求出 $F(x)$ 的经验分布函数如下:

$$F_n(x)=\begin{cases}0, & x\leqslant X_{(1)},\\ \dfrac{k}{n}, & X_{(k)}<x\leqslant X_{(k+1)}, k=1,2,\cdots,n-1,\\ 1, & x>X_{(n)},\end{cases} \tag{6.6.2}$$

其中 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 是样本 $X_1,\cdots,X_n$ 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

$$D_n=\sup_{-\infty<x<+\infty}|F_n(x)-F_0(x)|, \tag{6.6.3}$$

$D_n$ 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 $P\left(\lim\limits_{n\to\infty}D_n=0\right)=1$. 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 $D_n\geqslant c$ 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

$$p(D_0)=P(D_n>D_0|H_0) \tag{6.6.4}$$

就是在柯氏距离下, 样本 $X_1,\ldots,X_n$ 与理论分布 $F_0(x)$ 的拟合优度. 若指定一个阈值 $\alpha$(亦称检验水平), 则需定出一个常数 $D_{n,\alpha}$, 使得

$$p(D_{n,\alpha})=P(D_n>D_{n,\alpha}|H_0)=\alpha, \tag{6.6.5}$$

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.

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.

**Theorem 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

$$\lim_{n\to\infty}P\left(D_n\leqslant\frac{\lambda}{\sqrt{n}}\right)=K(\lambda)=\begin{cases}\sum\limits_{k=-\infty}^{\infty}(-1)^k e^{-2k^2\lambda^2}, & \lambda>0,\\ 0, & \lambda\leqslant 0.\end{cases} \tag{6.6.6}$$

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 $\alpha$ 时查表获得 $\lambda$, 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 $K(\lambda)=1-\alpha=0.95$ 查表得到 $\lambda=1.358$, 故 $D_{n,\alpha}=1.358/\sqrt{n}$; 若 $\alpha=0.01$, 由 $K(\lambda)=1-\alpha=0.99$, 从表中查出 $\lambda=1.628$, 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 $x_1,\cdots,x_n$ 按大小排列为 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$;

(2) 计算 $D_n$ 的值

$$D_n=\max_i\{\delta_i, i=1,2,\cdots,n\},$$

其中

$$\delta_i=\max\left\{\left|F_0\left(x_{(i)}\right)-\frac{i-1}{n}\right|,\left|F_0\left(x_{(i)}\right)-\frac{i}{n}\right|\right\},\quad i=1,2,\cdots,n; \tag{6.6.7}$$

(3) 给出检验水平 $\alpha$, 在 $n\leqslant 100$ 和 $n>100$ 时查不同的表, 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 $D_n\leqslant D_{n,\alpha}$ 则不足以否定 $H_0$.

需要说明的一点是 $\delta_i$ 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 $\delta_i(i=1,2,\cdots,n)$ 可由式(6.6.7)获得.

**Example 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 $D_n=\max\limits_i\{\delta_i,i=1,2,\cdots,10\}=0.166$. 由 $n=10,\alpha=0.10$, 查表得 $D_{10,0.10}=0.37>0.166=D_n$, 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 $\chi^2$ 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 $\chi^2$ 检验. 这是因为:① Pearson $\chi^2$ 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 $\sup\limits_i|F_n-F_0|$ 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: 科尔莫戈洛夫检验的计算

| $i$ | $x_{(i)}$ | $F_0\left(x_{(i)}\right)$ | $(i-1)/n$ | $i/n$ | $\delta_i$ |
|---|---|---|---|---|---|
| 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 $\chi^2$ 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,$\chi^2$ 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 $X_{i1},\cdots,X_{in_i}$ 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

$$H_0: F_1(x)=F_2(x),\quad -\infty<x<+\infty. \tag{6.6.8}$$

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

$$D^+_{n_1,n_2}=\sup_{-\infty<x<+\infty}\left(F_{1n_1}(x)-F_{2n_2}(x)\right),$$
$$D_{n_1,n_2}=\sup_{-\infty<x<+\infty}\left|F_{1n_1}(x)-F_{2n_2}(x)\right|.$$

苏联数学家斯米尔诺夫 (N.Smirnov) 于 1936 年证明了下列结果.

**Theorem 6.6.2.** 设原假设(6.6.8)成立, 则有

$$\lim_{\substack{n_1\to\infty\\ n_2\to\infty}} P\left(\sqrt{\frac{n_1n_2}{n_1+n_2}}D^+_{n_1,n_2}\leqslant x\right)=\begin{cases}1-e^{-2x^2}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$$
$$\lim_{\substack{n_1\to\infty\\ n_2\to\infty}} P\left(\sqrt{\frac{n_1n_2}{n_1+n_2}}D_{n_1,n_2}\leqslant x\right)=K(x),$$

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 $\lambda$ 的值可由表查出. 这就是**斯米尔诺夫检验**.

若假设检验问题为

$$H_0 : F_1(x) \leqslant F_2(x) \leftrightarrow K : F_1(x) > F_2(x), \quad x \in (-\infty, \infty),$$

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 $\chi^2$ 检验, 柯氏检验等当然可以使用. 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.

### 1.$W$ 检验 (Wilk 检验)

考虑检验问题

$$H_0 : X\text{服从正态分布} \leftrightarrow H_1 : X\text{不服从正态分布} \tag{6.6.9}$$

设 $X_1, \cdots, X_n$ 为来自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本,$X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ 为其次序统计量. 设 $Y_i = (X_i - \mu)/\sigma, i = 1, \cdots, n$, 则 $Y_1, \cdots, Y_n$i.i.d.$\sim N(0,1)$. 令

$$Y_{(i)} = \frac{X_{(i)} - \mu}{\sigma}, \quad e_i = X_{(i)} - E\left(X_{(i)}\right),$$
$$m_i = E\left(Y_{(i)}\right), \quad i = 1, 2, \cdots, n.$$

注意 $m_1, \cdots, m_n$ 是与 $\mu, \sigma^2$ 无关的确定的数. 显然有

$$X_{(i)} = \mu + \sigma m_i + e_i, \quad i = 1, 2, \cdots, n. \tag{6.6.10}$$

其中 $\boldsymbol{e} = (e_1, \cdots, e_n)'$ 是均值为 0, 协方差阵为 $\boldsymbol{V}$ 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 $(X_{(1)}, m_1)$, $(X_{(2)}, m_2)$,$\cdots$,$(X_{(n)}, m_n)$ 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 $\boldsymbol{X} = \left(X_{(1)}, \cdots, X_{(n)}\right)'$ 和 $\boldsymbol{m} = (m_1, \cdots, m_n)'$ 之间的相关系数 $R$,

$$R^2 = \frac{\left(\sum\limits_{i=1}^{n} \left(X_{(i)} - \overline{X}\right)(m_i - \overline{m})\right)^2}{\sum\limits_{i=1}^{n} \left(X_{(i)} - \overline{X}\right)^2 \sum\limits_{i=1}^{n} (m_i - \overline{m})^2}.$$

显然 $0 \leqslant R^2 \leqslant 1$, 当 $R^2$ 越接近 1,$\boldsymbol{X}$ 与 $\boldsymbol{m}$ 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,$R^2$ 接近于 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 $(Y_{(1)},\cdots,Y_{(n)})$ 与 $(-Y_{(n)},\cdots,-Y_{(1)})$ 有相同的联合分布, 从而 $Y_{(k)}$ 与 $-Y_{(n+1-k)}$ 同分布, 故 $m_k = -m_{n+1-k}, k=1,\cdots,n, \overline{m} = \sum\limits_{i=1}^{n} m_i/n = 0$, 于是

$$R^2 = \frac{\left(\sum\limits_{i=1}^{n} m_i X_{(i)}\right)^2}{\sum\limits_{i=1}^{n}\left(X_{(i)}-\overline{X}\right)^2 \sum\limits_{i=1}^{n} m_i^2} = \frac{\left[\sum\limits_{i=1}^{[n/2]} b_i(X_{(n+1-i)}-X_{(i)})\right]^2}{\sum\limits_{i=1}^{n}(X_{(i)}-\overline{X})^2}, \tag{6.6.11}$$

其中 $b_i = m_{n+1-i}/\|\boldsymbol{m}\|$, 此处 $\|\boldsymbol{m}\|^2 = \sum\limits_{i=1}^{n} m_i^2$. 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)

$$W = \frac{\left\{\sum\limits_{i=1}^{[n/2]} a_i(X_{(n+1-i)}-X_{(i)})\right\}^2}{\sum\limits_{i=1}^{n}(X_{(i)}-\overline{X})^2}. \tag{6.6.12}$$

在 $n \leqslant 50$ 时,$\{a_i : i \leqslant [n/2]\}$ 的值已制成表. 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. 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 $0 \leqslant W \leqslant 1$. 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 $\alpha$ 后, 检验问题(6.6.9)的 $W$ 检验是

$$\text{当 } W \leqslant W_\alpha \text{ 时否定 } H, \text{否则接受 } H, \tag{6.6.13}$$

其中 $W$ 按式(6.6.12)计算, 临界值 $W_\alpha$ 可由表查出.

**Example 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 $a_i$ 这一列的值由表根据 $n=10$ 查得. 经计算得

$$\sum_{i=1}^{10} x_{(i)}=834, \quad \frac{1}{10}\sum_{i=1}^{10} x_{(i)}=\overline{x}=\frac{1}{10}834=83.4,$$

$$\sum_{i=1}^{10}\left(x_{(i)}-\overline{x}\right)^2=\sum_{i=1}^{10} x_{(i)}^2-10\overline{x}^2=71736-10\times 6955.56=2180.4,$$

$$\sum_{i=1}^{5} a_i\left(x_{(11-i)}-x_{(i)}\right)=46.494, \quad \left[\sum_{i=1}^{5} a_i\left(x_{(11-i)}-x_{(i)}\right)\right]^2=2161.692.$$

于是有

$$W=\frac{\left[\sum\limits_{i=1}^{5} a_i\left(x_{(11-i)}-x_{(i)}\right)\right]^2}{\sum\limits_{i=1}^{10}\left(x_{(i)}-\overline{x}\right)^2}=\frac{2161.7}{2180.4}=0.99.$$

由 $\alpha=0.05$,$n=10$, 查表得 $W_{0.05}=0.842<0.99=W$, 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 $3\leqslant n\leqslant 50$ 的样本. 当 $n>50$ 时很难计算表中的相应值. 为此人们提出了 $D$ 检验. 达戈斯蒂诺 (D'Agostino) 建议在 $n>50$ 时用

$$D=\sum_{i=1}^{n}\left(i-\frac{n+1}{2}\right)X_{(i)}\Bigg/\left[n^3\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right]^{1/2} \tag{6.6.14}$$

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

$$E(D)\approx 0.28209479, \sqrt{\operatorname{Var}(D)}\approx\frac{0.02998598}{\sqrt{n}},$$

其中 $\operatorname{Var}(D)$ 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 $n \to \infty$ 时有

$$Y \xrightarrow{\mathscr{L}} N(0,1).$$

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.

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 $\alpha$ 的 $D$ **检验是**

$$\text{当 } Y \leqslant Y_{\alpha/2} \text{ 或 } Y \geqslant Y_{1-\alpha/2} \text{ 时, 否定 } H_0\text{; 否则就接受 } H_0, \tag{6.6.15}$$

其中 $Y_{\alpha/2}$ 和 $Y_{1-\alpha/2}$ 分别是 $Y$ 的 $\alpha/2$ 和 $1-\alpha/2$ 分位数, 其值可从表查出.

**Note 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

$$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}.$$

(2) 将数对 $(x_{(i)}, i/(n+1))(i = 1, \cdots, n)$ 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.

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 图, 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.

# Chapter 7

# 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

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 $\chi^2$ 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. 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.

**Example 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.

**Example 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 $\theta$, 经过一段时间后获得大量数据, 对 $\theta$ 作出估计. 就当日被抽查的那批产品的废品率 $\theta$ 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 $\theta$ 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 $\theta$" 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

$$P\left(\theta = \frac{i}{n}\right) = \pi_i, \quad i = 0, 1, 2, \cdots, n.$$

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 先验分布和后验分布

**Definition 7.1.1** (先验分布). 参数空间 $\Theta$ 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 $\theta$ 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 $\theta$ 的新认识. 这是由于样本 $X$ 也包含 $\theta$ 的信息, 因此一旦获得样本 $X$ 后, 人们对 $\theta$ 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.

