数理统计

Charles

2023 年 1 月

目录

1 基础知识 **1**

1.1 总体与样本 1

1.1.1 基本概念 1

1.1.2 样本数据的整理与显示 3

1.2 统计量及其分布 4

1.2.1 统计量 4

1.2.2 三大抽样分布 8

1.2.3 统计量的抽样分布 10

1.2.4 充分统计量 12

1.2.5 完备统计量 13

2 参数估计 **14**

2.1 点估计 14

2.1.1 估计的评价标准 14

2.1.2 矩估计 19

2.1.3 最大似然估计 19

2.2 区间估计 22

2.2.1 置信区间 22

2.2.2 正态总体参数的置信区间 23

2.2.3 大样本置信区间 25

3 假设检验 **26**

3.1 假设检验的基本思想与概念 26

3.1.1 假设检验的步骤 26

3.1.2 基本原理和逻辑 28

3.1.3 两个假设的选取 28

3.1.4 检验的 p 值 28

3.2 正态总体参数假设检验 29
3.2.1 单个正态总体均值的检验 29
3.2.2 两个正态总体均值差的检验 30
3.2.3 正态总体方差的检验 32
3.3 其他分布参数的假设检验 33
3.3.1 指数分布参数的检验 33
3.3.2 比率 p 的检验 34
3.3.3 大样本检验 35

Chapter 1

基础知识

1.1 总体与样本

1.1.1 基本概念

总体

研究对象的全体称为**总体**. 组成总体的每个成员称为**个体**.

总体和个体可以是具体的对象, 也可以是抽象的数据. 例如, 研究某校学生的月消费支出情况, 则总体可以是该校所有学生, 每个学生就是一个个体, 这时总体与个体指的是具体的人; 也可以以该校所有学生的月消费支出为总体, 每个学生的月消费支出就是个体, 这时总体与个体指的是数据.

本课常说的 "总体" 只包含所有可能的不同的研究对象和每个研究对象出现的频率两方面的信息. 等价地, 总体可用一个**分布** (称作总体分布) 来描述. 通常提到的 "总体", 指的就是一个分布, 有时也指服从该分布的随机变量.

总体可以分为**有限总体**和**无限总体**.

样本

从总体中随机地抽取 n 个个体称为总体的一个**样本**, n 称为**样本容量**, 样本中的个体称为**样品**.

样本具有二重性: 一方面, 样本是从总体中随机抽取的, 抽取前无法预知它们的数值, 因此样本是随机变量, 记为 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$; 另一方面, 在实际中对样本观察或者测量了之后, 就可以得出一组确定的数, 即样本观测值 (有时也称作样本), 记为 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$.

给出样本中每一个样品观测值的具体数值的样本称为**完全样本**; 样本观测值没有具体的数值只有一个范围的样本称为**分组样本**, 这是一种不完全样本.

为了能由样本对总体作出较可靠的推断, 就希望样本能很好地代表总体. 这就需要对抽样方法提出一些要求, 最常用的 “简单随机抽样” 有如下两个要求:

(1) 样本具有**随机性**: 要求总体中每个个体都有同等机会被选入样本. 即每个样品 X_i 与总体 X 具有相同分布;

(2) 样本要有**独立性**: 要求样本中每一件样品的取值不影响其他样品的取值. 即 X_1, X_2, $\cdots$, X_n 相互独立.

这样得到的样本称作**简单随机样本**, 简称样本. 其中的样品是**独立同分布**的, 所以简单随机样本也称为 iid 样本, 其共同分布就是总体分布. 设总体 X 的分布是 $F(x)$, 那么样本 $\boldsymbol{X}$ 的联合分布是

$$F_{\boldsymbol{X}}\left(x_1, x_2, \cdots, x_n\right)=\prod_{i=1}^n F\left(x_i\right).$$

对无限总体, 随机性与独立性容易实现, 关键在于排除有意或无意的人为干扰; 对有限总体, 只要总体所含个体数很大, 特别是与样本量相比很大时, 则独立性也可基本得到满足. 当总体个数 N 很大, 而样本量 n 不大 $(n/N \leqslant 0.1)$ 时可以把样本近似地看作简单随机样本.

经验分布函数

定义 1.1.1 (经验分布函数) 设 $x_1, x_2, \cdots, x_n$ 是从总体 $F(x)$ 中获得的一组样本, 定义

$$F_n(x)=\frac{1}{n} \sum_{i=1}^n I\left(x_i \leqslant x\right).$$

$F_n(x)$ 是一个分布函数, 称作该样本的**经验分布函数**.

经验分布是一个离散分布 $P(t_i)=f_i/n$, 其中 $t_1<t_2<\cdots<t_k$ 是 $x_1, x_2, \cdots, x_n$ 所有的不同值从小到大的排列, $f_1, f_2, \cdots, f_k$ 是相应的频数. 经验分布可表示为

$$F_n(x)=\begin{cases}0, & x<t_1, \\ \dfrac{1}{n} \displaystyle\sum_{j=1}^i f_j, & t_i \leqslant x<t_{i+1}\ (i=1,2, \cdots, k-1), \\ 1, & x \geqslant t_k.\end{cases}$$

经验分布函数具有确定性和随机性两种属性:

(1) 确定性: 当样本 $x_1, x_2, \cdots, x_n$ 给定时, 分布函数 $F_n(x)=\dfrac{1}{n} \displaystyle\sum_{i=1}^n I\left(x_i \leqslant x\right)$ 是一个确定的分布函数;

(2) 随机性: 对于给定的 x 值, $F_n(x)$ 是样本 $x_1, x_2, \cdots, x_n$ 的函数, 这时, 经验分布函数是一个随机变量; 同时考察所有的 x 值时, $F_n(x)$ 是一个依赖于样本的随机函数.

由伯努利大数定律, $n \rightarrow \infty$ 时, $F_n(x)$ 依概率收敛到 $F(x)$. 更深刻的定理还有格利文科定理.

定理 1.1.2 (格利文科 (Glivenko) 定理) 设 x_1, x_2, $\cdots$, x_n 是取自总体 $F(x)$ 的样本, $F_n(x)$ 是其经验分布函数, 当 $n \to \infty$ 时, 有

$$P\left(\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0\right) = 1.$$

格利文科定理表明, 当 n 相当大时, 经验分布函数是总体分布函数 $F(x)$ 的一个良好的近似. 经典统计学中一切统计推断都以样本为依据的理由就在于此.

用程序模拟标准正态总体不同样本量的经验分布函数如图 1.1, 可以看出, 样本量越大, 经验分布函数越接近总体分布函数, 且样本量很大时, 经验分布函数与总体分布函数已十分接近.

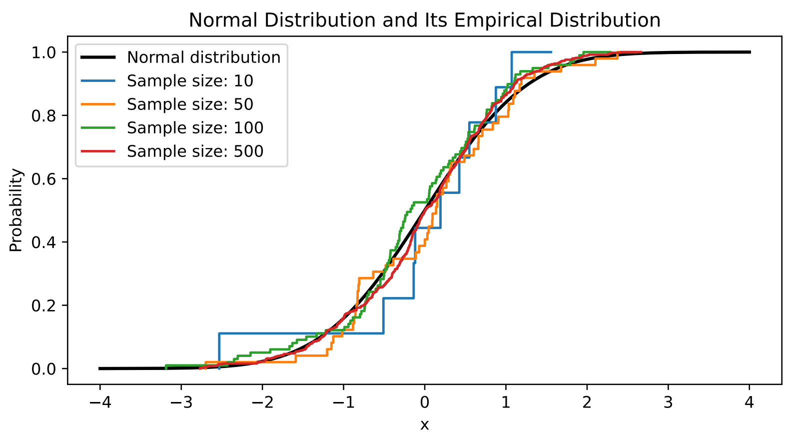

图 1.1: 经验分布函数与总体分布函数

1.1.2 样本数据的整理与显示

频数频率表

频数频率表的具体制作步骤:

(1) 对样本进行分组: 一般组数 k 通常在 5-20 个;

(2) 确定上下限: 左端点 a_0 略小于最小观测值, 右端点 a_k 略大于最大观测值;

(3) 确定每组组距: 组距 d 可定为 $d = \dfrac{a_k - a_0}{k}$;

(4) 确定每组组限: 各组区间端点为 a_0, $a_1 = a_0 + d$, $a_2 = a_0 + 2d$, $\cdots$, $a_k = a_0 + kd$, 形成左开右闭的分组区间 $(a_0, a_1]$, $(a_1, a_2]$, $\cdots$, $(a_{k-1}, a_k]$;

(5) 制表: 统计样本数据落入每个区间的个数—频数, 并列出其频数频率分布表.

直方图

直方图在组距相等场合常用宽度相等的长条矩形表示, 矩形的高低表示频数的大小. 在图形上, 横坐标表示变量的取值区间, 纵坐标可以是频数、频率或频率/组距, 这三者的差别仅在于

纵轴刻度的选择, 直方图本身并无变化.

茎叶图

茎叶图是把每一个数值分为两部分, 前面一部分称为茎, 后面部分称为叶, 然后画一条竖线, 在竖线的左侧写上茎, 右侧写上叶, 就得到了茎叶图. 茎叶图保留了数据中全部的信息. 在要比较两组样本时, 还可以使用背靠背的茎叶图.

1.2 统计量及其分布

1.2.1 统计量

定义 1.2.1 (统计量) 设 $x_1, x_2, \cdots, x_n$ 为取自某总体的样本, 若样本函数 $T = T(x_1, x_2, \cdots, x_n)$ 中不含有任何末知参数, 则称 T 为**统计量**.

尽管统计量与末知参数无关, 但其抽样分布可能与末知参数有关. 常用的统计量有样本均值、样本方差 (标准差)、样本矩及其函数、次序统计量、样本分位数等.

样本均值

定义 1.2.2 (样本均值) 设 $x_1, x_2, \cdots, x_n$ 为取自某总体的样本, 其算术平均值称为样本均值, 一般用 $\overline{x}$ 表示, 即

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

对于分组样本, 样本均值的近似公式为

$$\overline{x} = \frac{xf_1 + x_2 f_2 + \cdots + x_k f_k}{n} \ (n = \sum_{i=1}^{k} f_i),$$

其中 k 为组数, x_i 为第 i 组的组中值, f_1 为第 i 组的频数.

样本均值用来度量样本的中心位置.

性质 1.2.3 (样本均值的性质) (1) 若把样本中的数据与样本均值之差称为**偏差**, 则样本所有偏差之和为 0, 即 $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$;

(2) 样本值与样本均值的**偏差平方和**最小, 即在形如 $\sum_{i=1}^{n}(x_i - c)^2$ 的函数中, $\sum_{i=1}^{n}(x_i - \overline{x})^2$ 最小;

(3) 设总体 X 具有二阶矩, 即 $\mathbb{E}(X) = \mu$, $\mathrm{Var}(X) = \sigma^2 < \infty$, 则 $\mathbb{E}(\overline{x}) = \mu$, $\mathrm{Var}(\overline{x}) = \sigma^2/n$.

样本方差与样本标准差

定义 1.2.4 (样本方差与样本标准差) 设 $x_1, x_2, \cdots, x_n$ 为取自某总体的样本, 则它关于样本均值 $\overline{x}$ 的平均偏差平方和

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

称为样本方差. 其算术根 $s_n = \sqrt{s_n^2}$ 称为样本标准差.

在 n 不大时, 常用

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

作为**样本方差**, 其算术根 $s = \sqrt{s^2}$ 也称为**样本标准差**. 在实际中样本方差通常指的是 s^2.

对于分组数据, 样本方差的近似计算公式是

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i (x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} f_i x_i^2 - n\overline{x}^2\right).$$

样本方差用来度量样本的分散程度. 样本标准差与样本均值具有相同的度量单位

样本方差依赖于偏差平方和 $\sum_{i=1}^{n}(x_i - \overline{x})^2$. $n-1$ 是偏差平方和的自由度, 因为在 $\overline{x}$ 确定后, n 个偏差 $x_1 - \overline{x}$, $x_2 - \overline{x}$, $\cdots$, $x_n - \overline{x}$ 中只有 $n-1$ 个偏差可以自由变动, 而第 n 个则不能自由取值, 这是由于 $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$. 偏差平方和的常用简便公式是

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2.$$

定理 1.2.5 (样本方差 s^2 的期望) 设总体 X 具有二阶矩, 即 $\mathrm{Var}(X) = \sigma^2 < \infty$, x_1, x_2, $\cdots$, x_n 为从该总体得到的样本, 则 $\mathbb{E}\left(s^2\right) = \sigma^2$.

因此样本方差 s^2 是总体方差 σ^2 的无偏估计, 故 s^2 也称为无偏方差.

样本矩及其函数

定义 1.2.6 (样本矩) 设 $x_1, x_2, \cdots, x_n$ 是样本, k 为正整数, 则统计量

$$a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

称为样本 k **阶原点矩**, 特别, 样本一阶原点矩就是样本均值. 统计量

$$b_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k$$

称为样本 k **阶中心矩**. 特别, 样本二阶中心矩就是样本方差.

