



Mathematic Statistics

课程笔记

作者: Fir1247

组织: USTC

时间: December 30, 2023

版本: 1.1

联系方式: fa1247@mail.ustc.edu.cn 或者 QQ: 3105292483



在理性的基础上，所有的判断都是统计学。——C.R.Rao

目录

附录	i
0.1 常见分布	i
0.2 常用记号	ii
0.3 常用缩写	ii
第1章 数理统计预备知识	1
1.1 基础概念	1
1.2 抽样分布	4
1.3 统计三大分布	7
1.4 次序统计量	10
1.5 指数族	13
1.6 充分统计量	16
1.7 完全统计量	20
1.8 统计量的极限分布 *	24
第2章 点估计	26
2.1 矩估计	26
2.2 极大似然估计	29
2.3 点估计的优良性准则	33
2.4 一致最小方差无偏估计	38
2.5 Cramer-Rao 不等式	46
第3章 区间估计	50
3.1 基本概念	50
3.2 正态总体参数的置信区间	51
第4章 参数假设检验	53
4.1 基本概念	53
4.2 正态总体参数的假设检验——枢轴变量	56
4.3 似然比检验	58
4.4 一致最优检验	62
第5章 非参数假设检验	64
5.1 符号检验与符号秩和检验	64



5.2 Wilcoxon 两样本秩和检验.....	66
5.3 拟合优度检验.....	67
5.4 列联表中的独立性和齐一性检验	71
5.5 其他的非参数检验方法 *	73
第 6 章 贝叶斯方法简介	74
6.1 引言和若干基本概念.....	74
6.2 先验分布的确定	75
6.3 贝叶斯统计推断	76

附录

本节主要放一些很难记忆的内容，方便随时查阅。

0.1 常见分布

分布	密度函数 & 特征函数	期望	方差
离散均匀分布： m 个点概率相同 $X \sim U(m), m \in \mathbb{N}_+$	$P(X = k) = \frac{1}{m}, k = 1 \sim m$	$\frac{m+1}{2}$	$\frac{(m-1)(m+1)}{12}$
二项分布： 前 n 次成功次数 $X \sim b(n, p), n \in \mathbb{N}_+, p \in (0, 1)$	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0 \sim n$ $\varphi(t) = (1-p + pe^{it})^n$	np	$np(1-p)$
泊松分布： 单位时间内发生次数 $X \sim P(\lambda), \lambda > 0$	$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, k \in \mathbb{N}$	λ	λ
多项分布： n 个点不同概率 $X \sim M(n, p_1, \dots, p_n), p_i \in (0, 1)$	$P(X = k) = p_k, k = 1 \sim n$	略	略
负二项分布： 成功 r 次前失败次数 $X \sim NB(r, p), r \in \mathbb{N}_+, p \in (0, 1)$	$P(X = k) = \binom{k+r-1}{r-1} (1-p)^k p^r, k \in \mathbb{N}$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
超几何分布： N 白 M 黑抽 n 个中白球数 $X \sim H(N, M, n)$	$P(X = k) = \frac{\binom{N}{k} \binom{M}{n-k}}{\binom{N+M}{n}}$	$\frac{nN}{N+M}$	$\frac{nNM(N+M-n)}{(N+M)^2(N+M-1)}$

表 1: 离散型分布

分布	密度函数 & 特征函数	期望	方差
均匀分布: $X \sim U(a, b), -\infty < a < b < \infty$	$f(x) = \frac{1}{b-a} I_{(a,b)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布: $X \sim \exp(\lambda), \lambda > 0$	$f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}$ $\varphi(t) = \frac{\lambda}{\lambda - it}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态分布: $X \sim N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\mu = 0, \sigma^2 = 1$ 时 $\varphi(t) = e^{-\frac{t^2}{2}}$	μ	σ^2
Gamma 分布: $X \sim \Gamma(\alpha, \lambda), \alpha, \lambda > 0$	$f(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{\Gamma(\alpha)} I_{(0,\infty)}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Beta 分布: $X \sim B(a, b), a, b > 0$	$f(x) = \frac{x^{a-1} (1-x)^{b-1}}{B(a,b)} I_{(0,1)}$	$\frac{a}{a+b}$	$\frac{a(a+1)}{(a+b)(a+b+1)}$
Cauchy 分布:	$f(x) = \frac{1}{\pi(1+x^2)}$	不存在	不存在
χ^2 分布: 标准正态分布的方差 $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$	$f_n(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x} I_{(0,\infty)}$	n	$2n$
t 分布:			
F 分布:			