**Definition 7.1.2** (后验分布). 在获得样本 $X$ 后,$\theta$ 的**后验分布** (posterior distribution) 就是给定 $X=x$ 条件下 $\theta$ 的条件分布, 后验密度 $\pi(\theta|x)$ 有下列计算公式:

$$\pi(\theta|x)=\frac{f(x,\theta)}{m(x)}=\frac{f(x|\theta)\pi(\theta)}{\displaystyle\int_{\Theta} f(x|\theta)\pi(\theta)\mathrm{d}\theta}. \tag{7.1.1}$$

(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 $\theta$ 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 $\theta$ 的任何统计推断 (估计, 检验等) 必须基于且只能基于 $\theta$ 的后验分布.

### 7.1.3 古典学派和 Bayes 学派的争论

略.

### 7.1.4 历史

略.

## 7.2 先验分布的确定

### 7.2.1 主观概率

#### 1. 引言和定义

**Definition 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.

#### 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 $(\overline{A})$ 的可能性的 2 倍即 $P(A)=2/3$,$P(\overline{A})=1/3$, 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

$$P(A_1)=\frac{29}{37}=0.784, P(A_2)=\frac{6}{37}=0.162, P(A_3)=\frac{2}{37}=0.054.$$

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 $\theta$ 属于离散型随机变量, 即参数空间 $\Theta$ 是由有限或可列个点构成时, 可对 $\Theta$ 中每个点确定一个主观概率, 这就是前面所介绍的.

当参数 $\theta$ 是连续随机变量时, 即 $\Theta$ 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 $\theta$ 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 $\theta$ 为实轴上的区间时, 先把 $\theta$ 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线梯形的面积为 1).

**Example 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 $\theta$ 的概率分布.

**解:** 利用直方图来确定 $\theta$ 的概率分布, 按下述步骤:

(1) 把 $\theta$ 的区间 $(0, 35)$ 分成 7 个小区间, 每个小区间长为 5 个单位 (kg);

(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.

**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 |

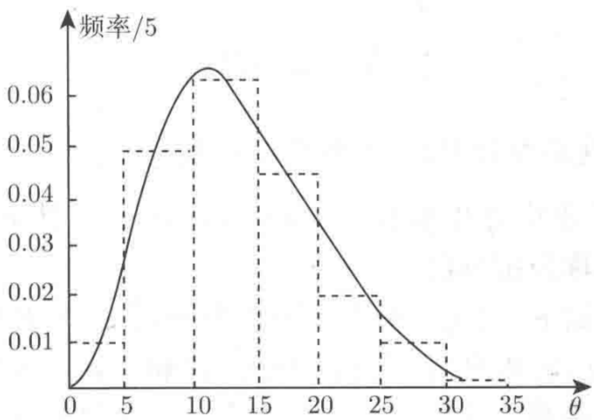



**图 7.2.1: 日销售量的直方图**

利用此直方图可以计算有关的概率. 例如,$P(20 \leqslant \theta \leqslant 21) = 1 \times \pi(20.5) = 0.03$.

**2. 相对似然法**

此法大多用于 $\Theta$ 为 $(-\infty, \infty)$ 的有限子区间的情形. 方法如下: 对 $\Theta$ 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 $\int_0^1 \tilde{\pi}(\theta)\mathrm{d}\theta \neq 1$. 记 $\pi(\theta)=c\tilde{\pi}(\theta)$, 使 $\int_0^1 \pi(\theta)\mathrm{d}\theta = c\int_0^1 \tilde{\pi}(\theta)\mathrm{d}\theta = 1$, 则 $\pi(\theta)$ 即为 $\theta$ 的先验密度.

**Note 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 $\Theta$ 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 $\Theta$ 为无限区间的情形.

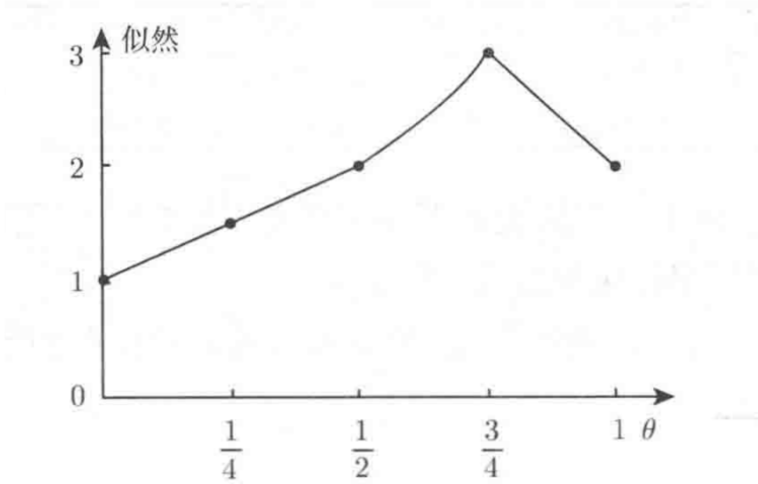



图 **7.2.2:** 似然

**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 $\mu$ 和 $\tau^2$ 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 $\alpha,\beta$ 为超参数. 对超参数为 $\alpha,\beta$ 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.

**Example 7.2.2.** 在例7.2.1中设参数 $\theta$ 为销售量, 选用正态分布 $N(\mu,\tau^2)$ 作为 $\theta$ 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 $\mu$ 和 $\tau^2$ 的问题. 这可用 "每日销售量统计表" 作出估计. 若对表中 $\theta$ 的每个小区间用其中点代替, 算得 $\mu$ 和 $\tau^2$ 的估计如下:

$$\begin{aligned}
\hat{\mu} =& 2.5\times 0.05 + 7.5\times 0.26 + \cdots + 32.5\times 0.01 = 13.45,\\
\hat{\tau}^2 =& (2.5-\hat{\mu})^2\times 0.05 + (7.5-\hat{\mu})^2\times 0.26 + \cdots\\
&+ (32.5-\hat{\mu})^2\times 0.01 = 36.85,
\end{aligned}$$

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

$$\begin{aligned}P(20 \leqslant \theta \leqslant 21) =& P\left(\frac{20-13.45}{\sqrt{36.85}} \leqslant \frac{\theta-13.45}{\sqrt{36.85}} \leqslant \frac{21-13.45}{\sqrt{36.85}}\right) \\ =& \Phi(1.24) - \Phi(1.08) = 0.8925 - 0.8508 = 0.0417.\end{aligned}$$

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.

**Example 7.2.3.** 设参数 $\theta$ 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;②0.25 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 $\mu$ 和 $\tau$ 的问题. 正态分布的中位数就是 $\mu$, 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 $\theta$ 的先验分布.

又若在本例中假定 $\theta$ 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 $\alpha$,$\beta$ 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 $\alpha$ 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 $\theta$ 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 $\Theta$ 中的任何一点 $\theta$ 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 $\Theta$ 为有限集, 即 $\theta$ 只可能取有限个值, 如 $\theta=\theta_i, i=1,2,\cdots,n$, 无信息先验给 $\theta$ 中的每个元素以概率 $1/n$, 即 $P(\theta=\theta_i)=1/n, i=1,2,\cdots,n$.

(2) 若 $\Theta$ 为 $\mathbb{R}$ 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 $\Theta$ 无界, 无信息先验如何选取? 例如, 样本分布为 $N(\theta,\sigma^2)$ 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 $\displaystyle\int_{\infty}^{\infty}\pi(\theta)\mathrm{d}\theta=\infty$. 这就引出广义先验分布的概念.

**Definition 7.2.2.** 设随机变量 $X\sim f(x|\theta),\theta\in\Theta$, 若 $\theta$ 的先验密度 $\pi(\theta)$ 满足下列条件:

(1) $\pi(\theta)\geqslant 0$ 且 $\displaystyle\int_{\Theta}\pi(\theta)\mathrm{d}\theta=\infty$;

(2) 后验密度 $\pi(\theta|x)$ 是正常的密度函数,

则称 $\pi(\theta)$ 为 $\theta$ 的**广义先验密度** (improper prior density).

**Note 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.

**Example 7.2.4.** 设 $X=(X_1,\cdots,X_n)$ 为从 $N(\theta,1)$ 总体中抽取的随机样本, 设 $\theta$ 的先验密度 $\pi(\theta)\equiv 1$, 求 $\theta$ 的后验密度.

**解:** 由式(7.1.1)可知

$$
\begin{aligned}
\pi(\theta|x)=&\frac{f(x|\theta)\pi(\theta)}{\displaystyle\int_{-\infty}^{\infty}f(x|\theta)\pi(\theta)\mathrm{d}\theta}=\frac{\exp\left\{-\frac{1}{2}\sum\limits_{i=1}^{n}(x_i-\theta)^2\right\}}{\displaystyle\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}\mathrm{d}\theta}\\
=&\sqrt{\frac{n}{2\pi}}\exp\left\{-\frac{n}{2}(\theta-\overline{x})^2\right\}.
\end{aligned}
$$

这是正态分布 $N(\overline{x},1/n)$ 的密度函数, 后验分布 $\pi(\theta|x)$ 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 $f(x-\theta),-\infty<\theta<\infty,\theta$ 为**位置参数** (locatior parameter). 对 $X$ 作平移变换得到 $Y=X+c$, 同时对 $\theta$ 也作平移变换得到 $\eta=\theta+c$. 显然 $Y$ 的密度函数有形式 $f(y-\eta),\eta$ 仍为位置参数. 所以 $(X,\theta)$ 与 $(Y,\eta)$ 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 $\theta$ 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 $\theta$ 为位置参数时, 其无信息先验密度取为常数或者 1.

**Example 7.2.5** (续例7.2.4). 设 $X = (X_1, \cdots, X_n)$ 为从总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, 其中 $\sigma^2$ 已知.$\theta$ 无任何先验信息可用, 求 $\theta$ 的后验分布.

**解:** 显然, 此时 $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ 为充分统计量且 $\overline{X} \sim N(\theta, \sigma^2/n)$, 即

$$p(\overline{x}|\theta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{n}{2\sigma^2}(\overline{x}-\theta)^2\right\}.$$

由于 $\theta$ 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 $N(\overline{x}, \frac{\sigma^2}{n})$ 的密度. 例如, 取 $\theta$ 的 Bayes 估计为后验均值, 则 Bayes 估计为 $\hat{\theta}_B = E(\theta|\overline{x}) = \overline{x}$. 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形式有 $\sigma^{-1}\varphi(x/\sigma)$,$\sigma > 0$ 为**刻度参数** (scale parameter). 对 $X$ 作变换 $Y = cX$, 同时对 $\sigma$ 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 $(X, \sigma)$ 和 $(Y, \eta)$ 统计问题的结构相同, 故主张 $\sigma$ 的无信息先验与 $\eta$ 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 $a$,$b$,$0 < a < b$,$c > 0$,$\sigma$ 落在 $[a, b]$ 内的先验概率, 应当等于 $\eta$ 落在 $[ca, cb]$ 内的先验概率. 不难看出, 这个事实只有在 $\sigma$ 的先验密度为

$$\pi(\sigma) = \frac{1}{\sigma}, 当\sigma > 0$$

时才能成立.

**Example 7.2.6.** 设总体 $X$ 为指数分布, 其密度为

$$f(x|\lambda) = \lambda^{-1} \exp\{-x/\lambda\}, x > 0,$$

其中 $\lambda > 0$ 为刻度参数. 令 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从上述分布中抽取的简单样本,$\lambda$ 的先验密度为无信息先验, 求其后验密度.

**解:** 记 $T=\sum\limits_{i=1}^{n}x_i$, 由式(7.1.1)可知 $\lambda$ 的后验密度为

$$\pi(\lambda|\boldsymbol{x})=\frac{\prod\limits_{i=1}^{n}f(x_i|\lambda)\pi(\lambda)}{\int_0^{\infty}\prod\limits_{i=1}^{n}f(x_i|\lambda)\pi(\lambda)\mathrm{d}\lambda}=\frac{\lambda^{-(n+1)}\exp\left\{-\frac{1}{\lambda}T\right\}}{\int_0^{\infty}\lambda^{-(n+1)}\exp\left\{-\frac{1}{\lambda}T\right\}\mathrm{d}\lambda}$$
$$=\frac{T^n}{\Gamma(n)}\lambda^{(n+1)}\exp\left\{-\frac{1}{\lambda}T\right\}.$$

若取其 Bayes 估计为后验均值, 则 Bayes 估计为

$$\hat{\lambda}_B=E(\lambda|\boldsymbol{x})=\frac{1}{n-1}\sum_{i=1}^{n}x_i,$$

其方差为 $\left(\sum\limits_{i=1}^{n}\right)\Big/[(n-1)^2(n-2)]$.