均值和方差反映了分布的中心和分散程度, 但是无法衡量分布的形状. 常用的统计量还有样本偏度和样本峰度, 它们都是样本中心矩的函数.

定义 1.2.7 (样本偏度) 设 $x_1, x_2, \cdots, x_n$ 是从总体 X 得到的样本, b_k 为样本 k 阶中心矩. 称

$$\widehat{\beta}_s = \frac{b_3}{b_2^{3/2}}$$

为**样本偏度**.

样本偏度 $\widehat{\beta}_s$ 反映了样本数据与对称性的偏离程度和偏离方向. 如果 $\widehat{\beta}_s = 0$, 表示样本对称; 如果 $\widehat{\beta}_s$ 明显大于 0, 表示样本的右尾长, 即样本中有几个较大的数, 总体分布是正偏的或右偏的; 如果 $\widehat{\beta}_s$ 明显小于 0, 表示分布的左尾长, 即样本中有几个特小的数, 总体分布是负偏的或左偏的, 如图 1.2.

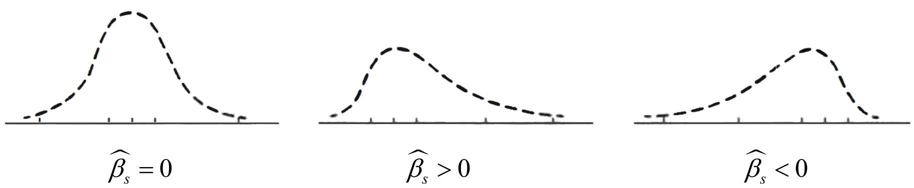

图 1.2: 样本偏度

定义 1.2.8 (样本峰度) 设 $x_1, x_2, \cdots, x_n$ 是从总体 X 得到的样本, b_k 为样本 k 阶中心矩. 称

$$\widehat{\beta}_k = \frac{b_4}{b_2^2} - 3$$

为**样本峰度**.

样本峰度 $\widehat{\beta}_k$ 是反映总体分布密度曲线在其峰值附近的陡峭程度和尾部粗细的统计量. $\widehat{\beta}_k > 0$ 时, 相比正态分布, 该总体中心附近更加陡峭, 尾部更厚, 称为尖顶分布 (尖峰厚尾); 反之, 中心附近平坦, 尾部更细, 称为平顶分布 (平顶薄尾).

补充: 一类重要的统计量

经验分布 $F_n(x)$ 作为函数不含末知参数, 因此是一个统计量.

$F_n(x)$ 是一个分布函数, 它的所有特征数都是统计量. 因为经验分布 $F_n(x)$ 只依赖于样本, 有了样本观察值之后, 它是完全确定的, 所以它的任何特征数不依赖于任何末知参数, 都是统计量. 比如分布 $F_n(x)$ 的均值是样本均值 $\overline{x}$; 分布 $F_n(x)$ 的原点矩和中心矩就是样本原点矩和样本中心矩.

次序统计量

定义 1.2.9 (次序统计量) 设 $x_1, x_2, \cdots, x_n$ 是总体 X 的一个样本. 对它们排序, 得

$$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}.$$

称 $(x_{(1)}, x_{(2)}, \cdots, x_{(n)})$ 为该样本的**次序统计量**. 称 $x_{(i)}$ 为第 i 个**次序统计量**. $x_{(1)}$ 和 $x_{(n)}$ 分别称为最小和最大次序统计量.

在一个 (简单随机) 样本中, x_1, x_2, $\cdots$, x_n 是独立同分布的, 而次序统计量 $x_{(1)}$, $x_{(2)}$, $\cdots$, $x_{(n)}$ 则既不独立, 分布也不相同.

样本中位数与样本分位数也是一个很常见的统计量, 它们是次序统计量的函数.

定义 1.2.10 (样本中位数与样本分位数) **样本中位数**

$$m_{0.5} = \begin{cases} x\left(\dfrac{n+1}{2}\right), & n \text{ 为奇数;} \\ \dfrac{1}{2}\left(x\left(\dfrac{n}{2}\right) + x\left(\dfrac{n}{2}+1\right)\right), & n \text{ 为偶数.} \end{cases}$$

更一般地, **样本 p 分位数**

$$m_p = \begin{cases} x_{([np+1])}, & \text{若 } np \text{ 不是整数;} \\ \dfrac{1}{2}\left(x_{(np)} + x_{(np+1)}\right), & \text{若 } np \text{ 是整数.} \end{cases}$$

定义 1.2.11 (样本分位数的线性插值定义) 样本 p 分位数也可以定义为

$$m_p^* = X_{(k)} + ((n+1)p - k)\left(X_{(k+1)} - X_{(k)}\right),$$

其中 k 表示 $(n+1)p$ 的整数部分. (这个定义与定义 1.2.10 下样本 p 分位数的渐近分布 (定理 1.2.26) 是一样的)

通常, 样本均值在概括数据方面具有一定的优势, 但样本均值受极端数值影响较大. 与之相对应, 中位数则不受极端值的影响. 因此, 当数据中含有极端值时, 使用中位数比使用均值更好. 中位数的这种抗干扰性在统计中称为具有**稳健性**.

次序统计量的应用之一是**五数概括**与**箱线图**.

五数概括指的是用五个数描: 最小观测值 $x_{\min} = x_{(1)}$, 最大观测值 $x_{\max} = x_{(n)}$, 中位数 $m_{0.5}$, 第一 4 分位数 $Q_1 = m_{0.25}$ 和第三 4 分位数 $Q_3 = m_{0.75}$ 来大致描述一批数据的轮廓.

五数概括的图形表示称为**箱线图**, 由箱子和线段组成, 其作法如下:

(1) 画一个箱子, 其两侧恰为第一 4 分位数和第三 4 分位数, 在中位数位置上画一条竖线, 它在箱子内. 这个箱子包含了样本中 50% 的数据;

(2) 在箱子左右两侧各引出一条水平线, 分别至最小值和最大值为止. 每条线段包含了样本中 25% 的数据.

箱线图可用来对样本数据分布的形状进行大致的判断, 图 1.3 是几个常见的箱线图, 分别对应左偏分布、对称分布和右偏分布.

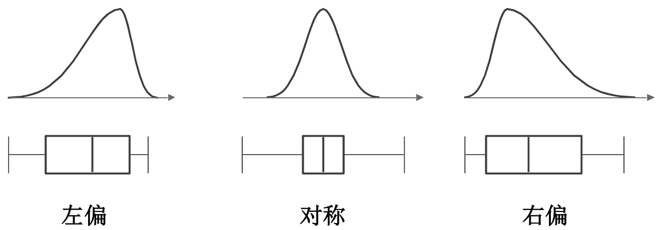

图 1.3: 箱线图及其对应的分布轮廓

1.2.2 三大抽样分布

χ^2 分布

定义 1.2.12 (χ^2 分布) 设 $X_1, X_2, \cdots, X_n$ 独立同分布于标准正态分布 $N(0,1)$, 则 $\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$ 的分布称为自由度为 n 的 χ^2 **分布**, 记为 $X^2 \sim \chi^2(n)$.

$\chi^2(n)$ 就是 $Ga(n/2, 1/2)$, 其密度函数是

$$p(x) = \frac{2^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \ (x > 0).$$

χ^2 分布是非对称的, 其可能取值是所有正数. 自由度 $n = 2$ 的 χ^2 分布 $\chi^2(2)$ 就是指数分布 $\mathrm{Exp}(1/2)$. 当自由度 $n > 2$ 时, χ^2 分布的密度函数曲线都是单峰曲线, 在 $n - 2$ 处取得峰值, 当 n 越大, 峰向右移动, 曲线变平缓, 如图 1.4.

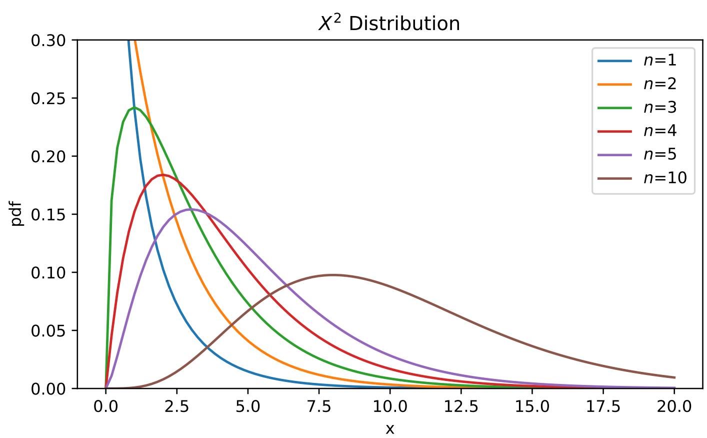

图 1.4: $\chi^2(n)$ 分布的密度函数

性质 1.2.13 (χ^2 分布的性质) (1) 若 $X \sim \chi^2(n)$, 则 $\mathbb{E}X = n$, $\mathrm{Var}X = 2n$;
(2) $\chi^2(n) * \chi^2(m) = \chi^2(n+m)$, 其中 $*$ 表示卷积.

F 分布

定义 1.2.14 (F 分布) 设随机变量 $X \sim \chi^2(n), Y \sim \chi^2(m)$, X 与 Y 独立, 则称 $F = \dfrac{X/n}{Y/m}$ 的分布是自由度为 n 与 m 的 F **分布**, 记为 $F \sim F(n,m)$, 其中 n 称为分子自由度, m 称为分母自由度.

$F(n,m)$ 的密度函数为

$$p(x) = \left(\frac{n}{m}\right)^{\frac{m}{2}} \frac{\left(1+\frac{n}{m}z\right)^{-\frac{m+n}{2}} z^{\frac{n}{2}-1}}{B\left(\frac{n}{2},\frac{m}{2}\right)} \ (x>0).$$

F 分布的密度函数的图像是一个只取非负值的分布, 当分子的自由度为 1 或 2 时, 其密度函数是单调递减函数, 其他情况下密度函数呈单峰的右偏分布, 如图 1.5.

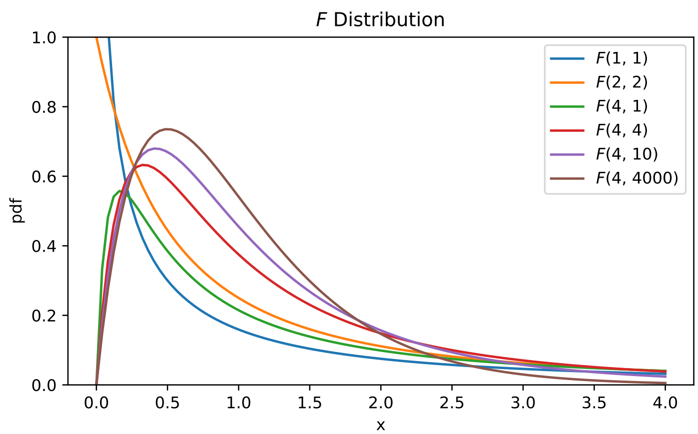

图 1.5: F 分布的密度函数

性质 1.2.15 (F 分布的性质) (1) $m>2$ 时, F 分布的数学期望存在, 且为 $m/(m-2)$;

(2) $m>4$ 时, F 分布的方差存在, 且为 $\dfrac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$;

(3) 若 $F \sim F(n,m)$, 则 $\dfrac{1}{F} \sim F(m,n)$, 进而若记分布 $F(n,m)$ 的 α 分位数为 $F_\alpha(n,m)$, 则 $F_\alpha(n,m) = \dfrac{1}{F_{1-\alpha}(m,n)}$.

t 分布

定义 1.2.16 (t 分布) 设随机变量 X_1 与 X_2 独立且 $X_1 \sim N(0,1)$, $X_2 \sim \chi^2(n)$, 则称 $t = \dfrac{X_1}{\sqrt{X_2/n}}$ 的分布为自由度为 n 的 t **分布**, 记为 $t \sim t(n)$.

$t(n)$ 分布的密度函数是

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

t 分布的密度函数的图像是一个关于纵轴对称的分布, 与标准正态分布的密度函数形状类似, 只是峰比标准正态分布低一些, 尾部的概率比标准正态分布的厚一些, 且随着自由度增加, t 分布的曲线更加接近标准正态分布的曲线, 如图 1.6.