表 2: 连续型分布

0.2 常用记号

1. k 阶 n 维样本中心矩: $m_{nk} = \frac{1}{n} \sum (X_i - \bar{X})^k$.
2. k 阶 n 维样本原点矩: $a_{nk} = \frac{1}{n} \sum X_i^k$.
3. k 阶总体矩: $\alpha_k = EX^k$, $\mu_k = E(X - \alpha_1)^k$.
4. 样本方差: $S^2 = \frac{n}{n-1} m_{n2} = \frac{1}{n-1} \sum (X_i - \bar{X})^2$.
5. 正态分布的 α 分位点: $X \sim N(0, 1)$, $P(X > u_\alpha) = \alpha$.

0.3 常用缩写

1. 随机变量: r.v.
2. 独立同分布: iid.
3. 特征函数: c.f.
4. (累计) 分布函数: c.d.f. 或者 d.f.
5. (概率) 密度函数: p.d.f.
6. (概率) 质量函数: p.m.f
7. 矩估计: MM 或 ME, 简记 M.
8. 极大似然估计: MLE, 简记 L.
9. 均方误差: MSE
10. 一致最小方差无偏估计: UMVUE

第 1 章 数理统计预备知识

1.1 基础概念

1.1.1 样本与样本分布

定义 1.1.1

1. 总体：研究对象的全体 $\rightarrow$ 某个指标取值的全体，以及这些值的概率。例如用一个随机变量来表达一个总体： $X \sim N(\mu, \sigma^2)$ ，正态总体； $X \sim B(n, p)$ ，二项分布总体。
2. 样本：从总体中按一定规则抽取的一些个体，记为 $X = (X_1, \dots, X_n)$ 。有两种抽样方式：有放回和不放回。这两种抽样方式都具有“代表性”——每一个个体都有可能被抽到、抽到的可能性相同。有放回抽样又叫做简单随机抽样： $X_1, X_2, \dots, X_n \text{ iid. } \sim F$ 。但不放回抽样， $X_1, \dots, X_n$ 之间不相互独立。
3. 二重性：在抽样方案实施之前，样本视为随机的，当抽样方案实施后视为若干数字，通常记作 $(X_1, \dots, X_n)$ 。

例 1.1.1.

有 N 个产品，其中有 M 个次品，不放回抽取 n 个样本 $(X_1, \dots, X_n)$ ，其中

$$A_i = \{\text{第 } i \text{ 个产品是次品}\}, X_i = I_{A_i}$$

$\sum X_i$ 满足超几何分布。从条件概率出发，可以得到

$$P(X_i = x_i, i = 1, 2, \dots, n) = P(X_1 = x_1)P(X_2 = x_2|X_1 = x_1) \cdots P(X_n = x_n|X_i = x_i, i = 1, 2, \dots, n-1)$$

例如 $n = 3, (x_1, x_2, x_3) = (1, 0, 1)$ ，上式为

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

一般样本空间从 N 个产品中有序取出 n 个样品的不同排列数目有

$$(N)_n := \frac{N!}{n!} = N(N-1)\cdots(N-n+1)$$

从 M 个次品中取出 a 个次品, 从 $N-M$ 个正品中取出 $n-a$ 个正品, 排列数目有

$$(M)_a \cdot (N-M)_{n-a}$$

于是

$$P(X_i = x_i, i = 1, 2, \dots, n) = \frac{(M)_a \cdot (N-M)_{n-a}}{(N)_n}, \sum x_i = a$$

如果是有放回抽取的, $X_1, \dots, X_n$ iid. 则

$$P(X_i = x_i, i = 1, 2, \dots, n) = \left(\frac{M}{N}\right)^a \left(\frac{N-M}{N}\right)^{n-a}, \sum x_i = a$$

以后我们谈抽样, 如不做特殊说明, 都为简单随机抽样。

例 1.1.2.