## *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 $\{f(x|\boldsymbol{\theta}),\boldsymbol{\theta}\in\mathit{\Theta}\}$ 满足 Cramer-Rao(C-R) 正则条件, 这里 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_p)$ 为 $p$ 维参数向量. 设 $X=(X_1,\cdots,X_n)$ 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

$$l(\theta|x)=\ln\left[\prod_{i=1}^{n}f(x_i|\boldsymbol{\theta})\right]=\sum_{i=1}^{n}\ln f(x_i|\boldsymbol{\theta}).$$

(2) 求样本的信息阵

$$\boldsymbol{I}(\boldsymbol{\theta})=(I_{ij}(\boldsymbol{\theta}))_{p\times p},\quad I_{ij}(\boldsymbol{\theta})=E_{\boldsymbol{X}|\boldsymbol{\theta}}\left\{-\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}\right\},\quad i,j=1,\cdots,p.$$

特别对 $p=1$, 即 $\theta$ 为单参数的情形

$$I(\theta)=E_{X|\theta}\left\{-\frac{\partial^2 l}{\partial\theta^2}\right\}.$$

(3) $\boldsymbol{\theta}$ 的无信息先验的密度为 $\pi(\boldsymbol{\theta})=[\det\boldsymbol{I}(\boldsymbol{\theta})]^{1/2}$, 其中 $\det\boldsymbol{I}(\boldsymbol{\theta})$ 表示 $p$ 阶方阵 $\boldsymbol{I}(\boldsymbol{\theta})$ 的行列式. 特别 $p=1$, 即单参数场合 $\pi(\theta)=[I(\theta)]^{1/2}$.

**Example 7.2.7.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从总体 $N(\mu,\sigma^2)$ 中抽取的简单样本, 记 $\theta=(\mu,\sigma)$, 求 $(\mu,\sigma)$ 的联合无信息先验.

**解:** 给定 $\boldsymbol{X}$ 时,$\theta$ 的对数似然函数是

$$l(\theta|\boldsymbol{x})=-\frac{n}{2}\log 2\pi-n\log\sigma-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

记 $\boldsymbol{I}(\theta)=(I_{ij}(\theta))_{p\times p}$, 则有

$$\begin{aligned}
I_{11}(\theta)=&E_{\boldsymbol{x}|\theta}\left\{-\frac{\partial^2 l(\theta|\boldsymbol{x})}{\partial\mu^2}\right\}=\frac{n}{\sigma^2},\\
I_{22}(\theta)=&E_{\boldsymbol{x}|\theta}\left\{-\frac{\partial^2 l(\theta|\boldsymbol{x})}{\partial\sigma^2}\right\}=-\frac{n}{\sigma^2}+\frac{3}{\sigma^4}E\left\{\sum_{i=1}^{n}(X_i-\mu)^2\right\}=\frac{2n}{\sigma^2},\\
I_{12}(\theta)=&I_{21}(\theta)=E_{\boldsymbol{x}|\theta}\left\{-\frac{\partial^2 l(\theta|\boldsymbol{x})}{\partial\mu\partial\sigma}\right\}=E\left\{\frac{2}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)\right\}=0.
\end{aligned}$$

故有

$$\boldsymbol{I}(\theta)=\begin{pmatrix}\frac{n}{\sigma^2}&0\\0&\frac{2n}{\sigma^2}\end{pmatrix},\quad[\det\boldsymbol{I}(\theta)]^{1/2}=\frac{\sqrt{2}n}{\sigma^2}.$$

所以,$(\mu,\sigma)$ 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 $(\mu,\sigma)$ 的联合无信息先验为 $1/\sigma^2$. 它的三个特例为

(1) 当 $\sigma$ 已知时,$I(\mu)=E\left\{-\frac{\partial^2 l(\theta|\boldsymbol{x})}{\partial\mu^2}\right\}=n/\sigma^2$, 故 $\pi_1(\mu)\equiv 1$;

(2) 当 $\mu$ 已知时,$I(\sigma)=E\left\{-\frac{\partial^2 l(\theta|\boldsymbol{x})}{\partial\sigma^2}\right\}=2n/\sigma^2$, 故 $\pi_2(\sigma)=1/\sigma$,$\sigma\in(0,\infty)$;

(3) 当 $\mu$ 和 $\sigma$ 独立时,$\pi(\mu,\sigma)=\pi_1(\mu)\pi_2(\sigma)=1/\sigma$,$\sigma\in(0,\infty)$.

由此可见, 当 $\mu$ 和 $\sigma$ 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 $\mu$ 和 $\sigma$ 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 $\mu$ 和 $\sigma$ 的联合无信息先验.

**Example 7.2.8.** 设 $\theta$ 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

$$P(X=x|\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x},\quad x=0,1,\cdots,n,$$

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

$$I(\theta)=E_{X|\theta}\left\{-\frac{\partial^2 l(\theta|X)}{\partial\theta^2}\right\}=E_{X|\theta}\left\{\frac{X}{\theta^2}+\frac{n-X}{(1-\theta)^2}\right\}=\frac{n}{\theta}+\frac{n}{1-\theta}=\frac{n}{\theta(1-\theta)}.$$

取 $\pi(\theta) \propto I(\theta)^{1/2}$ 得到正态化因子得到先验密度 $\pi(\theta)$, 它是一个 Beta 密度 $\mathrm{Be}(1/2, 1/2)$.

**Note 7.2.3.** 一般来说, 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.

**Definition 7.2.3.** 设 $\theta$ 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 $\pi(\theta) \in \mathscr{F}$ 及样本 $x$, 后验分布 $\pi(\theta|x)$ 仍属于 $\mathscr{F}$, 则称 $\mathscr{F}$ 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.

**Example 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 $\theta$ 的后验分布为 Beta 分布;

(2) 若取 $\theta$ 的先验分布为 Beta 分布 $\mathrm{Be}(a, b)$, 证明 $\theta$ 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 $\mathrm{Be}(1,1)$ 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 $\theta$ 的先验分布为 $\pi(\theta) = 1$, 当 $\theta \in (0,1)$, 故有

$$\pi(\theta|x) = \frac{\theta^x (1-\theta)^{n-x}}{\int_0^1 \theta^x (1-\theta)^{n-x} \mathrm{d}\theta}. \tag{7.2.1}$$

计算积分得到

$$\int_0^1 \theta^x (1-\theta)^{n-x} \mathrm{d}\theta = \frac{\Gamma(n-x+1)\Gamma(x+1)}{\Gamma(n+2)}.$$

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 $\theta$ 的后验分布是 Beta 分布 $\mathrm{Be}(x+1,n-x+1)$.

(2) 又若 $\theta\sim\mathrm{Be}(a,b)$, 则

$$\pi(\theta|x)=\frac{\theta^{x+a-1}(1-\theta)^{n-x+b-1}}{\displaystyle\int_0^1\theta^x(1-\theta)^{n-x}\theta^{a-1}(1-\theta)^{b-1}\mathrm{d}\theta}. \tag{7.2.2}$$

计算积分得到

$$\int_0^1\theta^{x+a-1}(1-\theta)^{n-x+b-1}\mathrm{d}\theta=\frac{\Gamma(x+a)\Gamma(n-x+b)}{\Gamma(n+a+b)}$$

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 $\theta$ 的后验分布是 Beta 分布 $\mathrm{Be}(x+a,n-x+b)$. 因此, 样本分布若为二项分布, 其参数 $\theta$ 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

$$\pi(\theta|x)=\frac{f(x|\theta)\pi(\theta)}{m(x)}=\frac{f(x|\theta)\pi(\theta)}{\int_\theta f(x|\theta)\pi(\theta)\mathrm{d}\theta}.$$

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 $\theta$ 的先验密度,$m(x)$ 是 $X$ 的边缘密度. 由于 $m(x)$ 与 $\theta$ 无关, 故可将 $1/m(x)$ 看成与 $\theta$ 无关的常数, 因此有

$$\pi(\theta|x)=\frac{f(x|\theta)\pi(\theta)}{m(x)}\propto f(x|\theta)\pi(\theta), \tag{7.2.3}$$

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 $\theta$ 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 $\theta$ 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 $\theta$ 有关的因子. 再写出先验密度 $\pi(\theta)$ 的核, 即 $\pi(\theta)$ 中仅与参数 $\theta$ 有关的因子.

(2) 类似式(7.2.3), 写出后验密度的核, 即

$$\pi(\theta|x)\propto f(x|\theta)\pi(\theta)\propto\{f(\theta|x)\text{的核}\}\cdot\{\pi(\theta)\text{的核}\}, \tag{7.2.4}$$

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.

**Note 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 $\theta$ 的先验分布为 $\mathrm{Be}(a, b)$. 求 $\theta$ 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

$$\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta) \propto \theta^{x+a-1}(1-\theta)^{n-x+b-1}.$$

易见上式的右边是"$\mathrm{Be}(x+a, n-x+b)$ 密度的核". 因此, 添加正则化因子得到后验密度

$$\pi(\theta \mid x) = \frac{\Gamma(n+a+b)}{\Gamma(x+a)\Gamma(n-x+b)}\theta^{(x+a)-1}(1-\theta)^{(n-x+b)-1}.$$

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.

**Example 7.2.10.** 设 $X \sim N(\theta, \sigma^2)$,$\sigma^2$ 已知而 $\theta$ 未知. 令 $\theta$ 的先验分布 $\pi(\theta)$ 是 $N(\mu, \tau^2)$, 其中 $\mu$ 和 $\tau^2$ 已知, 求 $\theta$ 的后验分布 $\pi(\theta|x)$.

**解:** 给定 $\theta$ 时 $X$ 的条件分布记为 $f(x|\theta)$, 令

$$A = \frac{1}{\sigma^2} + \frac{1}{\tau^2}, B = \frac{x}{\sigma^2} + \frac{\mu}{\tau^2}, C = \frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2}, \tag{7.2.5}$$

则有

$$\begin{aligned}
\pi(\theta|x) \propto & f(x|\theta)\pi(\theta) \propto \exp\left\{-\frac{1}{2}\left[\frac{(x-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2}\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[A\theta^2 - 2\theta B + C\right]\right\} \\
& = \exp\left\{-\frac{A}{2}\left(\theta - \frac{B}{A}\right)^2 - \frac{1}{2}\left(C - \frac{B^2}{A}\right)\right\} \\
& = \exp\left\{-\frac{1}{2\eta^2}(\theta - \mu(x))^2\right\} \cdot \exp\left\{-\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2+\tau^2}\right)\right\} \\
& \propto \exp\left\{-\frac{1}{2\eta^2}(\theta - \mu(x))^2\right\}
\end{aligned} \tag{7.2.6}$$

其中

$$\mu(x) = \frac{B}{A} = \frac{\sigma^2\mu + \tau^2 x}{\sigma^2 + \tau^2}, \quad \eta^2 = \frac{1}{A} = \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}, \quad C - \frac{B^2}{A} = \frac{(x-\mu)^2}{\sigma^2+\tau^2}.$$

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 $\theta$ 积分, 添加正则化常数得到 $X$ 的边缘密度

$$f(x)=\frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}}\exp\left\{-\frac{(\overline{x}-\mu)^2}{2(\sigma^2+\tau^2)}\right\}$$

若进一步假定 $X_1,\cdots,X_n$i.i.d.$\sim N(\theta,\sigma^2)$,$\sigma^2$ 已知,$\theta\sim N(\mu,\tau^2)$, 求 $\theta$ 的后验密度.

由于样本均值 $\overline{X}$ 为 $\theta$ 的充分统计量, 且给定 $\theta$ 时 $\overline{X}$ 的分布是 $N(\theta,\sigma^2/n)$, 故将上述结果中的 $x$ 用 $\overline{x}$ 代替,$\sigma^2$ 用 $\sigma^2/n$ 代替得到 $\theta$ 的后验分布

$$\pi(\theta|\overline{x})\sim N(\mu_n(\overline{x}),\eta_n^2),$$

其中

$$\mu_n(\overline{x})=\frac{\sigma^2/n}{\sigma^2/n+\tau^2}\mu+\frac{\tau^2}{\sigma^2/n+\tau^2}\overline{x},\eta_n^2=\frac{\tau^2\cdot\sigma^2/n}{\sigma^2/n+\tau^2}=\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}.$$

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 $\theta$ 的共轭先验分布族是正态分布族.