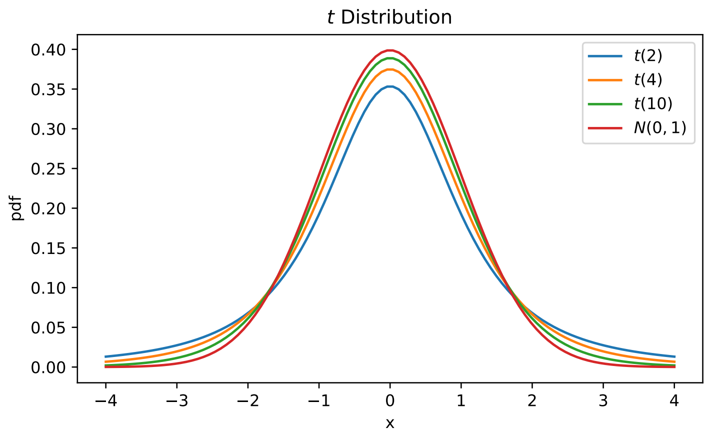

图 1.6: t 分布的密度函数

性质 1.2.17 (t 分布的性质) (1) t 分布是关于 y 轴对称的, 故若记 $t(n)$ 的 α 分位数为 $t_\alpha(n)$, 那么 $-t_\alpha(n)=t_{1-\alpha}(n)$;

(2) 自由度为 1 的 t 分布就是标准柯西分布, 它的均值不存在;

(3) $n>1$ 时, t 分布的数学期望存在且为 0;

(4) $n>2$ 时, t 分布的方差存在, 且为 $n/(n-2)$;

(5) 当自由度 n 较大 (如 $n\geqslant 30$) 时, t 分布可以用 $N(0,1)$ 近似;

(6) 若 $t\sim t(n)$, 则 $t^2\sim F(1,n)$.

三大抽样分布的基本性质及其统计量构造如表 1.1, 其中 $x_1,x_2,\cdots,x_n$ 和 $y_1,y_2,\cdots,y_m$ 是来自标准正态分布的两个相互独立的样本.

1.2.3 统计量的抽样分布

定义 1.2.18 (抽样分布) 统计量的分布称为**抽样分布**.

表 1.1: 三大抽样分布

分布	统计量构造	密度函数	期望	方差
$\chi^2(n)$	$\chi^2=x_1^2+x_2^2+\cdots+x_n^2$	$\dfrac{2^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}$	n	$2n$
$F(n,m)$	$F=\dfrac{(x_1^2+x_2^2+\cdots+x_n^2)/n}{(y_1^2+y_2^2+\cdots+y_m^2)/m}$	$\left(\dfrac{n}{m}\right)^{\frac{m}{2}}\dfrac{\left(1+\frac{n}{m}z\right)^{-\frac{m+n}{2}}z^{\frac{n}{2}-1}}{B\left(\frac{n}{2},\frac{m}{2}\right)}$	$\dfrac{m}{m-2}\ (m>2)$	$\dfrac{2m^2\left(n+m-2\right)}{n\left(m-2\right)^2\left(m-4\right)}\ (m>4)$
$t(n)$	$t=\dfrac{y_1}{\sqrt{(x_1^2+x_2^2+\cdots+x_n^2)/n}}$	$\dfrac{1}{\sqrt{n}B\left(\frac{1}{2},\frac{n}{2}\right)}\left(1+\dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$	$0\ (n>1)$	$\dfrac{n}{n-2}\ (n>2)$

样本均值的抽样分布

定理 1.2.19 (样本均值的抽样分布) 设 $x_1,x_2,\cdots,x_n$ 是来自某总体 X 的样本, $\overline{x}$ 为样本均值.

(1) 若总体分布为 $N\left(\mu,\sigma^2\right)$, 则 $\overline{x}$ 的精确分布是 $N\left(\mu,\sigma^2/n\right)$;

(2) 若总体分布末知, 但 $\mathbb{E}\left(X\right)=\mu,\mathrm{Var}\left(X\right)=\sigma^2$ 有限, 则当 n 较大时, $\overline{x}$ 的渐近分布为 $N\left(\mu,\sigma^2/n\right)$, 常记作 $\overline{x}\dot{\sim}N\left(\mu,\sigma^2/n\right)$.

样本方差的抽样分布

定理 1.2.20 (正态总体样本方差的抽样分布) 设 $x_1,x_2,\cdots,x_n$ 是来自正态总体 $N\left(\mu,\sigma^2\right)$ 的样本, 其样本均值 $\overline{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$, 样本方差 $s^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}\left(x_i-\overline{x}\right)^2$. 则

(1) $\overline{x}$ 与 s^2 相互独立;

(2) $\dfrac{(n-1)s^2}{\sigma^2}=\dfrac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}{\sigma^2}\sim\chi^2(n-1)$.

定理 1.2.21 (两独立正态总体样本方差之比的分布) 设 $x_1,\cdots,x_n$ 是来自总体 $N(\mu_1,\sigma_1^2)$ 的样本, $y_1,\cdots,y_m$ 是来自总体 $N\left(\mu_2,\sigma_2^2\right)$ 的样本, 且这两个样本独立. 它们的样本方差分别是 s_x^2 和 s_y^2. 那么

$$F=\frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2}\sim F(n-1,m-1).$$

当 $\sigma_1^2=\sigma_2^2$ 时,

$$F=\frac{s_x^2}{s_y^2}\sim F(n-1,m-1).$$

定理 1.2.22 (正态总体样本均值与样本标准差之比的分布) 设 $x_1,x_2,\cdots,x_n$ 是来自正态总体 $N\left(\mu,\sigma^2\right)$ 的一个样本, $\overline{x}$ 与 s 分别是其样本均值与样本标准差, 则有

$$t=\frac{\sqrt{n}(\overline{x}-\mu)}{s}\sim t(n-1).$$

次序统计量的抽样分布

如果总体是离散的, 那么次序统计量的分布一般比较复杂, 没有统一的表达式. 但对于连续总体, 次序统计量的分布具有简洁的表达式.

定理 1.2.23 (单个次序统计量的抽样分布) 设总体 X 的密度函数是 $p(x)$, 分布函数是 $F(x)$. $x_1, x_2, \cdots, x_n$ 是样本, 则第 k 个次序统计量 $x_{(k)}$ 的密度函数是

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} (1-F(x))^{n-k} p(x).$$

定理 1.2.24 (两个次序统计量的联合分布) 设总体 X 的密度函数是 $p(x)$, 分布函数是 $F(x)$. $x_1, x_2, \cdots, x_n$ 是样本, 次序统计量 $\left(x_{(i)}, x_{(j)}\right)$ $(i<j)$ 的联合分布的密度函数为

$$p_{ij}(y, z) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F(y)^{i-1} (F(z)-F(y))^{j-i-1} (1-F(z))^{n-j} p(y) p(z) \quad (y \leqslant z).$$

例 1.2.25 设总体分布是均匀分布 $U(0,1)$, 则其第 k 个次序统计量 $x_{(k)}$ 服从 Beta 分布 $\mathrm{Beta}(k, n-k+1)$, 且 $\mathbb{E}\left(X_{(k)}\right) = \dfrac{k}{n+1}$. 样本极差 $R_n = X_{(n)} - X_{(1)}$ 服从 Beta 分布 $\mathrm{Beta}(n-1, 2)$.

定理 1.2.26 (样本分位数的渐近分布) 设总体密度函数为 $p(x)$, x_p 为其 p 分位数, $p(x)$ 在 x_p 处连续且 $p\left(x_p\right)>0$, 则当 $n \rightarrow \infty$ 时样本 p 分位数 m_p 的渐近分布为

$$m_p \dot{\sim} \left(x_p, \frac{p(1-p)}{n \cdot p^2\left(x_p\right)}\right).$$

特别, 对样本中位数, 当 $n \rightarrow \infty$ 时近似地有

$$m_{0.5} \dot{\sim} N\left(x_{0.5}, \frac{1}{4n \cdot p^2\left(x_{0.5}\right)}\right).$$

1.2.4 充分统计量

充分统计量的直观含义是某个统计量包含了样本中关于感兴趣问题的所有信息, 统计上将这种 “样本加工不损失信息” 称作充分性.

定义 1.2.27 (充分统计量) T 称为关于 θ 的**充分统计量**, 如果样本 $x_1, x_2, \cdots, x_n$ 在给定统计量 T 时的条件分布 $F_\theta\left(x_1, \cdots, x_n \mid T\right)$ 与 θ 无关.

充分性原则: 在充分统计量存在的场合, 任何统计推断都可以基于充分统计量进行, 这可以简化统计推断的程序.

一个简单的判断统计量是否充分的办法是因子分解定理. 称 $f(x)$ 为随机变量 X 的**概率函数**: 在连续场合, $f(x)$ 表示 X 的概率密度; 在离散的场合, $f(x)$ 表示 X 的概率分布列.

定理 1.2.28 (因子分解定理) 设总体概率函数为 $f(x;\theta)$, $x_1, x_2, \cdots, x_n$ 为样本, 则 $T=T(x_1, x_2, \cdots, x_n)$ (T 可以是一维的, 也可以是多维的) 为充分统计量的充分必要条件是: 存在两个函数 $g(t,\theta)$ 和 $h\left(x_1, x_2, \cdots, x_n\right)$ 使得对任意的 θ 和任一组观测值 $x_1, x_2, \cdots, x_n$, 有

$$f\left(x_1, x_2, \cdots, x_n ; \theta\right)=g\left(T\left(x_1, x_2, \cdots, x_n\right), \theta\right) h\left(x_1, x_2, \cdots, x_n\right),$$

其中 $g(t,\theta)$ 是通过统计量 T 的取值而依赖于样本的.

正态总体下, $\left(\bar{x}, s^2\right)$ 是 $\theta=\left(\mu, \sigma^2\right)$ 的充分统计量.

定理 1.2.29 若 T 是充分统计量, 且存在某个函数 H, 使得 $T(\boldsymbol{x})=H(S(\boldsymbol{x}))$, 则统计量 S 也是充分统计量.

推论 1.2.30 设 $T=T(\boldsymbol{x})$ 是参数 θ 的充分统计量, $\Psi(\cdot)$ 是严格单调函数, 则 $S(\boldsymbol{x})=\Psi(T(\boldsymbol{x}))$ 也是 θ 的一个充分统计量.

定义 1.2.31 (极小充分统计量) 设 T^* 是一个充分统计量, 若对任意的充分统计量 T, 都存在函数 h, 使得 $T^*(\boldsymbol{x})=h(T(\boldsymbol{x}))$, 则 T^* 称为一个极小充分统计量. 极小充分统计量之间存在一一对应关系.

通常, 因子分解定理那个充分统计量就是极小充分统计量.

1.2.5 完备统计量

定义 1.2.32 (完备统计量) 设 $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)$ 是来自总体 $p(x;\theta)$ 的一个样本, $T=T(\boldsymbol{x})$ 为一统计量 (不必是充分的). 若对任何满足条件

$$\mathbb{E}_\theta(g(T(\boldsymbol{x})))=0\ (\forall \theta \in \Theta)$$

的 $g(T)$, 都有

$$P_\theta(g(T(\boldsymbol{x}))=0)=1\ (\forall \theta \in \Theta),$$

则称 T 是一个**完备统计量**.

统计量 T 的完备性不仅取决于 T, 还取决于样本 $\boldsymbol{x}$ 的分布. 完备统计量的函数亦是完备的, 但反之不真.

Chapter 2

参数估计

2.1 点估计

定义 2.1.1 (点估计) 设 $x_1, x_2, \cdots, x_n$ 是来自总体的一个样本, 用于估计末知参数 θ 的统计量 $\widehat{\theta} = \widehat{\theta}(x_1, x_2, \cdots, x_n)$ 称为 θ 的估计量, 或称为 θ 的**点估计**, 简称估计.

2.1.1 估计的评价标准

无偏性

定义 2.1.2 (无偏估计) 设参数 θ 的参数空间为 Θ. 设 $\widehat{\theta} = \widehat{\theta}(x_1, \cdots, x_n)$ 是 θ 的一个估计. 如果对于任意 $\theta \in \Theta$, 有

$$\mathbb{E}_\theta(\widehat{\theta}) = \theta,$$

则称 $\widehat{\theta}$ 为 θ 的**无偏估计**, 否则称为**有偏估计**.

无偏性等价于 $\mathbb{E}_\theta(\widehat{\theta} - \theta) = 0$, 表示该估计无系统偏差. 无偏性意味着, 当把重复抽样很多次之后, 所有的估计值的平均等于真实参数值.

对于任何总体, 样本均值 $\overline{x}$ 是总体均值 $\mathbb{E}(X)$ 的无偏估计. 当总体 k 阶 (原点) 矩存在时, 样本 k 阶原点矩

$$a_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k$$

是总体 k 阶 (原点) 矩 $\mu_k = \mathbb{E}\left(X^k\right)$ 的无偏估计. 样本方差

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

是总体方差 σ^2 的无偏估计.

不是所有的参数都存在无偏估计. 例如总体为二点分布 $b(1,p)$ $(0<p<1)$, $x_1,\cdots,x_n$ 是样本, 那么参数 $\theta=1/p$ 不存在无偏估计. 有偏估计并非一定不好.

无偏性不具有不变性, 即当 $\widehat{\theta}$ 是 θ 的无偏估计时, $g(\widehat{\theta})$ 一般不是 $g(\theta)$ 的无偏估计, 除非 $g(\theta)$ 是 θ 的线性函数 (**线性不变性**).