$X_1, \dots, X_n$ iid. 如果是离散的 r.v.

$$P_\theta(X_i = x_i, i = 1, \dots, n) = \prod_{i=1}^n P_\theta(X_i = x_i) = \prod_{i=1}^n p(x_i, \theta)$$

连续型 r.v.

$$f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$$

$$F_\theta(x_1, \dots, x_n) = \prod_{i=1}^n F_\theta(x_i)$$

1.1.2 统计推断

统计推断包括估计和假设检验。

定义 1.1.2

参数和参数空间: 记

$$\mathcal{F} = \{f(x_1, \dots, x_n, \theta) : \theta \in \Theta\}$$

称为样本的分布族, 其中 θ 可以是一个向量, Θ 被称为参数空间。

统计推断: 由样本对参数作推断, 由局部推断整体。

统计量: 由样本加工得到的量, 即关于样本的 (可测) 函数,

$$T = T(X_1, \dots, X_n)$$

简言之, 统计量 T 是关于 X 的可测函数。统计量当样本确定后就是一个数, 不能包含其他的变量。统计推断和决策理论的一个主要问题即是找出 T 的分布。

例 1.1.3.

$$T = \sum_{i=1}^n X_i - a$$

a 如果是一个确定的数, 那么 T 就是统计量, 否则不是。

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

是统计量, 因为由样本完全确定。

注 统计量也有二重性: 既是一个随机变量, 也是由样本完全确定的量。

常见统计量通常由样本矩和次序统计量构成。

样本矩是指

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad k \text{ 阶样本原点矩}$$

$$m_k = \frac{1}{n} (X_i - \mathbb{E}X)^k \quad k \text{ 阶样本中心矩}$$

并且几乎处处收敛于

$$a_k \xrightarrow{\text{a.s.}} \mathbb{E}X^k$$

$$m_k \xrightarrow{\text{a.s.}} \mathbb{E}(X - a)^k$$

此外还有一些样本矩构成的统计量:

$$\hat{\beta}_1 = \frac{m_3}{m_2^{\frac{3}{2}}} \xrightarrow{\text{a.s.}} \mathbb{E} \left(\frac{X - \mathbb{E}X}{\sigma} \right)^3 \quad \text{样本偏度}$$

$$\hat{\beta}_2 = \frac{m_4}{m_2^2} \xrightarrow{\text{a.s.}} \mathbb{E} \left(\frac{X - \mathbb{E}X}{\sigma} \right)^4 \quad \text{样本峰度}$$

次序统计量的相关内容会在后面详细介绍。**注** m_2 即方差, 也写成 S^2 , 和中学时期学的样本方差 S_X^2 (也是一个常用的统计量) 稍有不同:

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

这里的 $n-1$ 是自由度: 能自由变换的自变量个数。如果设 $y_i = x_i - \bar{x}$, 那么满足 $y_1 + \cdots + y_n = 0$, 作了一条限制, 所以自由度是 $n-1$ 而不是 n 。而回到 m_2 , 这里的 $\mathbb{E}X$ 是一个统计量, 和 X_i 共计 $n+1$ 个随机变量, 满足一条限制, 故自由度是 n 。

有关 $\bar{X}$ 和 S_X^2 的一些性质:

1. $\sum (X_i - \bar{X}) = 0$.
2. 令 $Y_i = aX_i + b$, 则 $S_Y^2 = a^2 S_X^2$.
3. $\min_{a \in \mathbb{R}} \sum (X_i - a)^2 = \sum (X_i - \bar{X})^2 = (n-1) S_X^2$.
4. $S_X^2 = \frac{1}{n-1} [\sum X_i^2 - n\bar{X}^2]$.

关于 $\bar{X}$ 和 S_X^2 的样本矩,

1. 如果同分布 F 有有限均值 μ , 则 $\mathbb{E}\bar{X} = \mu$.
2. 如果同分布 F 有有限方差 σ^2 , 则 $\text{Var}(\bar{X}) = \frac{1}{n} \sigma^2$, $\mathbb{E} S_X^2 = \sigma^2$.
3. 如果 $\mathbb{E}|X_1|^3$ 有限, 则 $\mathbb{E}(\bar{X}^3)$ 和 $\text{Cov}(\bar{X}, S_X^2)$ 有限。
4. 如果 $\mathbb{E}|X_1|^4$ 有限, 则 $\text{Var} S_X^2$ 有限。

1.2 抽样分布

1.2.1 复习内容

定义 1.2.1

统计量 $T = T(X_1, \dots, X_n)$ 的分布被称为抽样分布, 如果总体 $X \sim F(f)$, 则样本

$$(X_1, \dots, X_n) \sim \prod F(x_i) (\prod f(x_i))$$

为了研究分布, 引入工具:

$$g(t) = \mathbb{E}e^{itX} = \int e^{itX} dF(x) = \int e^{itX} f(x) dx \quad \text{特征函数}$$

另外还有矩母函数 $g_1(t) = \mathbb{E}e^{tX}$ 和生成函数 $g_2(t) = \mathbb{E}t^X$ 。但是特征函数用的最多, 因为它总是存在的。

例 1.2.1.