**Example 7.2.11.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从 Poisson 分布 $P(\lambda)$ 中抽取的简单样本,$\lambda$ 的先验分布为 Gamma 分布 $\Gamma(b,a)$. 证明给定 $\boldsymbol{X}=\boldsymbol{x}$ 时,$\lambda$ 的后验分布仍为 Gamma 分布.

**解:** $\lambda$ 的似然函数为

$$l(\lambda|\boldsymbol{x})=\frac{e^{-n\lambda}\cdot\lambda^{n\overline{x}}}{x_1!\cdots x_n!}\propto\lambda^{n\overline{x}}e^{-n\lambda},$$

其中 $\overline{x}=\sum\limits_{i=1}^{n}x_i/n$. 易见 $\lambda$ 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

$$\pi(\lambda|\boldsymbol{x})\propto l(\lambda|\boldsymbol{x})\pi(\lambda)\propto\lambda^{n\overline{x}+b-1}e^{-(n+a)\lambda}$$

上式右边是 $\Gamma(n\overline{x}+b,n+a)$ 密度的核. 添加正则化常数因子, 得到后验密度

$$\pi(\lambda|\boldsymbol{x})=\frac{(n+a)^{n\overline{x}+b}}{\Gamma(n\overline{x}+b)}\lambda^{n\overline{x}+b-1}e^{-(n+a)\lambda}.$$

此例表明当样本分布为参数 $\lambda$ 的 Poisson 分布时, 则 $\lambda$ 的共轭先验是 Gamma 分布族.

**Example 7.2.12.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从 Gamma 分布 $\Gamma(r,\lambda)$ 中抽取的简单样本, 其中 $r$ 已知. 假定 $\lambda$ 的先验分布为 Gamma 分布 $\Gamma(b,a)$, 证明给定 $\boldsymbol{X}$ 时,$\lambda$ 的后验分布仍为 Gamma 分布.

**解:** $\lambda$ 的似然函数为

$$l(\lambda|\boldsymbol{x}) = \frac{\lambda^{nr}}{(\Gamma(r))^n}\prod_{i=1}^{n} x_i^{r-1} \cdot \exp\left\{-\lambda n\overline{x}\right\} \propto \lambda^{nr}\exp\left\{-\lambda n\overline{x}\right\},$$

其中 $\overline{x} = \sum\limits_{i=1}^{n} x_i/n$. 而 $\lambda$ 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

$$\pi(\lambda|\boldsymbol{x}) \propto l(\lambda|\boldsymbol{x})\pi(\lambda) \propto \lambda^{nr+b-1}\exp\left\{-(a+n\overline{x})\lambda\right\}.$$

上式右边是 $\Gamma(nr+b, a+n\overline{x})$ 密度的核. 添加正则化常数因子, 得到后验密度

$$\pi(\lambda|\boldsymbol{x}) = \frac{(a+n\overline{x})^{nr+b}}{\Gamma(nr+b)}\lambda^{nr+b-1}\exp\left\{-(a+n\overline{x})\lambda\right\}.$$

此例表明当样本分布为 Gamma 分布 $\Gamma(r,\lambda)$, 其中 $r$ 已知时, 则 $\lambda$ 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 $\mathrm{Exp}(\lambda)$ 时,$\lambda$ 的共轭先验也是 Gamma 分布族.

**Definition 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 $X \sim \Gamma^{-1}(\alpha,\lambda)$.

**Example 7.2.13.** 设 $\boldsymbol{X} = (X_1,\cdots,X_n)$ 是从正态分布 $N(\mu,\sigma^2)$ 中抽取的简单样本, 其中 $\mu$ 已知, 求 $\sigma^2$ 的共轭先验分布.

**解:** $\sigma^2$ 的似然函数为

$$\begin{aligned} l(\sigma^2|\boldsymbol{x}) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\ &\propto (\sigma^2)^{-n/2}\exp\left\{-A/\sigma^2\right\}, \end{aligned}$$

其中 $A = \sum\limits_{i=1}^{n}(x_i-\mu)^2/2$, 可见似然函数 $l(\sigma^2|\boldsymbol{x})$ 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

$$\pi(\sigma^2) = \frac{\lambda^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-(\alpha+1)}e^{-\frac{\lambda}{\sigma^2}}, \quad \sigma^2 > 0,$$

则由式(7.2.4)可知

$$\pi(\sigma^2|\boldsymbol{x}) \propto l(\sigma^2|\boldsymbol{x})\pi(\sigma^2) \propto (\sigma^2)^{-(n/2+\alpha+1)} e^{-\frac{\lambda+A}{\sigma^2}}.$$

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

$$\pi(\sigma^2|\boldsymbol{x}) = \frac{(\lambda+A)^{n/2+\alpha}}{\Gamma(n/2+\alpha)} (\sigma^2)^{-(n/2+\alpha+1)} e^{-\frac{\lambda+A}{\sigma^2}}.$$

因此当样本分布为正态分布, 均值参数已知, 则 $\sigma^2$ 的共轭先验是逆 Gamma 分布族.

**Note 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 $N(\mu_n(\overline{x}), \eta_n^2)$, 其中

$$\mu_n(\overline{x}) = \frac{\tau^2}{\sigma^2/n+\tau^2}\overline{x} + \frac{\sigma^2/n}{\sigma^2/n+\tau^2}\mu = r\overline{x} + (1-r)\mu$$

是样本均值 $\overline{x}$ 和先验均值 $\mu$ 的加权平均. 若 $\sigma^2/n$ 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 $\overline{x}$ 决定, 反之若 $\sigma^2/n$ 很大 (即样本信息很少), 则后验均值主要由先验分布的均值 $\mu$ 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

### 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 $\theta$ 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.

**Definition 7.3.1.** 使后验密度 $\pi(\theta|x)$ 达到最大时 $\theta$ 的值称为后验分布的众数 (mode), 用后验分布的众数作为 $\theta$ 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 $\hat{\theta}_{\mathrm{MD}}$; 用后验分布的中位数 (median) 作为 $\theta$ 的估计, 称为 $\theta$ 的**后验中位数估计**, 记为 $\hat{\theta}_{\mathrm{ME}}$; 用后验分布的期望 (expectation) 作为 $\theta$ 的估计称为**后验期望估计** (或后验均值估计), 记为 $\hat{\theta}_{\mathrm{E}}$, 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 $\hat{\theta}_{\mathrm{B}}$ 来记.

**Note 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,$\theta$ 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.

**Example 7.3.1.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从正态总体 $N(\theta,\sigma^2)$ 中抽取的简单样本, 其中 $\sigma^2$ 已知. 取 $\theta$ 的先验为共轭先验分布 $N(\mu,\tau^2)$, 求 $\theta$ 的 Bayes 估计.

**解:** 由例7.2.10可知 $\pi(\theta|x)$ 是正态分布 $N(\mu_n(\boldsymbol{x}),\eta_n^2)$, 其中

$$\mu_n(\boldsymbol{x})=\frac{\sigma^2/n}{\sigma^2/n+\tau^2}\mu+\frac{\tau^2}{\sigma^2/n+\tau^2\overline{x}},\quad \eta_n^2=\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}.$$

由于后验分布 $N(\mu_n(\boldsymbol{x}),\eta_n^2)$ 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 $\theta$ 的 Bayes 估计为

$$\hat{\theta}_{\mathrm{B}}=\mu_n(\boldsymbol{x})=\frac{\sigma^2/n}{\sigma^2/n+\tau^2}\mu+\frac{\tau^2}{\sigma^2/n+\tau^2\overline{x}}.$$

如例7.2.10所述,$\hat{\theta}_{\mathrm{B}}$ 是 $\overline{x}$ 和 $\mu$ 加权平均, 是对样本信息和先验信息的综合.

**Example 7.3.2.** 为估计产品不合格品率 $\theta$, 今从一批产品中随机抽取 $n$ 件检查, 检查结果为 $X_1,\cdots,X_n$, 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 $X=\sum\limits_{i=1}^{n}X_i$. 若取 $\theta$ 的先验分布为共轭先验 $\mathrm{Be}(a,b)$, 求 $\theta$ 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 $\theta$ 的后验密度为

$$\pi(\theta|x)=\frac{\Gamma(n+a+b)}{\Gamma(x+a)\Gamma(b-x+n)}\theta^{x+a-1}(1-\theta)^{n-x+b-1},$$

即 $\pi(\theta|x)\sim\mathrm{Be}(x+a,n-x+b)$. 易知 $\theta$ 的后验众数估计为

$$\hat{\theta}_{\mathrm{MD}}=\frac{(x+a)-1}{(X+a)+(n-X+b)-2}=\frac{x+a-1}{n+a+b-2};$$

后验均值估计为

$$\hat{\theta}_{\mathrm{E}}=\frac{x+a}{n+a+b}.$$

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

$$\hat{\theta}_{\mathrm{MD}}=\frac{x}{n},\quad \hat{\theta}_{\mathrm{E}}=\frac{x+1}{n+2}.$$

对这两个估计作如下说明:①$\theta$ 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 $\theta$ 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$\theta$ 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 $\theta$ 的两种 Bayes 估计的比较表

| 试验号 | 样本量 $n$ | 不合格品数 $x$ | $\hat{\theta}_{\mathrm{MD}}=x/n$ | $\hat{\theta}_{\mathrm{MD}}=(x+1)/(n+2)$ |
|---|---|---|---|---|
| 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 $\hat{\theta}_{\mathrm{MD}}$ 皆为 0, 显示不出二者的差别, 而 $\hat{\theta}_{\mathrm{E}}$ 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,$\hat{\theta}_{\mathrm{MD}}=0$ 或 1, 这种估计未免太极端一些, 而 $\hat{\theta}_{\mathrm{E}}$ 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 $\hat{\theta}_{\mathrm{MD}}$ 与经典估计相同, 故 Bayes 估计 $\hat{\theta}_{\mathrm{E}}$ 显示出相对于经典估计的优点.

由此可见, 作为 $\theta$ 的估计, 后验均值估计常优于后验众数估计.

**Example 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 $\theta$ 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 $\theta$ 的后验众数估计.

**解:** $\theta$ 的后验密度为

$$\pi(\theta|x)=\frac{f(x|\theta)\pi(\theta)}{m(x)}=\frac{e^{-(x-\theta)}}{m(x)\pi(1+\theta^2)},\quad x\geqslant\theta.$$

要找使 $\pi(\theta|x)$ 最大化的 $\hat{\theta}$, 对后验密度关于 $\theta$ 求导数, 可得

$$\frac{\mathrm{d}\pi(\theta|x)}{\mathrm{d}\theta}=\frac{e^{-x}}{\pi m(x)}\left[\frac{e^{\theta}}{1+\theta^2}-\frac{2\theta e^{\theta}}{(1+\theta^2)^2}\right]=\frac{e^{-x}}{\pi m(x)}\cdot\frac{e^{\theta}(\theta-1)^2}{(1+\theta^2)^2},$$

故 $\pi(\theta|x)$ 的导数在 $\theta\leqslant x$ 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 $\theta_{\mathrm{MD}}=x$ 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 $\theta$ 的后验分布为 $\pi(\theta|x)$,$\theta$ 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):

$$\mathrm{PMSE}(\delta(x))=E^{\theta|x}[(\theta-\delta(x))^2]$$

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

$$\mathrm{PMSE}(\delta(x))=E^{\theta|x}[(\theta-\delta(x))^2]=D(\theta|x).$$

记 $\theta$ 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:

$$\begin{aligned}\mathrm{PMSE}(\delta(x))=&E^{\theta|x}[(\theta-\delta(x))^2]=E^{\theta|x}[(\theta-\mu^\pi(x))+(\mu^\pi(x)-\delta(x))]^2\\=&V^\pi(x)+(\mu^\pi(x)-\delta(x))^2\\\geqslant&V^\pi(x),\end{aligned}$$

且等号成立的充要条件是 $\delta(x)=\mu^\pi(x)$, 即 $\theta$ 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 $\theta$ 的 Bayes 估计.