定义 2.1.3 (渐近无偏估计) 设 $\widehat{\theta}_n=\widehat{\theta}_n\left(x_1,\cdots,x_n\right)$ 是 θ 的一个估计. 如果对于任意 $\theta\in\Theta$, 有

$$\lim_{n\to\infty}\mathbb{E}_\theta\left(\widehat{\theta}_n\right)=\theta,$$

则称 $\widehat{\theta}_n$ 为 θ 的**渐近无偏估计**.

渐近无偏性是好的估计量要满足的一个**基本要求**.

s_n^2 和 s 分别是 σ^2 和 σ 的渐近无偏估计.

有效性

定义 2.1.4 (有效性) 设 $\widehat{\theta}_1$ 和 $\widehat{\theta}_2$ 是 θ 的两个无偏估计. 如果对于任意 θ, 有

$$\operatorname{Var}_\theta\left(\widehat{\theta}_1\right)\leqslant\operatorname{Var}_\theta\left(\widehat{\theta}_2\right),$$

且至少对于一个 θ 不等号严格成立, 则称 $\widehat{\theta}_1$ 比 $\widehat{\theta}_2$ 更有效.

相合性

定义 2.1.5 (相合估计) 设 $\theta\in\Theta$ 为末知参数, $\widehat{\theta}_n=\widehat{\theta}_n\left(x_1,\cdots,x_n\right)$ 为 θ 的一个估计量, n 为样本量. 如果对任意的 $\varepsilon>0$, 有

$$\lim_{n\to\infty}P(|\widehat{\theta}-\theta|>\varepsilon)=0,$$

则称 $\widehat{\theta}$ 是参数 θ 的相合估计.

相合性也是对一个合理的估计的**基本要求**, 不满足相合性的估计量一般不予考虑.

参数的相合估计有很多, 如在正态总体 $N\left(\mu,\sigma^2\right)$ 下, $\bar{x}$ 是 μ 的相合估计; s^2 和 s_n^2 都是 σ^2 的相合估计.

定理 2.1.6 (相合性的充分条件) 设 $\widehat{\theta}_n=\widehat{\theta}_n\left(x_1,\cdots,x_n\right)$ 是参数 θ 的一个估计量, 若

$$\lim_{n\to\infty}\mathbb{E}\left(\widehat{\theta}_n\right)=\theta,\quad\lim_{n\to\infty}\operatorname{Var}\left(\widehat{\theta}_n\right)=0,$$

则 $\widehat{\theta}_n$ 是 θ 的相合估计.

定理 2.1.7 设 $\eta=g\left(\theta_1,\cdots,\theta_k\right)$ 是 $\left(\theta_1,\cdots,\theta_k\right)$ 的连续函数. 如果 $\widehat{\theta}_{n1}$, $\cdots$, $\widehat{\theta}_{nk}$ 分别是 θ_1, $\cdots$, θ_k 的相合估计, 则 $\widehat{\eta}_n=g\left(\widehat{\theta}_{n1},\cdots,\widehat{\theta}_{nk}\right)$ 是 η 的相合估计.

渐近正态性

定义 2.1.8 (渐近正态性) 参数 θ 的估计 $\widehat{\theta}_n$ 称为**渐近正态**的, 若存在序列 $\sigma_n(\theta) > 0$, 使得

$$\frac{\widehat{\theta}_n - \theta}{\sigma_n(\theta)} \xrightarrow{d} (\rightsquigarrow) N(0,1).$$

有时也称 $\widehat{\theta}_n$ 渐近地服从正态分布 $N\left(\theta, \sigma_n^2(\theta)\right)$, 记为

$$\widehat{\theta}_n \sim AN\left(\theta, \sigma_n^2(\theta)\right).$$

并将 $\sigma_n^2(\theta)$ 称为 $\widehat{\theta}_n$ 的渐近方差. 趋于零的数列 $\sigma_n(\theta)$ 表示 $\widehat{\theta}_n$ 依概率收敛到 θ 的速度.

渐近正态的估计是相合的. 渐近正态性常被用来对不同的相合估计进行比较, 主要比较其渐近方差大小. 渐近方差一般 $\neq$ 通常意义的方差.

最小方差无偏估计

相合性和渐近正态性 (主要是渐近方差) 是评价一个估计量在大样本 (即极限) 情况下优良性的两个重要标准. 基于小样本的评价标准最常用的评价指标是均方误差 (MSE).

定义 2.1.9 (均方误差) 参数 θ 的一个估计 $\widehat{\theta}$ 的**均方误差**定义为

$$\mathrm{MSE}(\widehat{\theta}) = \mathbb{E}(\widehat{\theta} - \theta)^2.$$

均方误差越小的估计越好.

均方误差有分解

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{Var}(\widehat{\theta}) + (\mathbb{E}\widehat{\theta} - \theta)^2.$$

均方误差由点估计的方差和偏差 $|\mathbb{E}(\widehat{\theta}) - \theta|$ 的平方两部分组成. 相比方差和偏差, 均方误差是更全面的评价标准. 对于无偏估计 $\widehat{\theta}$, 有 $\mathrm{MSE}(\widehat{\theta}) = \mathrm{Var}(\widehat{\theta})$, 此时, 有效性等价于均方误差越小越好的原则. 对于有偏估计, 其优良性不仅依赖于方差也依赖于偏差. 均方误差下某些有偏估计会优于无偏估计.

定义 2.1.10 (一致最小均方误差估计) 设有样本 $x_1, x_2, \cdots, x_n$, 对待估参数 θ, 设有一个估计类 Ω, 如果来自该类的点估计 $\widehat{\theta} = \widehat{\theta}(x_1, \cdots, x_n) \in \Omega$ 满足, 对于任意 θ 的值和任意估计 $\widetilde{\theta} \in \Omega$ 有

$$\mathrm{MSE}_\theta(\widehat{\theta}) \leqslant \mathrm{MSE}_\theta(\widetilde{\theta}).$$

称 $\widehat{\theta}$ 是该估计类中 θ 的**一致最小均方误差估计**.

不对估计做限制时一致最小均方误差估计通常不存在. 为使它存在, 可对估计提一些合理的限制, 常用的要求的是无偏性. 对于无偏估计, 均方误差等价于方差, 一致最小均方误差无偏估计等价于一致最小方差无偏估计.

定义 2.1.11 (一致最小方差无偏估计) 设有样本 $x_1, x_2, \cdots, x_n$, θ 是待估参数, Θ 是参数空间, $\widehat{\theta}$ 是 θ 的一个无偏估计. 如果对于任何无偏估计 $\widetilde{\theta}$, 对于任何 $\theta \in \Theta$ 都有

$$\mathrm{Var}_\theta(\widehat{\theta}) \leqslant \mathrm{Var}_\theta(\widetilde{\theta}).$$

称 $\widehat{\theta}$ 是该估计类中 θ 的**一致最小方差无偏估计**, 简记为 UMVUE.

定理 2.1.12 (UMVUE 的判断准则) 设 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 是来自某总体的一个样本, θ 是待估参数, Θ 是参数空间. 设 $\widehat{\theta}$ 是 θ 的一个无偏估计且 $\mathrm{Var}(\widehat{\theta}) < \infty$. 那么 $\widehat{\theta}$ 是 UMVUE 的充要条件是, 对于任意一个满足 $\mathbb{E}\{\varphi(\boldsymbol{x})\} = 0$ 和 $\mathrm{Var}\{\varphi(\boldsymbol{x})\} < \infty$ 的统计量 $\varphi(\boldsymbol{x})$, 都有

$$\mathrm{Cov}_\theta(\widehat{\theta}(\boldsymbol{x}), \varphi(\boldsymbol{x})) = \mathbb{E}_\theta(\widehat{\theta}(\boldsymbol{x})\varphi(\boldsymbol{x})) = 0 \ (\forall \theta \in \Theta).$$

这表明 θ 的 UMVUE 必与任一零的无偏估计不相关, 反之亦然.

性质 2.1.13 (UMVUE 的性质) (1) 设 $\widehat{\theta}_1$ 和 $\widehat{\theta}_2$ 分别是 θ_1 和 θ_2 的 UMVUE, 那么对于任意常数 c_1 和 c_2, 有 $c_1\widehat{\theta}_1 + c_2\widehat{\theta}_2$ 是 $c_1\theta_1 + c_2\theta_2$ 的 UMVUE;

(2) 设 $\widehat{\theta}_1$ 和 $\widehat{\theta}_2$ 都是 θ 的 UMVUE, 那么 $P\left(\widehat{\theta}_1 = \widehat{\theta}_2\right) = 1$.

定理 2.1.14 (充分性原则) 设总体概率函数是 $p(x, \theta)$, 设 $x_1, \cdots, x_n$ 是来自该总体的 iid 样本, $T = T(x_1, \cdots, x_n)$ 是充分统计量, 则对 θ 的任意无偏估计 $\widehat{\theta} = \widehat{\theta}(x_1, \cdots, x_n)$, 令 $\widetilde{\theta} = \mathbb{E}(\widehat{\theta} \mid T)$, 那么 $\widetilde{\theta}$ 也是 θ 的无偏估计, 且

$$\mathrm{Var}(\widetilde{\theta}) \leqslant \mathrm{Var}(\widehat{\theta}).$$

定理表明: 如果一个无偏估计不是充分统计量的函数, 那么将它在充分统计量给定的条件下求期望, 可得到一个新的无偏估计, 而且该估计的方差不超过原估计的方差. 等价地, 考虑 θ 的估计问题时, 只需要考察充分统计量的函数即可. 这种说法对于所有的统计推断问题都是正确的, 这就是所谓的充分性原则.

设 T 为一充分统计量, 则若 $g(\theta)$ 的 UMVUE 存在, 则它必能表为 T 的函数. 如果进一步知道能表为 T 的函数的 $g(\theta)$ 的无偏估计唯一, 则它就是 $g(\theta)$ 的 UMVUE. 事实上, 当且仅当充分统计量 T 还是完备的时, 能表为 T 的函数的 $g(\theta)$ 的无偏估计才唯一.

定理 2.1.15 设 $\boldsymbol{x} = (x_1, \cdots, x_n)$ 是来自总体 $p(x; \theta)$ 的样本, 设 $T = T(\boldsymbol{x})$ 为一个完备充分统计量, $\widehat{g}(T(\boldsymbol{x}))$ 为 $g(\theta)$ 的一个无偏估计, 满足 $\mathrm{Var}_\theta(\widehat{g}(T(\boldsymbol{x}))) < \infty, \forall \theta \in \Theta$, 则 $\widehat{g}(T(\boldsymbol{x}))$ 是 $g(\theta)$ 的唯一的 UMVUE (唯一性是指: 若 $\widehat{g}$ 和 $\widehat{g}_1$ 都是 $g(\theta)$ 的 UMVUE, 则 $P_\theta(\widehat{g} \neq \widehat{g}_1) = 0$).

定义 2.1.16 (费希尔信息量) 设总体 X 有密度函数 $p(x; \theta)$, $\theta \in \Theta$. 假设

$$I(\theta) = \int_{-\infty}^{\infty} \left(\frac{\partial \log p(x; \theta)}{\partial \theta}\right)^2 p(x; \theta) dx = \mathbb{E}\left(\frac{\partial \log p(X; \theta)}{\partial \theta}\right)^2$$

对于任意 $\theta \in \Theta$ 都存在, 称 $I(\theta)$ 为参数 θ 的**费希尔 (Fisher) 信息量**.

多维参数的费希尔信息量是费希尔信息矩阵, 其定义为:

$$I(\theta)=\int \frac{\partial \log p(x;\theta)}{\partial \theta}\frac{\partial \log p(x;\theta)}{\partial \theta^T}p(x;\theta)dx=\mathbb{E}\left(\frac{\partial \log p(X;\theta)}{\partial \theta}\frac{\partial \log p(X;\theta)}{\partial \theta^T}\right).$$

性质 2.1.17 (费希尔信息量的性质) 若 $p(x;\theta)$ 满足:

(1) 参数区间 Θ 是直线上的一个开区间;

(2) 支撑 $S=\{x:p(x;\theta)>0\}$ 与 θ 无关;

(3) 导数 $\dfrac{\partial p(x;\theta)}{\partial \theta}$ 对一切 $\theta\in\Theta$ 都存在;

(4) 积分与微分运算可交换次序, 即对任何 θ,

$$\frac{\partial}{\partial \theta}\int_{-\infty}^{\infty}p(x;\theta)dx=\int_{-\infty}^{\infty}\frac{\partial}{\partial \theta}p(x;\theta)dx$$

成立;

(5) 期望 $\mathbb{E}_\theta\left(\dfrac{\partial \log p(X;\theta)}{\partial \theta}\right)^2$ 存在, (条件 (1)-(5) 称为 **C-R 正则性**)

则

$$I(\theta)=\mathbb{E}\left(\frac{\partial \ln p(X;\theta)}{\partial \theta}\right)^2=-\mathbb{E}\left(\frac{\partial^2 \ln p(X;\theta)}{\partial \theta^2}\right).$$

$I(\theta)$ 越大说明每个样本所包含的末知参数 θ 的信息越多.