Γ -分布:

$$f(x, \lambda, \alpha) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} I_{(0, \infty)}(x)$$

其中

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

特殊值 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ 。那么如何求 Γ -分布的特征函数?

解答:

$$\begin{aligned} g(t) &= \int_0^\infty \frac{e^{itx} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-it)x} dx \\ &\stackrel{u=(\lambda-it)x}{=} \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{1}{(\lambda-it)^{\alpha-1}} \cdot \frac{1}{\lambda-it} \int_L u^{\alpha-1} e^{-u} du \\ &= \dots = \frac{1}{(1-\frac{it}{\lambda})^\alpha} \end{aligned}$$

这里的复积分路径是直线 $L: (\lambda-it)x, x \in (0, \infty)$, 根据全纯函数的性质, 可知这个积分可以当成实值积分 (积分区域 $(0, \infty)$) 来求。中间的推导过程跳过了。

特别地, 取 $\alpha = 1$, 则 Γ 分布为 $\exp(\lambda)$, 总体 $X \sim \exp(\lambda)$ 时,

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum X_i \\ g_{\bar{X}}(t) &= \mathbb{E}e^{it\bar{X}} = \mathbb{E}e^{\frac{it}{n} \sum X_i} = (\mathbb{E}e^{\frac{it}{n} X_1})^n \\ &= \frac{1}{(1-\frac{it}{n\lambda})^n} \end{aligned}$$

对比 $g(t)$ 可得

$$f_{\bar{X}}(x) = \frac{(n\lambda)^n x^{n-1} e^{-n\lambda x}}{(n-1)!} I_{(0, \infty)}(x)$$

定义 1.2.2

对于一个由随机变量构成的矩阵 $X = (X_{ij})_{n \times m}$, 定义 $\mathbb{E}X = (\mathbb{E}X_{ij})_{n \times m}$, 容易验证:

$$\mathbb{E}(A \cdot X \cdot B) = A \cdot \mathbb{E}X \cdot B$$

由此定义协方差矩阵: n 维随机向量 X, Y ,

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = (\text{Cov}(X_i, Y_j))_{m \times m}$$

1.2.2 来自正态总体的抽样分布

设总体 $\sim N(a, \sigma^2)$, 样本

$$(X_1, \dots, X_n) \sim \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2\right\}$$

总体的特征函数

$$g(t) = \mathbb{E}e^{itX} = e^{ita - \frac{t^2}{2}\sigma^2}$$

例 1.2.2.

如果 $X = (X_1, X_2) \sim N(a_1, a_2, \sigma_1^2, \sigma_2^2, \rho)$, 分布函数是

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{(x_1 - a_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - a_1)(x_2 - a_2)}{\sigma_1\sigma_2} + \frac{(x_2 - a_2)^2}{\sigma_2^2} \right]\right\}$$

很复杂, 但是用矩阵来定义二维乃至多维正态分布就方便得多。记

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = (\text{Cov}(X_i, X_j))_{2 \times 2}$$

于是

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

那么 p.d.f. 就能表示成

$$f(x_1, x_2) = \frac{1}{2\pi} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{a})' \Sigma^{-1}(\mathbf{x} - \mathbf{a})\right\}$$

定理 1.2.1

设随机变量 $X_1, \dots, X_n$ 相互独立, 且 $X_k \sim N(a_k, \sigma_k^2), k = 1, 2, \dots, n$ 。令 $c_1, \dots, c_n$ 为常数, 则

$$T = \sum_{k=1}^n c_k X_k \sim N(\mu, \tau^2)$$

其中

$$\mu = \sum_{k=1}^n c_k a_k, \tau^2 = \sum_{k=1}^n c_k^2 \sigma_k^2$$

证明 利用特征函数可以证明这个定理。

如果 $X = (X_1, \dots, X_n)'$, X_i 相互独立, $X_i \sim N(a_i, \sigma_i^2)$, 则

$$(X_1, \dots, X_n) \sim (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a_i)^2\right\}$$