**Example 7.3.4** (续例7.3.1)**.** 设 $X\sim N(0,\sigma^2),\sigma^2$ 已知,$\theta$ 未知, 设 $\theta$ 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 $\theta$ 的估计, 求估计量 $\hat\theta_E$ 的方差 $V^\pi(x)$, 并将其与 $\theta$ 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 $\eta^2$, 故知

$$V^\pi(x)=\eta^2=\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}.$$

而 $\delta(x)=x$ 的 PMSE 为

$$\begin{aligned}\mathrm{PMSE}(\delta(x))=&\mathrm{PMSE}(x)=V^\pi(x)+(\mu^\pi(x)-x)^2\\=&V^\pi(x)+\left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)^2(\mu-x)^2\geqslant V^\pi(x).\end{aligned}$$

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.

**Example 7.3.5.** 设 $X\sim N(\theta,\sigma^2)$, 其中 $\sigma^2$ 已知. 取 $\theta$ 的先验分布为无信息先验即 $\pi(\theta)\equiv 1$, 求 $\theta$ 的后验期望估计.

**解:** 在例7.2.4中取 $n=1$, 知 $\theta$ 的后验分布是 $N(x,\sigma^2)$,$\theta$ 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 $\theta$ 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.

**Example 7.3.6** (续例7.3.2)**.** 在例7.3.2中,$X\sim b(n,\theta)$,$\theta$ 的先验分布是 $\mathrm{Be}(a,b)$, 求 $\theta$ 后验期望估计的后验方差. 又当取 $\theta$ 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 $\theta$ 的后验分布是 $\mathrm{Be}(x+a, n-x+b)$, 故知 $\theta$ 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 $\theta$ 的先验分布是均匀分布 $U(0,1)$, 即 $\mathrm{Be}(1,1)$, 故 $\theta$ 的后验分布是 $\mathrm{Be}(x+1, n-x+1)$, 则后验均值估计和后验众数估计分别为

$$\hat{\theta}_E=\frac{x+1}{n+2}, \quad \hat{\theta}_{\mathrm{MD}}=\frac{x}{n}.$$

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 $\hat{\theta}_{\mathrm{MD}}$ 的后验均方误差是

$$\begin{aligned}
\mathrm{PMSE}(\hat{\theta}_{\mathrm{MD}}) =& V^\pi(x)+\left(\hat{\theta}_{MD}-\hat{\theta}_E\right)^2 \\
=& \frac{(x+1)(n-x+1)}{(n+2)^2(n+3)}+\left(\frac{x+1}{n+2}-\frac{x}{n}\right)^2 \\
\geqslant & V^\pi(x)
\end{aligned}$$

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 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_p)'$ 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,$\boldsymbol{\mu}^\pi(x)=E(\boldsymbol{\theta}|x)=(\mu_1^\pi(x),\cdots,\mu_p^\pi(x))'$, 估计量的精度用后验协方差阵 (记为 $\mathrm{Cov}^\pi(x)$) 来衡量

$$\mathrm{Cov}^\pi(x)=E^{\boldsymbol{\theta}|x}[(\boldsymbol{\theta}-\boldsymbol{\mu}^\pi(x))(\boldsymbol{\theta}-\boldsymbol{\mu}^\pi(x))'].$$

对 $\boldsymbol{\theta}$ 的任一估计 $\boldsymbol{\delta}(x)$, 其后验协方差阵可分解为

$$\mathrm{Cov}_\delta^\pi(x)=E^{\boldsymbol{\theta}|x}[(\boldsymbol{\theta}-\boldsymbol{\delta})(\boldsymbol{\theta}-\boldsymbol{\delta})']=\mathrm{Cov}^\pi(x)+(\boldsymbol{\mu}^\pi(x)-\boldsymbol{\delta})(\boldsymbol{\mu}^\pi(x)-\boldsymbol{\delta})'.$$

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 $\theta$ 的后验分布 $\pi(\theta|x)$ 获得后, 若存在区间 $[a,b]$, 使 $\theta$ 落在 $[a,b]$ 内的后验概率为 $1-\alpha$, 即

$$P(a \leqslant \theta \leqslant b|x) = 1-\alpha,$$

则称 $[a,b]$ 为 $\theta$ 的 Bayes 区间估计, 又称为可信区间. 这是对 $\theta$ 的后验分布为连续情形, 若 $\theta$ 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 $P(a \leqslant \theta \leqslant b|x) \geqslant 1-\alpha$, 这样的区间也是 $\theta$ 的 Bayes 可信区间. 定义如下:

**Definition 7.3.2** (可信区间)**.** 设参数 $\theta$ 的后验分布为 $\pi(\theta|x)$, 对给定的样本 $x$ 和 $0<\alpha<1$(通常 $\alpha$ 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

$$P(\hat{\theta}_1 \leqslant \theta \leqslant \hat{\theta}_2|x) \geqslant 1-\alpha,$$

则称 $[\hat{\theta}_1, \hat{\theta}_2]$ 为 $\theta$ 的可信水平为 $1-\alpha$ 的 Bayes **可信区间** (Bayesian credible interval). 而满足

$$P(\theta \geqslant \hat{\theta}_L|x) \geqslant 1-\alpha$$

的 $\hat{\theta}_L = \hat{\theta}_L(x)$ 称为 $\theta$ 的可信水平为 $1-\alpha$ 的**可信下限**, 而满足

$$P(\theta \leqslant \hat{\theta}_U|x) \geqslant 1-\alpha$$

的 $\hat{\theta}_U = \hat{\theta}_U(x)$ 称为 $\theta$ 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 $1-\alpha$ 后要求的可信区间, 如 $1-\alpha=0.90$ 的可信区间为 $[1.2, 2.0]$, 这时可以写为

$$P(1.2 \leqslant \theta \leqslant 2.0|x) = 0.90.$$

既可以说 "$\theta$ 属于这个区间的概率为 0.90", 也可以说 "$\theta$ 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 $\theta$ 为常数, 它要么在 $[1.2, 2.0]$ 之内, 要么在其外, 不能说 "$\theta$ 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 $\theta$". 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 $\theta$ 有关, 而它的分布与 $\theta$ 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.

**Example 7.3.7.** 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从正态总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, 其中 $\sigma^2$ 已知.$\theta$ 的先验分布是 $N(\mu, \tau^2)$, 其中 $\mu$,$\tau^2$ 已知. 求 $\theta$ 的可信水平 $1-\alpha$ 的可信区间.

**解:** 由例7.2.10已知 $\theta$ 的后验分布是 $N(\mu_n(x), \eta_n^2)$, 故有

$$P(\mu_n(\boldsymbol{x}) - \eta_n u_{\alpha/2} \leqslant \theta \leqslant \mu_n(\boldsymbol{x}) + \eta_n u_{\alpha/2} | \boldsymbol{x}) = 1 - \alpha,$$

其中 $u_{\alpha/2}$ 为 $N(0,1)$ 的上侧 $\alpha/2$ 分位数. 因此 $[\mu_n(\boldsymbol{x}) - \eta_n u_{\alpha/2}, \mu_n(\boldsymbol{x}) + \eta_n u_{\alpha/2}]$ 就是 $\theta$ 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 $\theta$ 为被测验儿童智商 IQ 真值,$\theta$ 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 $\theta$ 的后验分布是 $N(110.38, 8.32^2)$, 因此求得 $\theta$ 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 $\theta$ 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 "$\theta$ 落入区间 $[95.4, 134.6]$ 中的概率为 0.95", 也不能说 "$\theta$ 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 $\theta$ 的后验分布 $\pi(\theta \mid x)$ 后, 计算 $\Theta_0$ 和 $\Theta_1$ 的后验概率

$$\alpha_0 = P\left(\theta \in \Theta_0 \mid x\right), \quad \alpha_1 = P\left(\theta \in \Theta_1 \mid x\right).$$

比较 $\alpha_0$ 和 $\alpha_1$ 的大小决定接受 $H_0$ 还是 $H_1$.$\alpha_0$ 和 $\alpha_1$ 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:

$$\text{当 } \frac{\alpha_0}{\alpha_1} > 1 \text{ 时接受 } H_0, \text{ 否则拒绝 } H_0. \tag{7.3.1}$$

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.

**Example 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 $\theta$ 的先验分布为均匀分布 $U(0,1)$, 求下列检验问题:

$$H_0:\theta\leqslant 1/2\leftrightarrow H_1:\theta>1/2.$$

**解:** 由例7.2.9已知后验分布为 $\mathrm{Be}(x+1,n-x+1)$, 故有

$$\alpha_0=P(\Theta_0|x)=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\int_0^{1/2}\theta^x(1-\theta)^{n-x}\mathrm{d}\theta.$$

当取 $n=5$ 时可算得各种 $x$ 下的后验概率及后验机会比 $\alpha_0/\alpha_1$ 如下:

**表 7.3.2:** $\theta$ **的后验机会比**

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $\alpha_0$ | 63/64 | 57/64 | 42/64 | 22/64 | 7/64 | 1/64 |
| $\alpha_1$ | 1/64 | 7/64 | 22/64 | 42/64 | 57/64 | 63/64 |
| $\alpha_0/\alpha_1$ | 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 $H_0$.

**Example 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 $\theta$ 为这个孩子测验中的智商 IQ 真值,$\theta$ 的先验分布为 $N(100,225)$. 钙儿童测验得分 $x=115$, 求下列检验问题:

$$H_0:\theta\leqslant 100\leftrightarrow H_1:\theta>100.$$

**解:** 由例7.3.7可知 $\theta$ 的后验分布为 $N(110.38,8.32^2)$, 查标准正态分布表求得

$$\alpha_0=P(\theta\leqslant 100|x)=0.106,\quad \alpha_1=P(\theta>100|x)=0.894.$$

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

$$H_i:\theta\in\Theta_i,i=1,\cdots,k,\text{其中}\Theta_1\cup\Theta_2\cup\cdots\cup\Theta_k=\Theta,$$

其中 $\Theta$ 是参数空间,$\Theta_i(i=1,\cdots,k)$ 是 $\Theta$ 的非空真子集, 且两两不相交. 在求得 $\theta$ 的后验分布 $\pi(\theta|x)$ 后, 计算 $\Theta_i$ 的后验概率

$$\alpha_i = P(\Theta_i|x),\quad i=1,\cdots,k,$$

取其最大者, 则认为相应的假设成立.

**Example 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:

$$H_1:\theta\leqslant 90,\quad H_2: 90<\theta<110,\quad H_3:\theta\geqslant 110.$$

**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

$$P(\theta\leqslant 90|x=115)=0.007, P(90<\theta<110|x=115)=0.473,$$
$$P(\theta\geqslant 110|x=115)=0.520.$$

由于 $\Theta_3=\{\theta:\theta\geqslant 110\}$ 的后验概率最大, 故接受 $H_3$.

### *3. 预测推断

一般预测问题的典型情况是: 若 $X\sim f(x|\theta), \boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\theta$ 的后验分布, 于是 $g(z|\theta)\pi(\theta|x)$ 为给定 $X=x$ 的条件下 $(Z,\theta)$ 的联合分布, 把它对 $\theta$ 积分得到给定 $X=x$ 时 $Z$ 的边缘分布密度作为预测密度.

**Definition 7.3.3.** 设 $\theta$ 的后验密度是 $\pi(\theta|\boldsymbol{x})$. 给定 $\boldsymbol{x}$ 后, 随机变量 $Z$ 的**后验预测密度** (posterior predictive density) 定义为

$$p(z|\boldsymbol{x})=\int_\Theta g(z|\theta)\pi(\theta|\boldsymbol{x})\mathrm{d}\theta. \tag{7.3.2}$$

**Example 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 $\theta$, 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

$$f(x|\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x},\quad x=0,1,\cdots,n.$$

若取 $\theta$ 的先验分布为共轭先验 $\mathrm{Be}(a,b)$, 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

$$g(z|\theta)=\binom{k}{z}\theta^z(1-\theta)^{k-z},\quad z=0,1,\cdots,k.$$

于是在给定 $X=x$ 时, 随机变量 $Z$ 的后验预测密度为

$$\begin{aligned}p(z|x)=&\int_0^1\binom{k}{z}\theta^z(1-\theta)^{k-z}\pi(\theta|x)\mathrm{d}\theta.\\ =&c\int_0^1\theta^{z+x+a-1}(1-\theta)^{k-z+n-x+b-1}\mathrm{d}\theta\\ =&c\frac{\Gamma(z+x+a)\Gamma(k-z+n-x+b)}{\Gamma(k+n+a+b)},\end{aligned}$$

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 $\mathrm{Be}(1,1)=U(0,1)$, 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

$$\begin{aligned}&p(0|3)=0.1813,\quad &p(1|3)=0.3022,\quad &p(2|3)=0.2747,\\ &p(3|3)=0.1649,\quad &p(4|3)=0.0641,\quad &p(5|3)=0.02128.\end{aligned}$$

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

$$P_{Z|x}(0\leqslant Z\leqslant 3)=0.9231.$$

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

为了说明什么是决策 (判决) 问题, 请看下例.