定义 2.1.18 (积分和微商的可交换条件) 设 $T(x_1,\cdots,x_n)$ 是 $g(\theta)$ 的一个无偏估计, 导数 $g'(\theta)=\dfrac{dg(\theta)}{d\theta}$ 存在, 对于任意 $\theta\in\Theta$, 对

$$g(\theta)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T\left(x_1,\cdots,x_n\right)\prod_{i=1}^{n}p\left(x_i;\theta\right)dx_1\cdots dx_n$$

的微商可在积分号下进行, 即

$$\begin{aligned}g'(\theta)&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T\left(x_1,\cdots,x_n\right)\frac{\partial}{\partial \theta}\prod_{i=1}^{n}p\left(x_i;\theta\right)dx_1\cdots dx_n\\&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T\left(x_1,\cdots,x_n\right)\left(\frac{\partial}{\partial \theta}\ln\left(\prod_{i=1}^{n}p\left(x_i;\theta\right)\right)\right)\prod_{i=1}^{n}p\left(x_i;\theta\right)dx_1\cdots dx_n.\end{aligned}$$

对于离散总体, 将上述积分改为求和.

定理 2.1.19 (C-R 不等式) 假设总体概率函数 $p(x;\theta)$ 满足 C-R 正则性条件. 设 $x_1,\cdots,x_n$ 是来自该总体的样本, $T=T\left(x_1,\cdots,x_n\right)$ 是 $g(\theta)$ 的任何一个无偏估计, 导数 $g'(\theta)=\dfrac{dg(\theta)}{d\theta}$ 存在, 积分和微商可交换条件成立. 则有

$$\mathrm{Var}(T)\geqslant\frac{g'(\theta)^2}{nI(\theta)}.$$

$g'(\theta)^2/(nI(\theta))$ 称为 $g(\theta)$ 的无偏估计的方差的 **C-R 下界**, 简称 $g(\theta)$ 的 C-R 下界. 特别, 对 θ 的无偏估计 $\widehat{\theta}$, 有 $\mathrm{Var}(\widehat{\theta})\geqslant(nI(\theta))^{-1}$.

若估计 T 使得 C-R 不等式的等号成立, 则称 T 是 $g(\theta)$ 的**有效估计**. 有效估计一定是 UMVUE.

2.1.2 矩估计

定义 2.1.20 (替换原理和矩法估计) **替换原理**:
- 用样本矩去替换总体矩 (矩可以是原点矩也可以是中心矩);
- 用样本矩的函数去替换相应的总体矩的函数.

这一方法所给出的估计称作**矩法估计**.

矩法估计的实质是用经验分布函数去替换总体分布, 计算相应的特征数. 其合理性的理论基础是格利文科定理.

根据这一原理, 总体分布末知时也可以对各种参数进行估计:
- 用样本均值 $\overline{x}$ 和方差 s^2 分别估计总体均值 $\mathbb{E}(X)$ 和方差 $\mathrm{Var}(X)$;
- 用样本 p 分位数估计总体 p 分位数, 用样本中位数估计总体中位数;
- 用样本偏度估计总体偏度, 样本峰度估计总体峰度.

概率函数已知时, 设总体具有已知的概率函数 $f(x;\theta_1,\cdots,\theta_k)$, $(\theta_1,\cdots,\theta_k)\in\Theta$ 是末知参数; 设总体的 k 阶原点矩 μ_k 存在; 设 $x_1,\cdots,x_n$ 是样本, 记前 k 阶样本原点矩为 a_j; 如果 $\theta_1,\cdots,\theta_k$ 可以表示成 $\mu_1,\cdots,\mu_k$ 的函数 $\theta_j(\mu_1,\cdots,\mu_k)$, 那么:
- θ_j 的矩估计是 $\widehat{\theta}_j=\theta_j(a_1,\cdots,a_k)$, $j=1,2,\cdots,k$;
- 若关心的是 $\theta_1,\cdots,\theta_k$ 的函数 $\eta=g(\theta_1,\cdots,\theta_k)$, 则 η 的矩估计是 $\widehat{\eta}=g\left(\widehat{\theta}_1,\cdots,\widehat{\theta}_k\right)$.

矩估计可能不唯一, 通常应当尽量选择低阶矩为参数构造矩估计. 矩估计一般都是相合的.

2.1.3 最大似然估计

基本逻辑

定理 2.1.21 (K-L 不等式) 设 $p(x)$, $q(x)$ 是两个概率函数且 $X\sim p(x)$, 则

$$\mathbb{E}\log q(X)\leqslant\mathbb{E}\log p(X),$$

且等号仅在 $p=q$ 时成立.

最大似然估计的基本逻辑:

(1) 设 $X\sim p(x,\theta)$, 其中 θ 末知;

(2) 则根据 K-L 不等式,

$$l(u):=\mathbb{E}_\theta\log p(X;u)\leqslant\mathbb{E}_\theta\log p(X;\theta)=l(\theta),$$

形式上, 函数 $l(u)$ 在 $u=\theta$ 处取得极大值:

$$\theta=\operatorname{Arg}\max\{l(u)\};$$

(3) 估计 $l(u)$ 得到

$$\widehat{l}(u)=\frac{1}{n}\sum_{i=1}^{n}\log p\left(X_i;u\right),$$

并将其最大化:

$$\widehat{\theta}=\operatorname{Arg}\max_u\sum_{i=1}^{n}\log p\left(X_i;u\right)=\operatorname{Arg}\max_u\prod_{i=1}^{n}p\left(X_i;u\right).$$

最大似然估计的定义

定义 2.1.22 (最大似然估计) 设总体的概率函数为 $p(x;\theta)$, 其中 $\theta\in\Theta$ 是一维或多维末知参数, Θ 是参数空间. 设 $x_1,\cdots,x_n$ 是来自该总体的样本. 样本的联合概率函数看成 θ 的函数, 用 $L\left(\theta;x_1,\cdots,x_n\right)$ 表示, 简记为 $L(\theta)$,

$$L(\theta)=L\left(\theta;x_1,\cdots,x_n\right)=\prod_{i=1}^{n}p\left(x_i;\theta\right)$$

称 $L(\theta)$ 为参数 θ 的**似然函数**.

如果某统计量 $\widehat{\theta}=\widehat{\theta}\left(x_1,\cdots,x_n\right)$ 满足

$$L(\widehat{\theta})=\max_{\theta\in\Theta}L(\theta),$$

则称 $\widehat{\theta}$ 是 θ 的**最大似然估计**, 简记为 MLE.

称 $\ell(\theta)=\ln(L(\theta))$ 为 θ 的对数似然函数. 由于 $\ln(x)$ 是 x 的单调函数, 所以最大似然估计可以通过最大化 $\ell(\theta)$ 或者 $L(\theta)$ 得到.

当 $L(\theta)$ 是可微函数时, 求导是求最大似然的最常用方法, 此时基于 $\ell(\theta)$ 求最大似然估计要更简单. 如果似然函数 $L(\theta)$ 乘以一个与 θ 无关的量, 或者对数似然 $\ell(\theta)$ 加上一个 θ 无关的量, 不会影响最大似然估计. 所以在计算时, 可剔除与 θ 无关的量.

定理 2.1.23 (最大似然估计的不变性) 如果 $\widehat{\theta}$ 是 θ 的最大似然估计, 则对任何函数 $g(\theta)$, 其最大似然估计是 $g(\widehat{\theta})$. 这个性质称作最大似然估计的**不变性**.

定理 2.1.24 (最大似然估计的相合性与渐近正态性) 设总体概率函数为函数 $p(x;\theta)$, $\theta\in\Theta$, Θ 为非退化区间, 若总体分布满足**正规性条件**:

(1) $p(x;\theta)$ 可识别, 即 $\theta_1\neq\theta_2\Longrightarrow p\left(x;\theta_1\right)$ 与 $p\left(x;\theta_2\right)$ 是两个不同的概率函数;

(2) $\forall x$, 偏导 $\dfrac{\partial\log p}{\partial\theta}$, $\dfrac{\partial^2\log p}{\partial\theta^2}$ 和 $\dfrac{\partial^3\log p}{\partial\theta^3}$ 对所有的 $\theta\in\Theta$ 都存在;

(3) $\forall\theta\in\Theta$, 存在满足

$$\int_{-\infty}^{\infty}H_1(x)dx<\infty,\quad\sup_{\theta\in\Theta}\int_{-\infty}^{\infty}H_2(x)p(x;\theta)dx<\infty$$

的函数 $H_1(x)$ 和 $H_2(x)$ 使得

$$\left|\frac{\partial p}{\partial\theta}\right|+\left|\frac{\partial^2 p}{\partial\theta^2}\right|<H_1(x),\quad\left|\frac{\partial^3\log p}{\partial\theta^3}\right|<H_2(x);$$

(4) 对于任意 $\theta \in \Theta$, 费希尔信息量满足 $0 < I(\theta) < \infty$.

若 $x_1, \cdots, x_n$ 是来自该总体 $p(x;\theta)$ 的 iid 样本, 则

- 未知参数 θ 的最大似然估计 $\widehat{\theta}_n$ 存在, 具有相合性;
- 任何似然方程的解 $\widehat{\theta}_n$, 若其相合, 则具有渐近正态性, $\widehat{\theta} \sim AN\left(\theta, \dfrac{1}{nI(\theta)}\right)$.

最大似然估计的概率意义

对离散总体, 似然函数 (看作 θ 的函数)

$$L(\theta) = P\left(X_1 = x_1, \cdots, X_n = x_n; \theta\right) = \prod_{i=1}^{n} P\left(X_i = x_i; \theta\right)$$

正好是样本观测值出现的概率.

对连续总体, 分布 $F(x;\theta)$, 密度 $p(x;\theta)$. 虽然样本观测值出现的概率为 0, 但样本出现在样本观测值附近, 即事件

$$\left\{X_1 \in \left(x_1, x_1 + dx_1\right), \cdots, X_n \in \left(x_n, x_n + dx_n\right)\right\},$$

可以近似为

$$\prod_{i=1}^{n}\left(F\left(x_i + dx_i; \theta\right) - F\left(x_i; \theta\right)\right) \approx \prod_{i=1}^{n} p\left(x_i; \theta\right) dx_i = \prod_{i=1}^{n} p\left(x_i; \theta\right) \cdot \prod_{j=1}^{n} dx_j,$$

其中 $\prod_{j=1}^{n} dx_j$ 不依赖于 θ, 与似然函数 $\prod_{i=1}^{n} p\left(x_i; \theta\right)$ 成比例.

最大似然估计的概率意义是: 既然当前的样本值已经发生, 那么这组观测值发生的概率一定很大; 当前的样本值给定时, 这组观测值发生的概率是参数的函数, 如果一个参数值使得当前样本观测值出现的概率越大, 就可认为它越像真值; 保守起见, 应取使得当前样本值出现的概率达到最大的参数值作为参数的估计, 这就是最大似然估计.

EM 算法

- 设观测数据 $y = (y_1, \cdots, y_n)$ (不完全数据) 的联合分布是 $p(y, \theta)$, 根据情况, 引入适当的随机变量 z, 构成完全数据 $\{y, z\}$, 给定 θ 和 y, 导出 z 的条件密度 $p(z \mid y, \theta)$;
- 完全数据的联合密度为 $p(y, z \mid \theta) = p(z \mid y, \theta) p(y \mid \theta)$, 基于完全数据的对数似然为 $\ell(\theta) = \ln\left(p(y, z \mid \theta)\right)$;
- E 步 (条件期望): 计算 $Q\left(\theta \mid y, \theta^{(i)}\right) = \mathbb{E}_z\left(\ln\left(p(y, z \mid \theta)\right) \mid y, \theta^{(i)}\right)$;
- M 步 (取最大值): 计算 $\theta^{(i+1)} = \operatorname{Arg}\max\limits_{\theta} Q\left(\theta \mid y, \theta^{(i)}\right)$;
- 直到某种精度要求满足, 通常要求迭代到前后两次参数之间的距离小于一个给定的阈值, 比如 10^{-4}.

2.2 区间估计

2.2.1 置信区间

置信区间的概念

定义 2.2.1 (置信区间) 设 θ 是总体的一个参数, 其参数空间为 Θ, $x_1, x_2, \cdots, x_n$ 是来自该总体的样本, 对给定的一个 α $(0 < \alpha < 1)$, 假设有两个统计量 $\widehat{\theta}_L = \widehat{\theta}_L(x_1, x_2, \cdots, x_n)$ 和 $\widehat{\theta}_U = \widehat{\theta}_u(x_1, x_2, \cdots, x_n)$, 若对任意的 $\theta \in \Theta$, 有

$$P_\theta\left(\widehat{\theta}_L \leqslant \theta \leqslant \widehat{\theta}_U\right) \geqslant 1-\alpha,$$

则称随机区间 $[\widehat{\theta}_L, \widehat{\theta}_U]$ 为 θ 的置信水平为 $1-\alpha$ 的**置信区间**, 或简称 $[\widehat{\theta}_L, \widehat{\theta}_U]$ 是 θ 的 $1-\alpha$ 置信区间, $\widehat{\theta}_L$ 和 $\widehat{\theta}_U$ 分别称为 θ 的 (双侧) **置信下限**和**置信上限**.