记 $\mathbf{x} = (x_1, \dots, x_n)'$, $\mathbf{a} = (a_1, \dots, a_n)'$, 则

$$(X_1, \dots, X_n) \sim (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{a})'(\mathbf{x} - \mathbf{a})\right\}$$

若 $Y = (Y_1, \dots, Y_n)' = AX$, $A = (a_{ij})_{n \times n}$, 那么

$$Y_i = \sum_{j=1}^n a_{ij} X_j, i = 1, \dots, n$$

概率论学过以下定理:

定理 1.2.2

如果 $Y = \varphi(X)$, $X \sim f(x)dx$, 那么

$$Y \sim f(\varphi^{-1}(y)) \left| \frac{dx}{dy} \right| \cdot dy$$

推广到多元的情况: $Y_i = \varphi_i(X_1, \dots, X_n), i = 1, \dots, n$, $(X_1, \dots, X_n) \sim f(x_1, \dots, x_n)dx_1 \cdots dx_n$, 如果可以反解出 $X_i = \psi_i(Y_1, \dots, Y_n)$, 那么

$$Y \sim f(\psi_1, \dots, \psi_n) \left| \frac{dx_1 \cdots dx_n}{dy_1 \cdots dy_n} \right| dy_1 \cdots dy_n$$

这里其实是雅可比矩阵:

$$\frac{dx_1 \cdots dx_n}{dy_1 \cdots dy_n} = \frac{\partial(X_1, \dots, X_n)}{\partial(Y_1, \dots, Y_n)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

特别地, 如果 $Y = AX$, $X \sim f(x) = f(x_1, \dots, x_n)$, 反解很简单:

$$Y \sim f(A^{-1}a) \cdot |A|^{-1}$$

如果 A 是一个正交阵, $A' = A^{-1}$, 则

$$Y \sim f(A'a)$$

引理 1.2.1

$X_1, \dots, X_n$ 相互独立, $X_i \sim N(a_i, \sigma^2)$, 如果 $Y = AX$, 其中 A 是一个正交阵, 记 $a = (a_1, \dots, a_n)'$, $x = (x_1, \dots, x_n)'$, 则 $Y \sim N_n(Aa, \sigma^2 I_n)$.

证明 因为

$$X \sim (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} (x - a)'(x - a) \right\}$$

所以

$$\begin{aligned} Y = AX &\sim (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} (A^{-1}a - a)'(A^{-1}a - a) \right\} \\ &= (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} (a - b)'(a - b) \right\} \\ &= (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - b_i)^2 \right\} \end{aligned}$$

其中 $b = \mathbb{E}Y = A \cdot a$. 于是 $Y_1, \dots, Y_n$ 相互独立, 且 $Y_i \sim N(b_i, \sigma^2)$.

注 这个引理也是回归分析作业 Week2 的第一题。

如果 A 并不是正交阵, 那么结论为: $Y \sim N_n(Aa, \sigma^2 AA^T)$.

定理 1.2.3

如果 $X_1, \dots, X_n$ 相互独立, $X_i \sim N(a, \sigma^2)$, $Y = AX$, A 是正交阵, 记 $a = (a, \dots, a)'$, $b = Aa = (b_1, \dots, b_n)'$, 记 $\mathbf{1} = (1, \dots, 1)'$, 则

1. $\bar{X} \sim N(a, \frac{\sigma^2}{n})$.
2. $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$.
3. $\bar{X}$ 和 s^2 相互独立。

证明 在引理中取 $A = \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix}$, $A^{-1} = A'$, 且令 $\mathbf{a}_1 = \frac{1}{\sqrt{n}}\mathbf{1}$, 由正交阵的定义, $\mathbf{a}'_1 \mathbf{a}_i = 0$, 即 $\mathbf{a}_i$ 各分量之和等于 0, 令 $Y = AX$, 则

$$\mathbb{E}Y = A(\mathbf{a}_1) = a \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix} \mathbf{1} = a \begin{pmatrix} \sqrt{n} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