**Example 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 ($\theta_1$) 与看跌 ($\theta_2$). 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 | $a_1$ | $a_2$ | $a_3$ |
|---|---|---|---|
| $\theta_1$ | 5000 | 3000 | 1000 |
| $\theta_2$ | $-10000$ | $-8000$ | 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.

**1. 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** $\mathscr{X}$ 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,$\theta$ 是未知参数,$\Theta$ 为**参数空间**,$\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\mathscr{A}$. 在估计问题中,$\mathscr{A}$ 由一切估计量 $\delta(x)$ 构成, 常取 $\mathscr{A}=\Theta$. 在检验问题中 $\mathscr{A}$ 只有两个行动组成, 即 $\mathscr{A}=\{a_0,a_1\}$, 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 $\Theta\times\mathscr{A}$ 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 $a\in\mathscr{A}$ 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 $\Theta$ 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.

**2. 风险函数和一致最优决策**

**Definition 7.4.1.** 定义于样本空间 $\mathscr{X}$ 内而取值于行动空间 $\mathscr{A}$ 内的函数 $\delta=\delta(\boldsymbol{x})$ 称为决策函数或判决函数 (decision rules).

设 $\delta$ 是采取的决策行动, 若参数为 $\theta$, 则由此造成的损失是 $L(\delta,\theta)$, 这个量与样本 $\boldsymbol{X}$ 有关, 因而是随机的. 故采取行动 $\delta$ 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.

**Definition 7.4.2.** 设 $\delta$ 是一个决策函数, 称平均损失

$$R(\delta,\theta)=E[L(\delta(\boldsymbol{X}),\theta)]=\int_{\mathscr{X}}L(\delta(\boldsymbol{x}),\theta)f(x,\theta)\mathrm{d}\boldsymbol{x} \tag{7.4.1}$$

为 $\delta$ 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.

**Definition 7.4.3** (一致最优决策函数)**.** 记 $R(\delta,\theta)$ 为决策函数 $\delta$ 的风险函数. 若存在一个决策函数 $\delta^*$, 使得对任一决策函数 $\delta$ 有

$$R(\delta^*,\theta)\leqslant R(\delta,\theta),\quad \text{一切}\theta\in\Theta, \tag{7.4.2}$$

则称 $\delta^*$ **为一致最优解或一致最优决策函数**.

**Note 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性准则获得的估计.7.5 节将要介绍的同变性也属于这一类.

**3.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.

**Definition 7.4.4.** 设 $R(\delta,\theta)$ 为 $\delta$ 的风险函数,$H(\theta)$ 为 $\theta$ 的先验分布 (若存在密度, 用 $\pi(\theta)$ 表示), 则称

$$R_H(\delta) = E^{\theta}[R(\delta,\theta)] = \int_{\Theta} R(\delta,\theta)\mathrm{d}H(\theta) \tag{7.4.3}$$

为 $\delta$ 的 Bayes **风险** (Bayes risk).

由定义可见 Bayes 风险是将风险函数以 $\theta$ 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.

**Definition 7.4.5.** 设 $\delta_1$ 和 $\delta_2$ 为 $\theta$ 的两个决策函数, 若 $R_H(\delta_1) \leqslant R_H(\delta_2)$, 则称 $\delta_1$ 在 Bayes 风险下优于 $\delta_2$. 若存在 $\delta^*$, 使得对任一决策函数 $\delta$ 有

$$R_H(\delta^*) \leqslant R_H(\delta) \tag{7.4.4}$$

则称 $\delta^*$ 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 $\theta$ 时, 随机变量 $X \sim f(x|\theta)$,$\theta$ 的先验分布是 $H(\theta)$(如有密度, 用 $\pi(\theta)$ 记之),$\theta$ 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.

**Definition 7.4.6.** 设 $H(\theta|x)$ 为 $\theta$ 的后验分布函数,$L(\delta,\theta)$ 为损失函数, 则称

$$R(\delta|x) = E^{\theta|x}[L(\delta,\theta)] = \int_{\Theta} L(\delta,\theta)\mathrm{d}H(\theta|x) \tag{7.4.5}$$

为决策函数 $\delta$ 的**后验风险** (posterior risk).

若存在决策函数 $\delta^*$, 使得对任一决策函数 $\delta$ 有

$$R(\delta^*|x) = \min_{\delta} R(\delta|x), \tag{7.4.6}$$

则称 $\delta^*$ 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

$$R_H(\delta) = E^{\theta}[R(\delta,\theta)] = \int_{\Theta} R(\delta,\theta)\mathrm{d}H(\theta) = \int_{\Theta}\left[\int_{\mathscr{X}} L(\delta,\theta)\mathrm{d}F(x|\theta)\right]\mathrm{d}H(\theta)$$

$$
\begin{aligned}
&=\int_{\mathscr{X}}\left[\int_{\Theta} L(\delta,\theta)\mathrm{d}H(\theta|x)\right]\mathrm{d}F_m(x)=\int_{\mathscr{X}} R(\delta|x)\mathrm{d}F_m(x)\\
&=E^X[R(\delta|x)],
\end{aligned}
\tag{7.4.7}
$$

其中 $F_m(x)$ 是 $X$ 的边缘分布 (如有密度, 以 $f_m(x)$ 记之),$E^\theta$ 表示“关于 $\theta$ 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 $\theta$ 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.

**Theorem 7.4.1.** 对任何样本值 $x$, 若存在决策函数 $\delta_H(x)$, 满足条件

$$R(\delta_H|x)=\inf_{\delta\in\mathscr{A}} R(\delta|x),$$

则 $\delta_H$ 为先验分布 $H(\theta)$ 之下的 Bayes 解.

**证:** 设 $\delta$ 为任一非随机化决策函数, 由已知条件可知

$$R(\delta|x)=\int_{\Theta} L(\delta,\theta)H(\mathrm{d}\theta|x)\geqslant\int_{\Theta} L(\delta_H,\theta)H(\mathrm{d}\theta|x)=R(\delta_H|x)$$

对一切 $x\in\mathscr{X}$ 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

$$R_H(\delta)=\int_{\mathscr{X}} R(\delta|x)\mathrm{d}F_m(x)\geqslant\int_{\mathscr{X}} R(\delta_H|x)\mathrm{d}F_m(x)=R_H(\delta_H),$$

即使 Bayes 风险达到最小的决策函数为 $\delta_H$, 因此它是 Bayes 解.

**Note 7.4.2.** 当 $\theta$ 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**

**Theorem 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,$\theta$ 的 Bayes 估计为其后验期望. 即

$$\hat{\theta}_B(x) = E(\theta|x). \tag{7.4.8}$$

**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

$$R(a|x) = E[(\theta-a)^2|x] = \int_{\Theta}[\theta^2 - 2a\theta + a^2]H(\mathrm{d}\theta|x).$$

令

$$\frac{\mathrm{d}R(a|x)}{\mathrm{d}a} = -2\int_{\Theta}\theta H(\mathrm{d}\theta|x) + 2a = 0,$$

解方程得 $a = \hat{\theta}_B(x) = \int_{\Theta}\theta H(\mathrm{d}\theta|x) = E(\theta|x)$, 且 $\frac{\mathrm{d}^2}{\mathrm{d}a^2}[R(a|x)] = 2 > 0$. 定理得证.

**Example 7.4.2.** 设 $X \sim N(\theta,\sigma^2)$, 其中 $\sigma^2$ 已知,$\theta$ 的先验分布是 $N(\mu,\tau^2)$. 求平方损失下 $\theta$ 的 Bayes 估计.

**解:** 由例7.2.10可知 $\theta$ 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 $\theta$ 的 Bayes 估计是

$$\hat{\theta}_B(x) = E(\theta|x) = \mu(x) = \frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}x.$$

**Example 7.4.3.** 设 $\boldsymbol{X} = (X_1,\cdots,X_n)$ 是从 Poisson 分布 $P(\theta)$ 中抽取的简单样本. 取 $\theta$ 的先验分布为 $\Gamma(\alpha,\lambda)$, 即 $\pi(\theta) = \left(\frac{\lambda^\alpha}{\Gamma(\alpha)}\right)\theta^{\alpha-1}e^{-\lambda\theta}I_{[\theta>0]}$, 其中 $\lambda>0,\alpha>0$ 已知, 求平方损失下 $\theta$ 的 Bayes 估计.

**解:** 由例7.2.11可知,$\theta$ 的后验分布为 $\Gamma(n\overline{x}+\alpha, n+\lambda)$, 故知 Bayes 估计

$$\hat{\theta}_B(x) = \frac{n\overline{x}+\alpha}{n+\lambda} = \frac{n}{n+\lambda}\overline{x} + \frac{\lambda}{n+\lambda}\cdot\left(\frac{\alpha}{\lambda}\right),$$

其中 $\overline{x}$ 为样本均值,$\alpha/\lambda$ 为先验分布的均值, 可见 $\theta$ 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 $\overline{x}$ 在 Bayes 估计中起主导作用, 当 $\lambda \gg n$ 时, 先验均值 $\alpha/\lambda$ 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**

**Theorem 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,$\theta$ 的 Bayes 估计为

$$\hat{\theta}_B = \frac{E[w(\theta)(\theta)|x]}{E[w(\theta)|x]}, \tag{7.4.9}$$

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

$$R(a|x) = E[w(\theta)(\theta-a)^2|x] = \int_\Theta \left[w(\theta)\theta^2 - 2\theta w(\theta)a + w(\theta)a^2\right] H(\mathrm{d}\theta|x).$$

令

$$\frac{\mathrm{d}}{\mathrm{d}a}[R(a|x)] = -2\int_\Theta \theta w(\theta)H(\mathrm{d}\theta|x) + 2a\int_\Theta w(\theta)H(\mathrm{d}\theta|x) = 0,$$

解方程得

$$a = \hat{\theta}_B = \frac{\int_\Theta \theta w(\theta)H(\mathrm{d}\theta|x)}{\int_\Theta w(\theta)H(\mathrm{d}\theta|x)} = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)},$$

且 $\frac{\mathrm{d}^2}{\mathrm{d}a^2}[R(a|x)] > 0$. 定理得证.

**Example 7.4.4.** 设 $\boldsymbol{X} = (X_1,\cdots,X_n)$ 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 $\theta$ 的先验分布服从逆 Gamma 分布 $\Gamma^{-1}(\alpha,\ \lambda)$, 即先验密度是

$$\pi(\theta) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\lambda/\theta}, & \theta>0, \\ 0, & \text{其他}. \end{cases}$$

求 $\theta$ 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

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta) \propto \theta^{-n}e^{-n\overline{x}/\theta}.$$

类似于例7.2.13, 可知后验分布

$$\pi(\theta|\boldsymbol{x}) \propto f(\boldsymbol{x}|\theta)\pi(\theta) \propto \theta^{-(n+\alpha+1)}e^{-(n\overline{x}+\lambda)/\theta},$$

添加正则化常数

$$\pi(\theta|\boldsymbol{x}) = \frac{(n\overline{x}+\lambda)^{n+\alpha}}{\Gamma(n+\alpha)}\cdot\theta^{-(n+\alpha+1)}e^{-(n\overline{x}+\lambda)/\theta}I_{(0,\infty)}(\theta).$$

$\theta$ 的 Bayes 估计由式(7.4.9)给出, 其中

$$
\begin{aligned}
E(\theta w(\theta)|\boldsymbol{x}) =& E(\theta^{-1}|\boldsymbol{x}) = \int_0^\infty \theta^{-1}\cdot \pi(\theta|\boldsymbol{x})\mathrm{d}\theta = \frac{n+\alpha}{n\overline{x}+\lambda},\\
E(w(\theta)|\boldsymbol{x}) =& E(\theta^{-2}|\boldsymbol{x}) = \int_0^\infty \theta^{-2}\cdot \pi(\theta|\boldsymbol{x})\mathrm{d}\theta = \frac{(n+\alpha+1)(n+\alpha)}{(n\overline{x}+\lambda)^2},
\end{aligned}
$$

因此得到

$$
\hat{\theta}_B = \frac{E(\theta w(\theta)|\boldsymbol{x})}{E(w(\theta)|\boldsymbol{x})} = \frac{(n+\alpha)/(n\overline{x}+\lambda)}{(n+\alpha+1)(n+\alpha)/(n\overline{x}+\lambda)^2} = \frac{n\overline{x}+\lambda}{n+\alpha+1}.
$$

**3. 绝对值损失下的 Bayes 解**

**Theorem 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,$\theta$ 的 Bayes 估计为后验中位数.

证明略.