置信水平 $1-\alpha$ 的频率解释: 在大量重复使用 θ 的置信区间 $[\widehat{\theta}_L, \widehat{\theta}_U]$ 时, 平均而言, 在这大量的区间估计观测值中, 至少有 $100(1-\alpha)\%$ 包含 θ. 如图 2.1 是用程序模拟的正态总体 $N(1,1)$ 下 100 组容量为 20 的样本的 $1-\alpha=0.9$ 和 $1-\alpha=0.5$ 的置信区间, 直观她显示了该种频率意义.

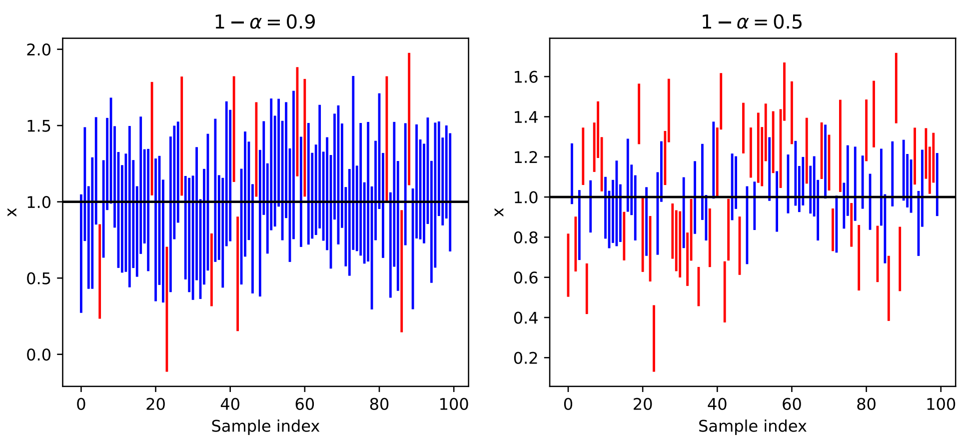

图 2.1: 不同置信水平的置信区间

定义 2.2.2 (同等置信区间) 设 θ 是总体参数, 参数空间为 Θ. $x_1, \cdots, x_n$ 是来自该总体的样本. 对于给定 α $(0<\alpha<1)$, 假设有两个统计量 $\widehat{\theta}_L$ 和 $\widehat{\theta}_U$. 对于任意 $\theta \in \Theta$ 满足

$$P_\theta\left(\widehat{\theta}_L \leqslant \theta \leqslant \widehat{\theta}_U\right) = 1-\alpha,$$

则称 $[\widehat{\theta}_L, \widehat{\theta}_U]$ 为 θ 的 $1-\alpha$ **同等置信区间**.

定义 2.2.3 (置信下限) 设 $\widehat{\theta}_L=\widehat{\theta}_L\left(x_1,x_2,\cdots,x_n\right)$ 是统计量, 对给定的 $\alpha\in(0,1)$ 和任意的 $\theta\in\Theta$, 有

$$P_\theta\left(\widehat{\theta}_L\leqslant\theta\right)\geqslant 1-\alpha\ (\forall\theta\in\Theta),$$

则称 $\widehat{\theta}_L$ 为 θ 的置信水平为 $1-\alpha$ 的 (单侧) **置信下限**. 假如等号对一切 $\theta\in\Theta$ 成立, 则称 $\widehat{\theta}_L$ 为 θ 的 $1-\alpha$ **同等置信下限**.

定义 2.2.4 (置信上限) 设 $\widehat{\theta}_U=\widehat{\theta}_U\left(x_1,x_2,\cdots,x_n\right)$ 是统计量, 对给定的 $\alpha\in(0,1)$ 和任意的 $\theta\in\Theta$, 有

$$P_\theta\left(\widehat{\theta}_v\geqslant\theta\right)\geqslant 1-\alpha,$$

则称 $\widehat{\theta}_U$ 为 θ 的置信水平为 $1-\alpha$ 的 (单侧) **置信上限**. 若等号对一切 $\theta\in\Theta$ 成立, 则称 $\widehat{\theta}_v$ 为 θ 的 $1-\alpha$ **同等置信上限**.

枢轴量法

构造末知参数 θ 的置信区间的最常用的方法是**枢轴量法**, 其步骤如下:

(1) 设法构造一个样本和 θ 的函数 $G=G\left(x_1,x_2,\cdots,x_n,\theta\right)$ 使得 G 的分布不依赖于末知参数. 一般称具有这种性质的 G 为**枢轴量**.

(2) 适当地选择两个常数 c, d, 使对给定的 α $(0<\alpha<1)$, 有

$$P(c\leqslant G\leqslant d)=1-\alpha.$$

在离散场合, 上式等号改为大于等于 ($\geqslant$).

(3) 假如能将 $c\leqslant G\leqslant d$ 进行不等式等价变形化为 $\widehat{\theta}_L\leqslant\theta\leqslant\widehat{\theta}_U$, 则有

$$P_\theta\left(\widehat{\theta}_L\leqslant\theta\leqslant\widehat{\theta}_U\right)=1-\alpha,$$

这表明 $[\widehat{\theta}_L,\widehat{\theta}_U]$ 是 θ 的 $1-\alpha$ 同等置信区间.

枢轴量法间的关键在于构造枢轴量 G. 枢轴量的寻找一般从 θ 的点估计出发, 一般应尽量选用 MLE.

满足要求的 c, d 可以有很多, 选择的目的是希望置信区间的平均长度 $E_\theta\left(\widehat{\theta}_U-\widehat{\theta}_L\right)$ 尽可能短. 不过在不少场合很难做到这一点. 故常这样选择 c 和 d, 使得两个尾部概率各为 $\alpha/2$, 即

$$P_\theta(G<c)=P_\theta(G>d)=\frac{\alpha}{2},$$

这样得到的置信区间称为**等尾置信区间**.

2.2.2 正态总体参数的置信区间

单个正态总体参数的 $1-\alpha$ 置信区间

设 $x_1,x_2,\cdots,x_m$ 是来自 $N\left(\mu,\sigma^2\right)$ 的样本.

1. σ 已知时 μ 的置信区间:

$$\left[\overline{x}-u_{1-\alpha/2}\sigma/\sqrt{n},\overline{x}+u_{1-\alpha/2}\sigma/\sqrt{n}\right]\left(\overline{x}\pm u_{1-\alpha/2}\sigma/\sqrt{n}\right);$$

2. σ 末知时 μ 的置信区间:

$$\overline{x}\pm t_{1-\alpha/2}(n-1)s/\sqrt{n};$$

3. μ 末知时 σ^2 的置信区间:

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

σ 的置信区间:

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

两个正态总体下的 $1-\alpha$ 置信区间

设 $x_1,x_2,\cdots,x_m$ 是来自 $N\left(\mu_1,\sigma_1^2\right)$ 的样本, $y_1,y_2,\cdots,y_n$ 是来自 $N\left(\mu_2,\sigma_2^2\right)$ 的样本.

1. σ_1^2 和 σ_2^2 已知时 $\mu_1-\mu_2$ 的置信区间:

$$\overline{x}-\overline{y}\pm u_{1-a/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}};$$

2. $\sigma_1^2=\sigma_2^2=\sigma^2$ 末知时 $\mu_1-\mu_2$ 的置信区间:

$$\overline{x}-\overline{y}\pm\sqrt{\frac{m+n}{mn}}s_w t_{1-\alpha/2}(m+n-2),$$

其中 $s_w^2=\dfrac{(m-1)s_x^2+(n-1)s_y^2/c}{m+n-2}$;

3. $\sigma_2^2/\sigma_1^2=c$ 已知时 $\mu_1-\mu_2$ 的置信区间:

$$\overline{x}-\overline{y}\pm\sqrt{\frac{mc+n}{mn}}s_w t_{1-a/2}(m+n-2),$$

其中 $s_w^2=\dfrac{(m-1)s_x^2+(n-1)s_y^2/c}{m+n-2}$;

4. 当 m 和 n 都很大时 $\mu_1-\mu_2$ 的近似置信区间:

$$\overline{x}-\overline{y}\pm u_{1-\alpha/2}\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}};$$

5. 一般情况下 $\mu_1-\mu_2$ 的一种近似置信区间:

$$\overline{x}-\overline{y}\pm s_0 t_{1-\alpha/2}(l),$$

其中 $s_0^2=s_x^2/m+s_y^2/n$, $l=\dfrac{s_0^4}{\frac{s_x^4}{m^2(m-1)}+\frac{s_y^4}{n^2(n-1)}}$, l 一般不为整数, 可以取与 l 最接近的整数代替之;

6. μ_1, μ_2 未知时 σ_1^2/σ_2^2 的置信区间:

$$\left[\frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{1-\alpha/2}(m-1,n-1)}, \frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{\alpha/2}(m-1,n-1)}\right].$$

2.2.3 大样本置信区间

基于 MLE 的近似置信区间

在最大似然估计的场合, 密度函数 $p(x;\theta)$ 中的参数 θ 常有一列估计量 $\widehat{\theta}_n=\widehat{\theta}_n(x_1, x_2, \cdots, x_n)$, 并有渐近正态分布 $N\left(\theta,\sigma_n^2(\theta)\right)$, 其中渐近方差 $\sigma_n^2(\theta)$ 是参数 θ 和样本量 n 的函数. 在一般场合, 若用 MLE $\widehat{\theta}_n$ 代替 $\sigma_n(\theta)$ 中的 θ, 上式仍然成立, 因为 MLE $\widehat{\theta}_n$ 还是 θ 的相合估计. 此时对给定的置信水平 $1-\alpha$ $(0<\alpha<1)$, 利用标准正态分布的分位数, 可得

$$P\left(-u_{1-a/2}<\frac{\widehat{\theta}_n-\theta}{\sigma_n\left(\widehat{\theta}_n\right)}<u_{1-a/2}\right)\approx 1-\alpha$$

从而可得 θ 的近似 $1-\alpha$ 的等尾置信区间:

$$\widehat{\theta}_n\pm u_{1-\alpha/2}\sigma_n\left(\widehat{\theta}_n\right).$$

基于中心极限定理的近似置信区间

在独立同分布样本场合, 只要总体均值 μ 与总体方差 σ^2 存在, 无论总体分布是什么, 据中心极限定理, 其样本均值 $\overline{x}$ 有渐近正态分布, 即

$$\overline{x}\dot{\sim}N\left(\mu,\sigma^2/n\right)$$

由此立即可得总体均值 μ 的近似 $1-\alpha$ 的等尾置信区间:

$$\overline{x}\pm u_{1-\alpha/2}\sigma/\sqrt{n}$$

若其中 σ 未知, 用 σ^2 的相合估计 (譬如样本方差 s^2) 替代即可.

例 2.2.5 (比例 p 的置信区间) 设 $x_1,x_2,\cdots,x_n$ 是来自二点分布 $b(1,p)$ 的样本, p 的 $1-\alpha$ 置信区间近似为:

$$\left[\overline{x}-u_{1-\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}, \overline{x}+u_{1-\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\right].$$

Chapter 3

假设检验

3.1 假设检验的基本思想与概念

3.1.1 假设检验的步骤

一般情况下, 假设检验的步骤如下:

1. 根据实际问题, 建立统计假设 H_0 vs H_1;

2. 选取一个合适的检验统计量 $T(\boldsymbol{x})$, 使当 H_0 成立时 (或 H_0 中某个具体参数下), T 的分布完全已知, 并根据 H_0 及 H_1 的特点, 确定拒绝域 W 的形式;

3. 选择合适的显著性水平 α;

4. 确定具体的拒绝域 W;

5. 由样本观测值 x_1, x_2, $\cdots$, x_n, 计算检验统计量的 $T(x_1, x_2, \cdots, x_n)$, 由 $T(x_1$, x_2, $\cdots$, $x_n)$ 是否属于 W, 作出最终判断.

建立假设

设有来自某参数分布 $F(x, \theta)$ $(\theta \in \Theta)$ 的样本 $x_1, \cdots, x_n$, 其中 Θ 为参数空间. 设非空集合 $\Theta_0 \subset \Theta$, 则命题 "$H_0 : \theta \in \Theta_0$" 称为一个假设, 或**原假设**或者零假设; 令 $\Theta_1 = \Theta - \Theta_0$, 则命题 "$H_1 : \theta \in \Theta_1$" 称为 H_0 的对立假设或者**备择假设**. **假设检验问题**可描述为

$$H_0 : \theta \in \Theta_0 \text{ vs } H_1 : \theta \in \Theta_1.$$

如果 Θ_0 仅包含一个点, 则称之为**简单原假设**, 否则称为**复合原假设** (备择假设也有简单和复合之分). 简单原假设可写成 $H_0 : \theta = \theta_0$. 这时备择假设有三种可能:

$$H_1' : \theta \neq \theta_0 \quad H_1'' : \theta < \theta_0 \quad H_1''' : \theta > \theta_0.$$

H_0 vs H_1' 称作双侧假设或双边假设 (备择假设在原假设两侧); H_0 vs H_1'' 和 H_0 vs H_1''' 称作单侧假设或单边假设 (备择假设在原假设单侧).

选择检验统计量, 临界值与拒绝域

对于假设的检验就是指这样的一个法则: 当有了具体的样本后, 按该法则就可决定是接受 H_0 还是拒绝 H_0, 即检验就等价于把样本空间划分成两个互不相交的部分 W 和 $\overline{W}$, 当样本属于 W 时, 拒绝 H_0; 否则接受 H_0. 于是, 我们称 W 为该检验的**拒绝域**, 而 $\overline{W}$ 称为**接受域**. 由样本对原假设进行检验总是通过一个统计量完成的, 该统计量称为检验统计量.

当拒绝域确定了, 检验的**判断准则**跟着也确定了:

- 如果 $(x_1, x_2, \cdots, x_n) \in W$, 则拒绝 H_0;
- 如果 $(x_1, x_2, \cdots, x_n) \in \overline{W}$, 则接受 H_0.

一个拒绝域 W 唯一确定一个检验法则, 反之, 一个检验法则也唯一确定一个拒绝域.

选择显著性水平: 两类错误

假设检验可能犯如下两种错误: 当 $\theta \in \Theta_0$ 时, 样本却落人了拒绝域 W, 采取了拒绝 H_0 的错误决策, 称这样的错误为**第一类错误**, 或**拒真错误**; 当 $\theta \in \Theta_1$ 时, 样本却落人了接受域 $\overline{W}$, 采取了接受 H_0 的错误决策, 称这样的错误为**第二类错误**, 或**取伪错误**, 如表 3.1.

表 3.1: 检验的两类错误

观测数据情况	总体情况	
	H_0 为真	H_1 为真
$(x_1, x_2, \cdots, x_n) \in W$	犯第一类错误	正确
$(x_1, x_2, \cdots, x_n) \in \overline{W}$	正确	犯第二类错误

犯第一类错误概率 $\alpha(\theta) = P_\theta\{X \in W\}$, $\theta \in \Theta_0$; 犯第二类错误概率 $\beta(\theta) = P_\theta\{X \in \overline{W}\}$, $\theta \in \Theta_1$.

定义 3.1.1 (势函数) 设检验问题

$$H_0 : \theta \in \Theta_0 \text{ vs } H_1 : \theta \in \Theta_1$$

的拒绝域为 W, 则样本观测值 $\boldsymbol{x}$ 落在拒绝域 W 内的概率称为该检验的**势函数**或者**功效函数**, 记为

$$g(\theta) = P_\theta(X \in W) \ (\theta \in \Theta = \Theta_0 \cup \Theta_1).$$

犯两类错误的概率都是参数 θ 的函数, 并可由势函数得到:

$$\begin{cases} \alpha(\theta) = g(\theta), & \theta \in \Theta_0; \\ \beta(\theta) = 1 - g(\theta), & \theta \in \Theta_1. \end{cases}$$

在样本量一定的条件下不可能找到一个使 $\alpha(\theta)$, $\beta(\theta)$ 都小的检验. 通常的做法是仅限制犯第一类错误的概率, 这就是费希尔的显著性检验.

定义 3.1.2 (显著性检验) 对检验问题 $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, 如果一个检验满足对任意的 $\theta \in \Theta_0$, 都有

$$g(\theta) \leqslant \alpha,$$

则称该检验是显著性水平为 α 的**显著性检验**, 简称水平为 α 的检验.

显著性检验就是要控制犯第一类错误的概率 α, 但也不能使得 α 过小 (α 过小会导致 β 过大), 在适当控制 α 中制约 β. 最常用的选择是 $\alpha = 0.05$, 有时也选择 $\alpha = 0.10$ 或 $\alpha = 0.01$.

给出拒绝域

确定显著性水平之后, 可根据拒绝域的形状, 确定具体的拒绝域.

做出判断

在有了明确的拒绝域 W 后, 根据样本观测值可以作出判断.

3.1.2 基本原理和逻辑

统计假设检验所依据的**基本原理**是**小概率原理**: 小概率事件在一次试验中是不会发生的.

假设检验的基本逻辑 (努力证明原假设 H_0 不对): 由拒绝域 W 的定义, 在原假设成立时 $P((x_1, \cdots, x_n) \in W) \leqslant \alpha$. 类似于反证法, 先假设原假设 H_0 成立. 因为 α 通常很小, 所以此时 W 是一个小概率事件. 根据小概率原理, 观测到的样本 $(x_1, \cdots, x_n)$ 落入拒绝域 W 是不会发生的. 如果上述事件发生了, 则产生矛盾, 因此我们比较有把握的认为原假设不成立, 从而拒绝原假设; 如果上述事件没有发生, 则没有产生矛盾, 因此没有足够的证据表明 H_0 不对, 所以无法拒绝原假设.

3.1.3 两个假设的选取

假设检验中两个假设的选取是个重要问题, 通常应遵循以下原则:

- 应当把如果它成立但是误判为不成立时后会造成严重后果的命题选为原假设;
- 应当把分析人员想证明它正确的命题作为备择假设, 把分析人员努力证明它不正确的命题作为原假设;
- 应当把大众普遍认为它成立的命题作为原假设, 因为原假设不能轻易拒绝, 除非有足够的证据表明它不对.

3.1.4 检验的 p 值

定义 3.1.3 (检验的 p 值) 在一个假设检验问题中, 利用样本观测值能够拒绝原假设的最小的显著性水平, 称为**检验的 p 值**.

p 值是根据样本计算得到的数值, 所以 p 值是统计量. 由检验的 p 值与显著性水平 α 进行比较可以很容易作出检验的结论:

- 如果 $p \leqslant \alpha$, 则在显著性水平 α 下拒绝 H_0;
- 如果 $p > \alpha$, 则在显著性水平 α 下接受 H_0.

建立拒绝域, 然后考察样本观察值是否落入拒绝域; 与由样本观测值计算检验的 p 值, 然后与显著性水平比较两种进行检验的方法是等价的.

3.2 正态总体参数假设检验

下面检验的显著性水平都是 α.

3.2.1 单个正态总体均值的检验

设 $x_1, x_2, \cdots, x_n$ 是来自 $N\left(\mu, \sigma^2\right)$ 的样本, 考虑如下三种关于 μ 的检验问题:

$$\begin{array}{rl}
\text{I} & H_0: \mu \leqslant \mu_0 \text{ vs } H_1: \mu > \mu_0, \\
\text{II} & H_0: \mu \geqslant \mu_0 \text{ vs } H_1: \mu < \mu_0, \\
\text{III} & H_0: \mu = \mu_0 \text{ vs } H_1: \mu \neq \mu_0,
\end{array}$$

其中 μ_0 是已知常数.

$\sigma = \sigma_0$ 已知时的 u 检验

u 检验统计量

$$u = \frac{\overline{x} - \mu_0}{\sigma_0 / \sqrt{n}} \sim N(0,1).$$

拒绝域与 p 值如表 3.2, 其中 $u_0 = \dfrac{\sqrt{n}\left(\overline{x} - \mu_0\right)}{\sigma_0}$.

表 3.2: u 检验

问题	拒绝域	p 值
I	$W_{\text{I}} = \{u \geqslant u_{1-\alpha}\}$	$p_{\text{I}} = P\left(u \geqslant u_0\right) = 1 - \Phi\left(u_0\right)$
II	$W_{\text{II}} = \{u \leqslant u_\alpha\}$	$p_{\text{II}} = P\left(u \leqslant u_0\right) = \Phi\left(u_0\right)$
III	$W_{\text{III}} = \left\{\lvert u\rvert \geqslant u_{1-\alpha/2}\right\}$	$p_{\text{III}} = P\left(\lvert u\rvert \geqslant \lvert u_0\rvert\right) = 2\left(1 - \Phi\left(\lvert u_0\rvert\right)\right)$

实际中也可能会遇到以下的假设:

$$\begin{array}{rl}
\text{IV} & H_0: \mu = \mu_0 \text{ vs } H_1: \mu > \mu_0, \\
\text{V} & H_0: \mu = \mu_0 \text{ vs } H_1: \mu < \mu_0.
\end{array}$$

问题 IV 的拒绝域与问题 I $H_0: \mu \leqslant \mu_0$ vs $H_1: \mu > \mu_0$ 相同, 都是 $W = \{u \geqslant u_{1-\alpha}\}$. 问题 V 的拒绝域与检验问题 II $H_0: \mu \geqslant \mu_0$ vs $H_1: \mu < \mu_0$ 相同, 都是 $W = \{u \leqslant u_\alpha\}$.

σ 未知时的 t 检验

t 检验统计量

$$t=\frac{\sqrt{n}\left(\overline{x}-\mu_0\right)}{s}\sim t(n-1).$$

拒绝域与 p 值如表 3.3, 其中 $t_0=\dfrac{\sqrt{n}\left(\overline{x}-\mu_0\right)}{s}$.

表 3.3: t 检验

问题	拒绝域	p 值
I	$W_{\mathrm{I}}=\{t\geqslant t_{1-\alpha}(n-1)\}$	$p_{\mathrm{I}}=P\left(t\geqslant t_0\right)$
II	$W_{\mathrm{II}}=\{t\leqslant t_{\alpha}(n-1)\}$	$p_{\mathrm{II}}=P\left(t\leqslant t_0\right)$
III	$W_{\mathrm{III}}=\left\{\lvert t\rvert\geqslant t_{1-\alpha/2}(n-1)\right\}$	$p_{\mathrm{III}}=P\left(\lvert t\rvert\geqslant\lvert t_0\rvert\right)$

3.2.2 两个正态总体均值差的检验

设 $x_1,x_2,\cdots,x_m$ 是来自正态总体 $N\left(\mu_1,\sigma_1^2\right)$ 的样本, $y_1,y_2,\cdots,y_n$ 是来自另一个正态总体 $N\left(\mu_2,\sigma_2^2\right)$ 的样本, 两个样本相互独立. 考虑如下三类检验问题:

$$\begin{array}{ll}
\mathrm{I} & H_0:\mu_1-\mu_2\leqslant 0 \text{ vs } H_1:\mu_1-\mu_2>0,\\
\mathrm{II} & H_0:\mu_1-\mu_2\geqslant 0 \text{ vs } H_1:\mu_1-\mu_2<0,\\
\mathrm{III} & H_0:\mu_1-\mu_2=0 \text{ vs } H_1:\mu_1-\mu_2\neq 0.
\end{array}$$

σ_1, σ_2 已知时的两样本 u 检验

u 检验统计量

$$u=\frac{\overline{x}-\overline{y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\sim N(0,1).$$

拒绝域与 p 值如表 3.4, 其中 $u_0=\dfrac{\overline{x}-\overline{y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}$

表 3.4: u 检验

问题	拒绝域	p 值
I	$W_{\mathrm{I}}=\{u\geqslant u_{1-\alpha}\}$	$p_{\mathrm{I}}=P\left(u\geqslant u_0\right)=1-\Phi\left(u_0\right)$
II	$W_{\mathrm{II}}=\{u\leqslant u_{\alpha}\}$	$p_{\mathrm{II}}=P\left(u\leqslant u_0\right)=\Phi\left(u_0\right)$
III	$W_{\mathrm{III}}=\left\{\lvert u\rvert\geqslant u_{1-\alpha/2}\right\}$	$p_{\mathrm{III}}=P\left(\lvert u\rvert\geqslant\lvert u_0\rvert\right)=2\left(1-\Phi\left(\lvert u_0\rvert\right)\right)$

$\sigma_1=\sigma_2=\sigma$ 但未知时的两样本 t 检验

t 检验统计量

$$t=\frac{\overline{x}-\overline{y}}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(n+m-2).$$

其中 $s_w^2=\dfrac{(m-1)s_x^2+(n-1)s_y^2/c}{m+n-2}$. 拒绝域与 p 值如表 3.5, 其中 $t_0=\dfrac{\overline{x}-\overline{y}}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}}$.