**Example 7.4.5.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从均匀分布 $U(0,\theta)$ 中抽取的简单样本,$\theta$ 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 $\theta$ 在绝对值损失下的 Bayes 估计.

**解:** $X_1,\cdots,X_n$ 的联合密度 (或称为似然函数) 为

$$
f(\boldsymbol{x}|\theta)=\frac{1}{\theta^n}I_{(0,\theta)}(x_{(n)}),
$$

其中 $x_{(n)}=\max\{x_1,\cdots,x_n\}$. 由式(7.2.3)可知

$$
\pi(\theta|\boldsymbol{x})\propto f(\boldsymbol{x}|\theta)\pi(\theta)\propto\frac{1}{\theta^{n+\alpha+1}},\quad \theta>\max\{x_{(n)},\theta_0\}.
$$

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:

$$
\pi(\theta|x)=\frac{(\alpha+n)\theta_1^{\alpha+n}}{\theta^{\alpha+n+1}},\quad \theta>\theta_1;\quad H(\theta|x)=1-\left(\frac{\theta_1}{\theta}\right)^{\alpha+n},\quad \theta>\theta_1.
$$

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 $\theta$ 的 Bayes 估计为后验均值, 即

$$
\hat{\theta}_{n_1}=\frac{\alpha+n}{\alpha+n-1}\theta_1.
$$

这个估计要比经典方法中极大似然估计 $\hat{\theta}=\max(x_1,\cdots,x_n)$ 大些, 因为

$$\frac{\alpha+n}{\alpha+n-1}>1.$$

**Note 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 $\mathscr{A}=\{a_1,\cdots,a_r\}$. 令在参数为 $\theta$, 行动为 $a_i$ 下的损失函数是 $L(a_i,\theta)$,$i=1,\cdots,r$. 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:

$$H_0:\theta\in\Theta_0\leftrightarrow H_1:\theta\in\Theta_1\text{且}\Theta_0\cup\Theta_1=\Theta.$$

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.

$$L(a_i,\theta)=\begin{cases}0, & \theta\in\Theta_i,\\ 1, & \theta\notin\Theta_i,\end{cases}$$

其后验风险为

$$R(a_0|x)=\int_\Theta L(a_0,\theta)\pi(\theta|x)\mathrm{d}\theta=\int_{\Theta_1}\pi(\theta|x)\mathrm{d}\theta=P(\Theta_1|x),$$

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 $R(a_0|x)\geqslant R(a_1|x)$, 则否定 $H_0$, 即接受 $H_1$. 等价地, 若

$$P(\Theta_1|x)\geqslant P(\Theta_0|x)\text{时否定}H_0,\text{接受}H_1. \tag{7.4.10}$$

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

$$R(a_0|x) > R(a_1|x) \tag{7.4.11}$$

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

$$P(\Theta_1|x) > \frac{k_1}{k_0+k_1}\text{时否定}H_0\text{, 否则接受}H_0.$$

**Example 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 $\theta$ 为这个孩子测验中的智商 IQ 真值,$\theta$ 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

$$H_0 : \theta \leqslant 105 \leftrightarrow H_1 : \theta > 105.$$

**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

$$\begin{aligned} P(\Theta_0|x = 115) =& P(\theta \leqslant 105|x = 115) = \Phi(-5.38/8.32) = 0.2578, \\ P(\Theta_1|x = 115) =& 1 - P(\Theta_0|x = 115) = 0.7422. \end{aligned}$$

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:

$$H_0 : \theta_1 - \theta_2 < -\varepsilon, \quad H_1 : \theta_1 - \theta_2 > \varepsilon, \quad H_2 : |\theta_1 - \theta_2| \leqslant \varepsilon,$$

其中 $\varepsilon > 0$ 的选择使得当 $|\theta_1 - \theta_2| \leqslant \varepsilon$ 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.

**Example 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:

$$H_1:\ \theta<90,\quad H_2:90\leqslant\theta\leqslant 110,\quad H_3:\ \theta>110.$$

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:

$$L(a_1,\theta)=\begin{cases}0, & \theta<90,\\ \theta-90, & 90\leqslant\theta\leqslant 110,\\ 2(\theta-90), & \theta>110,\end{cases}$$

$$L(a_2,\theta)=\begin{cases}90-\theta, & \theta<90,\\ 0, & 90\leqslant\theta\leqslant 110,\\ \theta-110, & \theta>110,\end{cases}$$

$$L(a_3,\theta)=\begin{cases}2(110-\theta), & \theta<90,\\ 110-\theta, & 90\leqslant\theta\leqslant 110,\\ 0, & \theta>110.\end{cases}$$

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

$$\begin{aligned}R(a_1|x)=E^{\theta|x}[L(a_1,\theta)]=&\int_{90}^{110}(\theta-90)\pi(\theta|x)\mathrm{d}\theta+2\int_{110}^{\infty}(\theta-90)\pi(\theta|x)\mathrm{d}\theta\\ =&6.49+27.83=34.23,\\ R(a_2|x)=E^{\theta|x}[L(a_2,\theta)]=&\int_{-\infty}^{90}(90-\theta)\pi(\theta|x)\mathrm{d}\theta+\int_{110}^{\infty}(\theta-110)\pi(\theta|x)\mathrm{d}\theta\\ =&0.02+3.53=3.55,\\ R(a_3|x)=E^{\theta|x}[L(a_3,\theta)]=&\int_{-\infty}^{90}2(110-\theta)\pi(\theta|x)\mathrm{d}\theta+\int_{90}^{110}(110-\theta)\pi(\theta|x)\mathrm{d}\theta\\ =&0.32+2.95=3.27.\end{aligned}$$

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 $\theta\geqslant 110$.

***3. 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 $\theta$ 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 $\theta$ 不属于 $C(x)$ 引起的损失. 按后验风险最小的原则, 应使

$$R(C(x)|x)=E^{\theta|x}[L(C(x),\theta)]=m_1[d_2(x)-d_1(x)]+m_2P_{\theta|x}(\theta\notin C(x))$$

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 $\delta$ 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 $\delta$ 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.

**Definition 7.5.1.** 设 $\delta_1$ 和 $\delta_2$ 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 $\delta_1$ 优于 $\delta_2$. 如果存在某个决策函数 $\delta^*$, 对任何决策函数 $\delta$ 都有

$$M(\delta^*)\leqslant M(\delta),$$

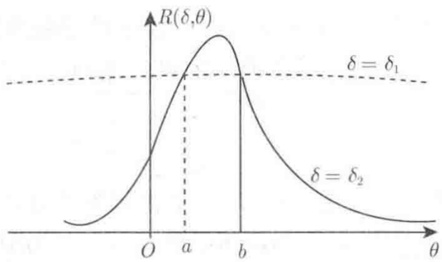
则称 $\delta^*$ 为该统计决策问题的 Minimax 解, 也称 $\delta^*$ 为 Minimax 决策函数. 当统计决策问题为估计或检验时, 也称 $\delta^*$ 为 Minimax 估计或 Minimax 检验.

**Note 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,$\delta_1$ 优于 $\delta_2$. 但是仔细观看二者的风险函数, 发现对大多数 $\theta$ 而言,$\delta_2$ 优于 $\delta_1$, 仅当 $a<\theta<b$ 时,$\delta_1$ 优于 $\delta_2$. 如果没有足够的先验信息说

明 $\theta$ 处在 $a,b$ 之间, 就很难说 $\delta_1$ 优于 $\delta_2$ 了. 因此 Bayes 学派认为, 只是在人们对 $\theta$ 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.

**Theorem 7.5.1.** 设 $\delta^*$ 为在先验分布 $H(\theta)$ 之下的 Bayes 解, 且 $\delta^*$ 的风险函数为常数 $c$, 即对任何 $\theta\in\Theta$ 有 $R(\delta^*,\theta)=c$, 则 $\delta^*$ 为一个 Minimax 解. **证:** 采用反证法. 若不然,$\delta^*$ 不是 Minimax 解, 则存在估计量 $\delta$ 使得

$$\sup_{\theta\in\Theta} R(\delta,\theta)<\sup_{\theta\in\Theta} R(\delta^*,\theta)\equiv c. \tag{7.5.1}$$

故 $R(\delta,\theta)<c$ 对一切 $\theta\in\Theta$ 成立, 这时将两边关于 $\theta$ 的先验分布 $H(\theta)$ 求平均可得

$$R_H(\delta)=\int_\Theta R(\delta,\theta)\mathrm{d}H(\theta)<c\int_\Theta \mathrm{d}H(\theta)=\int_\Theta R(\delta^*,\theta)\mathrm{d}H(\theta)=R_H(\delta^*),$$

即 $R_H(\delta)<R_H(\delta^*)$(其中 $R_H(\delta)$ 为 $\delta$ 的 Bayes 风险,$R_H(\delta^*)$ 为 $\delta^*$ 的 Bayes 风险), 这与 $\delta^*$ 为 Bayes 解矛盾.

**Example 7.5.1.** 设 $X\sim b(n,\theta)$,$\theta$ 的先验分布是 $\mathrm{Be}(a,b)$. 损失函数为 $L(d,\theta)=(\theta-d)^2$, 求 $\theta$ 的 Minimax 估计.

**解:** 由例7.2.9可知 $\theta$ 的后验分布 $\pi(\theta|x)$ 是 $\mathrm{Be}(x+a,n-x+b)$, 故 $\theta$ 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 $\theta$ 的 Minimax 估计.

二项分布中参数 $\theta$ 的传统的估计 $\delta_2=\overline{X}$ 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 $\delta_2$ 的风险函数的图形与图7.5.1类似, 所不同的是 $\delta_1$ 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 $\theta$ 而言,$\delta_2$ 优于 $\delta_1$, 仅当 $a<\theta<b$ 时,$\delta_1$ 优于 $\delta_2$. 如果没有足够的先验信息说明 $\theta$ 以较大的概率落在 $a,b$ 之间, 就很难说 $\delta_1$ 优于 $\delta_2$ 了. 因此, 除非有某种特殊的原因, 人们宁可用 $\delta_2$ 而不用 $\delta_1$.

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.

**Theorem 7.5.2.** 设 $\delta_k$ 为一个统计决策问题在先验分布 $H_k$ 之下的 Bayes 解, 假定 $\delta_k$ 的 Bayes 风险为 $r_k, k=1,2,\cdots$, 且有

$$\lim_{k\to\infty} r_k=r<\infty. \tag{7.5.2}$$

又设 $\delta^*$ 为同一问题的一个决策函数, 满足条件

$$M(\delta^*)\leqslant r, \tag{7.5.3}$$

则 $\delta^*$ 为此统计决策问题的 Minimax 解. **证:** 反证法. 若不然,$\delta^*$ 不是 Minimax 解, 则必存在决策函数 $\delta$ 使得

$$M(\delta)<M(\delta^*)\leqslant r=\lim_{k\to\infty} r_k.$$

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

$$\begin{aligned}R_{H_k}(\delta) &= \int_\Theta R(\delta,\theta)\mathrm{d}H_k(\theta) \leqslant \int_\Theta M(\delta)\mathrm{d}H_k(\theta) = M(\delta)\\ &< r_k = \int_\Theta R(\delta_k,\theta)dH_k(\theta) = R_{H_k}(\delta_k).\end{aligned}$$

这与 $\delta_k$ 为先验分布 $H_k(\theta)$ 之下的 Bayes 解矛盾.

**Example 7.5.2.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $\theta$ 的 Minimax 估计.

**解:** 找一串 $\theta$ 的先验分布 $\{H_k\}$,$H_k$ 为 $N(0,k^2), k=1,2,\cdots$. 由例7.2.10可知 $\theta$ 的后验分布是 $N(\mu_k(\boldsymbol{x}),\eta^2)$. 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

$$\delta_k(x)=\mu_k(x)=\frac{nk^2}{nk^2+1}\overline{x},$$

其风险函数

$$\begin{aligned}R(\delta_k,\theta) =& E\left[\left(\frac{nk^2\overline{X}}{nk^2+1}-\theta\right)^2\right]\\ =& D\left(\frac{nk^2\overline{X}}{nk^2+1}\right)+\left(\frac{nk^2}{nk^2+1}E(\overline{X})-\theta\right)^2\\ =& \frac{nk^4}{(nk^2+1)^2}+\frac{\theta^2}{(nk^2+1)^2}=\frac{nk^4+\theta^2}{(nk^2+1)^2},\end{aligned}$$

其中均值 $E$ 和方差 $D$ 是关于"给定 $\theta$ 时 $X$ 的分布"计算的. 故 $\delta_k$ 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

$$\lim_{k\to\infty} r_k=\lim_{k\to\infty}\frac{k^2}{nk^2+1}=\frac{1}{n}=r.$$

若取 $\delta^*=\overline{X}$, 其风险函数 $R(\delta^*,\theta)=E(\overline{X}-\theta)^2=1/n\leqslant r$, 故由定理7.5.2可知,$\delta^*=\overline{X}$ 为 $\theta$ 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 $X_1,\cdots,X_n$, 假定它们相互独立, 同服从正态分布 $N(a,\sigma^2)$. 用 $\delta(X_1,\cdots,X_n)$ 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 $X_1',\cdots,X_n'$,$X_i'=X_i+c$,$c>0$,$i=1,2,\cdots,n$, 用这批数据去估计物重 $a$, 即用 $\delta(x_1',\cdots,x_n')=\delta(x_1+c,\cdots,x_n+c)$ 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 $\delta(x_1+c,\cdots,x_n+c)-c$ 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 $\delta$ 满足条件: 对一切实数 $c_i$ 有

$$\delta(x_1',\cdots,x_n')=\delta(x_1+c,\cdots,x_n+c)=\delta(x_1,\cdots,x_n)+c, \tag{7.6.1}$$

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 $X_1,\cdots,X_n$, 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 $\delta(x_1,\cdots,x_n)$, 若把测量单位改变 (如将m 改为cm), 则得到数据为 $X_1',\cdots,X_n'$,$X_i'=cX_i$,$i=1,2,\cdots,n$,$1<c<\infty$. 用 $\delta(x_1',\cdots,x_n')=\delta(cx_1,\cdots,cx_n)$ 去估计 $b$ 显然偏高, 它的平均值应当是 $cb$. 因此, 应当用 $\delta(x_1',\cdots,x_n')/c$ 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 $\delta$ 满足条件

$$\delta(x_1',\cdots,x_n')=\delta(cx_1,\cdots,cx_n)=c\delta(x_1,\cdots,x_n),c>0. \tag{7.6.2}$$

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.

**Definition 7.6.1.** 设 $\delta(x_1,\cdots,x_n)$ 是参数 $\theta$ 的一个估计量. 假如对样本作了某种变换, 估计量 $\delta(x_1,\cdots,x_n)$ 仍保持此变换的统计特性, 则称 $\delta(x_1,\cdots,x_n)$ 是在该变换下 $\theta$ 的同变估计. 在变换是明确无误的情形下, 可简称 $\delta(x_1,\cdots,x_n)$ 是参数 $\theta$ 的同变估计.

例如, 若对样本作平移变换, 则位置参数 $\theta$ 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 $\theta$ 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 $\delta$ 的风险函数, 则有如下定义.

**Definition 7.6.2.** 设 $\delta^*(x_1,\cdots,x_n)$ 是某特定变换下 $\theta$ 的一个同变估计, 如果对此种变换下的任一同变估计 $\delta(x_1,\cdots,x_n)$, 有

$$R(\delta^*,\theta)\leqslant R(\delta,\theta),\quad \forall\theta\in\Theta,$$

则称 $\delta^*(x_1,\cdots,x_n)$ 是此种变换下 $\theta$ 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.

**Corollary 7.6.1.** 设 $X=(X_1,\cdots,X_n)$ 为从正态总体 $N(a,\sigma^2)$ 中抽取的简单样本, 而 $f(X_1,\cdots,X_n)$ 满足条件:$f(X_1+c,\cdots,X_n+c)=f(X_1,\cdots,X_n)$, 对任何实数 $c$, 则 $\overline{X}$ 与 $f(X_1,\cdots,X_n)$ 独立.

**Example 7.6.1.** 设 $X_1,\ldots,X_n$ 为从正态总体 $N(a,\sigma^2)$ 中抽取的简单样本, 则样本均值 $\overline{X}$ 是 $a$ 的平移变换下最优同变估计.

**证:** 设 $\delta(\boldsymbol{X})=\delta(X_1,\ldots,X_n)$ 为 $a$ 的任一位置同变估计 (即满足式(7.6.1)), 显然,$\overline{X}$ 也是 $a$ 的位置同变估计. 记

$$\delta_0(\boldsymbol{X})=\delta(\boldsymbol{X})-\overline{X},$$

则 $\delta_0$ 满足:$\delta_0(X_1+c,\ldots,X_n+c)=\delta_0(X_1,\ldots,X_n)$. 由引理7.6.1, 可知 $\overline{X}$ 与 $\delta_0(\boldsymbol{X})$ 独立. 若取损失函数为平方损失, 则有

$$R(\delta(\boldsymbol{X}),a)=E_a[\delta(\boldsymbol{X})-a]^2=E_a[(\delta(\boldsymbol{X})-\overline{X})+(\overline{X}-a)]^2.$$

由于 $\overline{X}$ 与 $\delta_0(\boldsymbol{X})$ 独立, 故有

$$\begin{aligned}R(\delta(\boldsymbol{X}),a)=&E_a(\overline{X}-a)^2+E_a[\delta_0^2(\boldsymbol{X})]+2E_a(\overline{X}-a)\cdot E_a[\delta_0(\boldsymbol{X})]\\=&E_a(\overline{X}-a)^2+E_a[\delta_0^2(\boldsymbol{X})]\\\geqslant&E_a(\overline{X}-a)^2=R(\overline{X},a),\text{对一切}a.\end{aligned}$$

这就证明了 $\overline{X}$ 是 $a$ 的一切位置同变估计中风险一致最小者, 因而 $\overline{X}$ 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.

**Definition 7.6.3.** 记 $R(\delta,\theta)$ 为决策函数 $\delta$ 的风险函数. 若存在决策函数 $\delta_1$ 使得

(1) 对任何 $\theta\in\Theta$, 有 $R(\delta_1,\theta)\leqslant R(\delta,\theta)$;

(2) 至少存在一个 $\theta_0\in\Theta$, 使 $R(\delta_1,\theta_0)<R(\delta,\theta_0)$,

则称 $\delta_1$ 一致地优于 $\delta$, 而称 $\delta$ 是**不可容许的决策函数**. 反之, 若不存在一致地优于 $\delta$ 的决策函数, 则称 $\delta$ 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.

**Theorem 7.6.1.** 在 Bayes 决策问题中, 设 $\delta_H$ 为先验分布 $H(\theta)$ 下的 Bayes 解, 若 $H(\theta)$ 和 $\delta_H$ 满足下列条件:

(1) 先验分布 $H(\theta)$ 对参数空间 $\Theta$ 的任何非空开子集有正概率;

(2) $\delta_H$ 的 Bayes 风险 $R_H(\delta_H)<\infty$;

(3) 对任何决策函数 $\delta$,$R(\delta,\theta)$ 为 $\theta$ 的连续函数, 则 $\delta_H$ 为可容的.

**Note 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.

证明略.

**Theorem 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 $\delta_H$ 是唯一的, 则它也是可容许的.

**证:** 反证. 若 $\delta_H$ 不是可容许的, 则必存在另一个估计 $\delta^*\neq\delta_H$, 使得

$$R(\delta^*,\theta)\leqslant R(\delta_H,\theta),\quad \text{一切}\quad \theta\in\Theta,$$

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

$$R_H(\delta^*)=\int_\Theta R(\delta^*,\theta)\mathrm{d}H(\theta)\leqslant\int_\Theta R(\delta_H,\theta)\mathrm{d}H(\theta)=R_H(\delta_H).$$

故 $\delta^*=\delta_H$ 与 $\delta_H$ 的唯一性矛盾.

**Theorem 7.6.3.** 在统计决策问题中, 假如 $\delta_0$ 为参数 $\theta$ 的唯一 Minimax 估计, 则 $\delta_0$ 也是 $\theta$ 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

$$R(\delta^*,\theta)\leqslant R(\delta_0,\theta),\quad \text{一切}\quad \theta\in\Theta,$$

且严格不等式至少对某个 $\theta\in\Theta$ 成立, 因此有

$$\sup_{\theta\in\Theta}R(\delta^*,\theta)\leqslant\sup_{\theta\in\Theta}R(\delta_0,\theta),$$

从而 $\delta^*$ 也是 $\theta$ 的 Minimax 估计, 与唯一性矛盾.

**Example 7.6.2.** 设随机变量 $X \sim b(n,\theta)$,$\theta \sim \mathrm{Be}(a,b)$, 损失函数 $L(d,\theta)=(d-\theta)^2$. 证明参数 $\theta$ 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 $x$,$\theta$ 的后验分布是 $\mathrm{Be}(a+x,b+n-x)$. 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 $\theta$ 的 Minimax 估计.

易验证 $\theta$ 的先验分布 $\mathrm{Be}(a,b)$ 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 $\delta$ 的风险函数

$$R(\delta,\theta)=E(\delta-\theta)^2=\sum_{i=0}^{n}(\delta(i)-\theta)^2\binom{n}{x}\theta^i(1-\theta)^{n-i}$$

是 $\theta$ 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 $\theta$ 的 Minimax 估计(7.6.3)是可容许的.

**Example 7.6.3.** 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 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 $c\overline{X}$ 是 $\theta$ 的可容许估计.

**证:** 事实上, 若取 $\theta$ 的先验分布 $H(\theta)$ 为 $N(0,\tau^2)$, 其中 $\tau^2$ 为已知常数, 取 $c=n\tau^2/(1+n\tau^2)$, 则由例7.2.10可知 $\theta$ 的 Bayes 估计是后验均值, 即

$$\delta_H(\boldsymbol{X})=\frac{n\tau^2}{1+n\tau^2}\overline{X}=c\overline{X}$$

不难验证定理7.6.1的条件 (1)-(3) 皆成立, 故 $\delta_H(\boldsymbol{X})$ 是 $\theta$ 的可容许估计.

此处不能用定理7.6.1的方法直接证明 $\theta$ 的常见估计 $\overline{X}=X/n$ 的可容许性. 但 $\overline{X}$ 确是 $\theta$ 的可容许估计, 详细信息可查看文献.

### 2. James-Stein 估计

设 $X_1,\cdots,X_n$ 为从 $p$ 维正态分布 $N(\boldsymbol{\theta},\boldsymbol{I}_p)$ 中抽取的简单样本, 取损失函数 $L(\boldsymbol{\theta},\boldsymbol{d})=\|\boldsymbol{d}-\boldsymbol{\theta}\|^2$, 其中 $\boldsymbol{\theta},\boldsymbol{d}$ 皆为 $p$ 维向量.$\boldsymbol{\theta}$ 的通常估计 $\overline{\boldsymbol{X}}=\sum\limits_{i=1}^{n}X_i/n$ 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 $\overline{\boldsymbol{X}}$ 也是可容许的.C.Stein 在 1955 年证明: 当 $p\geqslant 3$, 样本均值

向量 $\overline{X}$ 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 $\overline{\boldsymbol{X}}$ 更优的估计

$$\boldsymbol{\theta}_{\mathrm{JS}} = \left(1 - \frac{p-2}{\|\overline{\boldsymbol{X}}\|^2}\right)\overline{\boldsymbol{X}}.$$

这个估计称为 James-Stein 估计.