表 3.5: t 检验

问题	拒绝域	p 值
I	$W_{\mathrm{I}}=\{t\geqslant t_{1-\alpha}(m+n-2)\}$	$p_{\mathrm{I}}=P\left(t\geqslant t_0\right)$
II	$W_{\mathrm{II}}=\{t\leqslant t_{\alpha}(m+n-2)\}$	$p_{\mathrm{II}}=P\left(t\leqslant t_0\right)$
III	$W_{\mathrm{III}}=\left\{\lvert t\rvert\geqslant t_{1-\alpha/2}(m+n-2)\right\}$	$p_{\mathrm{III}}=P\left(\lvert t\rvert\geqslant\lvert t_0\rvert\right)$

σ_1 和 σ_2 未知且 m, n 很大时的大样本 u 检验

检验统计量

$$u=\frac{\overline{x}-\overline{y}}{\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}}}\dot{\sim}N\left(\frac{\mu_1-\mu_2}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}},1\right).$$

拒绝域与 p 值如表 3.6, 其中 $u_0=\dfrac{\overline{x}-\overline{y}}{\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}}}$.

表 3.6: 大样本 u 检验

问题	拒绝域	p 值
I	$W_{\mathrm{I}}=\{u\geqslant u_{1-\alpha}\}$	$p_{\mathrm{I}}=P\left(u\geqslant u_0\right)=1-\Phi\left(u_0\right)$
II	$W_{\mathrm{II}}=\{u\leqslant u_{\alpha}\}$	$p_{\mathrm{II}}=P\left(u\leqslant u_0\right)=\Phi\left(u_0\right)$
III	$W_{\mathrm{III}}=\left\{\lvert u\rvert\geqslant u_{1-\alpha/2}\right\}$	$p_{\mathrm{III}}=P\left(\lvert u\rvert\geqslant\lvert u_0\rvert\right)=2\left(1-\Phi\left(\lvert u_0\rvert\right)\right)$

σ_1 和 σ_2 未知且 m, n 较小时的近似 t 检验

检验统计量

$$t=\frac{\overline{x}-\overline{y}}{\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}}}\dot{\sim}t(l).$$

其中 $s_0^2 = s_x^2/m + s_y^2/n$, $l = \dfrac{s_0^4}{\frac{s_x^4}{m^2(m-1)} + \frac{s_y^4}{n^2(n-1)}}$, l 一般不是整数, 可取最接近的整数代替. 拒绝域与 p 值如表 3.7, 其中 $t_0 = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{s_s^2}{m} + \frac{s_y^2}{n}}}$.

表 3.7: 近似 t 检验

问题	拒绝域	p 值
I	$W_{\mathrm{I}} = \{t \geqslant t_{1-\alpha}(l)\}$	$p_{\mathrm{I}} = P\left(t \geqslant t_0\right)$
II	$W_{\mathrm{II}} = \{t \leqslant t_{\alpha}(l)\}$	$p_{\mathrm{II}} = P\left(t \leqslant t_0\right)$
III	$W_{\mathrm{III}} = \left\{\lvert t\rvert \geqslant t_{1-\alpha/2}(l)\right\}$	$p_{\mathrm{III}} = P\left(\lvert t\rvert \geqslant \lvert t_0\rvert\right)$

成对数据检验

在对两个总体均值进行比较时, 有时数据是成对出现的, 此时若采用二样本 t 检验所得出的结论有可能是不对的. 对于成对数据, 应当尽量采用成对数据的 t 检验.

记 $d_i = x_i - y_i$, $\mu = \mu_1 - \mu_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2$, 那么 $d_i \sim N\left(\mu, \sigma^2\right)$. 原始的两样本检验问题 I, II, III 分别等价于如下的单样本检验问题

$$
\begin{aligned}
\mathrm{I}' &\quad H_0 : \mu \leqslant 0 \text{ vs } H_1 : \mu > 0,\\
\mathrm{II}' &\quad H_0 : \mu \geqslant 0 \text{ vs } H_1 : \mu < 0,\\
\mathrm{III}' &\quad H_0 : \mu = 0 \text{ vs } H_1 : \mu \neq 0,
\end{aligned}
$$

再用 3.2.1 的方法检验. 这种方法检验目的更加明确, 数据的误差更小, 功效更高 (从而第二类错误概率更小).

3.2.3 正态总体方差的检验

对方差检验时通常假定 μ 未知.

单个正态总体方差的 χ^2 检验

设 $x_1, x_2, \cdots, x_n$ 是来自 $N\left(\mu, \sigma^2\right)$ 的样本, 对方差亦可考虑如下三个检验问题:

$$
\begin{aligned}
\mathrm{I} &\quad H_0 : \sigma^2 \leqslant \sigma_0^2 \text{ vs } H_1 : \sigma^2 > \sigma_0^2,\\
\mathrm{II} &\quad H_0 : \sigma^2 \geqslant \sigma_0^2 \text{ vs } H_1 : \sigma^2 < \sigma_0^2,\\
\mathrm{III} &\quad H_0 : \sigma^2 = \sigma_0^2 \text{ vs } H_1 : \sigma^2 \neq \sigma_0^2,
\end{aligned}
$$

其中 σ_0^2 是已知常数.

χ^2 检验统计量

$$\chi^2 = (n-1)s^2/\sigma_0^2 \sim \chi^2(n-1).$$

拒绝域与 p 值如表 3.8, 其中 $\chi_0^2=(n-1)s^2/\sigma_0^2$.

表 3.8: χ^2 检验

问题	拒绝域	p 值
I	$W_{\mathrm{I}}=\left\{\chi^2 \geqslant \chi_{1-\alpha}^2(n-1)\right\}$	$p_{\mathrm{I}}=P\left(\chi^2 \geqslant \chi_0^2\right)$
II	$W_{\mathrm{II}}=\left\{\chi^2 \leqslant \chi_{1-\alpha}^2(n-1)\right\}$	$p_{\mathrm{II}}=P\left(\chi^2 \leqslant \chi_0^2\right)$
III	$W_{\mathrm{III}}=\left\{\chi^2 \leqslant \chi_{\alpha/2}^2(n-1)\ 或 \chi^2 \geqslant \chi_{1-\alpha/2}^2(n-1)\right\}$	$p_{\mathrm{III}}=2\min\left\{P\left(\chi^2 \geqslant \chi_0^2\right), P\left(\chi^2 \leqslant \chi_0^2\right)\right\}$

两个正态总体方差比的 F 检验

设 $x_1,x_2,\cdots,x_m$ 是来自 $N\left(\mu_1,\sigma_1^2\right)$ 的样本, $y_1,y_2,\cdots,y_n$ 是来自 $N\left(\mu_2,\sigma_2^2\right)$ 的样本. 考虑如下三个假设检验问题:

$$
\begin{array}{ll}
\mathrm{I} & H_0:\sigma_1^2 \leqslant \sigma_2^2 \text{ vs } H_1:\sigma_1^2>\sigma_2^2, \\
\mathrm{II} & H_0:\sigma_1^2 \geqslant \sigma_2^2 \text{ vs } H_1:\sigma_1^2<\sigma_2^2, \\
\mathrm{III} & H_0:\sigma_1^2=\sigma_2^2 \text{ vs } H_1:\sigma_1^2 \neq \sigma_2^2.
\end{array}
$$

F 检验统计量

$$
F=\frac{s_x^2}{s_y^2} \sim F(m-1,n-1).
$$

拒绝域与 p 值如表 3.9, 其中 $F_0=s_x^2/s_y^2$.

表 3.9: F 检验

问题	拒绝域	p 值
I	$W_{\mathrm{I}}=\{F \geqslant F_{1-a}(m-1,n-1)\}$	$p_{\mathrm{I}}=P(F \geqslant F_0)$
II	$W_{\mathrm{II}}=\{F \leqslant F_\alpha(m-1,n-1)\}$	$p_{\mathrm{II}}=P(F \leqslant F_0)$
III	$W_{\mathrm{III}}=\left\{F \leqslant F_{\alpha/2}(m-1,n-1)\ 或 F \geqslant F_{1-\alpha/2}(m-1,n-1)\right\}$	$p_{\mathrm{III}}=2\min\{P(F \geqslant F_0), P(F \leqslant F_0)\}$

3.3 其他分布参数的假设检验

3.3.1 指数分布参数的检验

设 $x_1,\cdots,x_n$ 是来自指数分布 $\operatorname{Exp}(1/\theta)$ 的样本, θ 是其均值. 考虑关于 θ 的如下检验问题:

$$
\begin{array}{ll}
\mathrm{I} & H_0:\theta \leqslant \theta_0 \text{ vs } H_1:\theta>\theta_0, \\
\mathrm{II} & H_0:\theta \geqslant \theta_0 \text{ vs } H_1:\theta<\theta_0, \\
\mathrm{III} & H_0:\theta=\theta_0 \text{ vs } H_1:\theta \neq \theta_0,
\end{array}
$$

其中 θ_0 是确定的常数.

检验统计董

$$\chi^2 = \frac{2n\overline{x}}{\theta_0} \sim \chi^2(2n).$$

拒绝域与 p 值如表 3.10, 其中 $\chi_0^2 = \dfrac{2n\overline{x}}{\theta_0}$.

表 3.10: 指数分布参数的检验

问题	拒绝域	p 值
I	$W_{\mathrm{I}} = \left\{\chi^2 \geqslant \chi^2_{1-\alpha}(2n)\right\}$	$p_{\mathrm{I}} = P\left(\chi^2 \geqslant \chi_0^2\right)$
II	$W_{\mathrm{II}} = \left\{\chi^2 \leqslant \chi^2_{\alpha}(2n)\right\}$	$p_{\mathrm{II}} = P\left(\chi^2 \leqslant \chi_0^2\right)$
III	$W_{\mathrm{III}} = \left\{\chi^2 \leqslant \chi^2_{\alpha/2}(2n) \text{ 或} \chi^2 \geqslant \chi^2_{1-\alpha/2}(2n)\right\}$	$p_{\mathrm{III}} = 2\min\left\{P\left(\chi^2 \leqslant \chi_0^2\right), P\left(\chi^2 \geqslant \chi_0^2\right)\right\}$

3.3.2 比率 p 的检验

假设做了 n 次独立试验, 以 x 表示该事件发生的次数, 那么 $x \sim b(n,p)$. 考虑如下检验问题:

$$\begin{array}{ll} \mathrm{I} & H_0: p \leqslant p_0 \text{ vs } H_1: p > p_0, \\ \mathrm{II} & H_0: p \geqslant p_0 \text{ vs } H_1: p < p_0, \\ \mathrm{III} & H_0: p = p_0 \text{ vs } H_1: p \neq p_0. \end{array}$$

检验统计量

$$x = x \sim b(n,p).$$

拒绝域与 p 值如表 3.11, 其中 x_0 为 x 的观测值, c_0 满足

$$\sum_{i=0}^{c_0} \binom{n}{i} p_0^i \left(1-p_0\right)^{n-i} < \alpha < \sum_{i=0}^{c_0+1} \binom{n}{i} p_0^i \left(1-p_0\right)^{n-i}.$$

表 3.11: 指数分布参数的检验

问题	拒绝域	p 值
I	$W_{\mathrm{I}} = \{x \geqslant c_0\}$	$p_{\mathrm{I}} = P\left(x \geqslant x_0\right)$
II	$W_{\mathrm{II}} = \{x \leqslant c_0\}$	$p_{\mathrm{II}} = P\left(x \leqslant x_0\right)$
III	$W_{\mathrm{III}} = \left\{x \leqslant c_1 \text{ 或} x \geqslant c_2\right\}$	$p_{\mathrm{III}} = 2\min\left\{P\left(x \leqslant x_0\right), P\left(x \geqslant x_0\right)\right\}$

c_1, c_2 的确定较复杂, 在离散场合使用 p 值作检验较为简便.

3.3.3 大样本检验

如果样本量较大, 还常用渐近正态分布构造检验统计量, 获得大样本检验. 其一般思路如下: 设 $x_1, x_2, \cdots, x_n$ 是来自某总体分布 $F(x;\theta)$ 的样本, 又设该总体均值为 θ, 方差为 θ 的函数, 记为 $\sigma^2(\theta)$. 现要对下列三类假设检验问题:

$$\begin{array}{ll} \text{I} & H_0: \theta \leqslant \theta_0 \text{ vs } H_1: \theta > \theta_0, \\ \text{II} & H_0: \theta \geqslant \theta_0 \text{ vs } H_1: \theta < \theta_0, \\ \text{III} & H_0: \theta = \theta_0 \text{ vs } H_1: \theta \neq \theta_0, \end{array}$$

寻找大样本检验方法. 在样本容量 n 充分大时, 利用中心极限定理知 $\overline{x} \dot{\sim} N(\theta, \sigma^2(\theta)/n)$, 故可采用如下检验统计量:

$$u = \frac{\sqrt{n}\,(\overline{x} - \theta_0)}{\sqrt{\sigma^2(\widehat{\theta})}} \dot{\sim} N\left(\frac{\theta - \theta_0}{\sigma/\sqrt{n}}, 1\right),$$

其中 $\widehat{\theta}$ 为 θ 的 MLE, 检验的拒绝域和近似的 p 值与单样本正态总体场合 (3.2.1) 一样.