而

$$Y_1 = \frac{1}{\sqrt{n}} \sum X_i = \sqrt{n}\bar{X}$$

$$Y_j \sim N(0, \sigma^2)$$

且 Y_i 之间相互独立, 于是

$$\frac{Y_j}{\sigma} \sim N(0, 1) \stackrel{\text{def of } \chi^2}{\Rightarrow} \sum_{j=2}^n Y_j^2 \sim \chi_{n-1}^2$$

同时,

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

由于 A 为正交阵, 所以 $X'X = Y'Y$,

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

所以

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

1.3 统计三大分布

1.3.1 χ^2 分布

定义 1.3.1

设 $X_1, \dots, X_n \text{ iid. } \sim N(0, 1)$, 则

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

就是自由度为 n 的 χ^2 分布, 记作 χ_n^2 .

基本性质:

1. d.f. 为

$$\begin{aligned} F_n(x) &= P(\chi^2 \leq x) = \int \cdots \int_{x_1^2 + \cdots + x_n^2 \leq x} (2\pi)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n x_i^2\right\} dx_1 \cdots dx_n \\ &\stackrel{r^2 = x_1^2 + \cdots + x_n^2}{=} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1}} \int_0^{\sqrt{x}} e^{-\frac{r^2}{2}} r^{n-1} dr \end{aligned}$$

2. p.d.f. 为

$$f_n(x) = \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} I_{(0,+\infty)}$$

3. 记 k 阶原点矩为 α_k , k 阶中心矩为 μ_k , 则

$$\alpha_k = \mathbb{E}X^k = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_0^\infty x^{\frac{n}{2}+k-1} e^{-\frac{1}{2}x} dx$$

$$\stackrel{x=2y}{=} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_0^\infty 2^{\frac{n}{2}+k} y^{\frac{n}{2}+k-1} e^{-y} dy = \frac{2^k \Gamma(\frac{n}{2}+k)}{\Gamma(\frac{n}{2})} = n(n+2)\cdots(n+2k-2)$$

特殊地, 期望 $\mathbb{E}X = \alpha_1 = n$, 所以

$$\mu_k = \mathbb{E}(X-n)^k = \mathbb{E} \sum_{0 \leq i \leq k} \binom{k}{i} (-1)^i n^i X^{k-i} = \sum_{0 \leq i \leq k} \binom{k}{i} (-1)^i n^i \alpha_{k-i}$$

方差 $\text{Var}(X) = \mu_2 = 2n$, 还有另一种算法:

$$\begin{aligned} \text{Var}(X) &= \text{Var} \sum_{i=1}^n (X_i^2) \\ &= n \cdot \text{Var}(X_1^2) = n(\mathbb{E}X_1^4 - (\mathbb{E}X_1^2)^2) = n(3-1) = 2n \end{aligned}$$

4. 如果 $Z \sim N(0,1)$, 则 $Z^2 \sim \chi_1^2$.

5. $X \sim \chi_m^2$, $Y \sim \chi_n^2$, 如果 X 和 Y 独立, 则 $X+Y \sim \chi_{m+n}^2$, 这是由定义直接得到的.

6. χ_n^2 是特殊的 Gamma 分布: $\text{Gamma}(\frac{n}{2}, \frac{1}{2})$.

7. χ_n^2 分布的特征函数为 $\varphi(t) = (1-2it)^{-\frac{n}{2}}$.

8. 设 $X_1, \dots, X_n$ i.i.d. $\sim \exp(\lambda)$, 则 $2\lambda n \bar{X} \sim \chi_{2n}^2$.

9. χ^2 分布最重要的性质是与正态总体的联系, 见定理 1.2.3.

1.3.2 t 分布

定义 1.3.2

$x \sim N(0,1)$, $Y \sim \chi_n^2$, X 和 Y 独立, 则

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

的分布称为自由度为 n 的 t 分布, 记作 t_n .

基本性质:

1. 令 $\eta = \sqrt{T/n}$, $Y \sim f(x)dx$,

$$\eta \sim f(\eta^2) \cdot 2n\eta d\eta := f_1(t)dt$$

于是

$$\xi = \frac{X}{\eta} \sim \int_0^\infty t \cdot h_1(xt) f_1(t) dt = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

这里的 h_1 是标准正态分布的密度函数, 得到 p.d.f. 为

$$f_n(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

2. 当 $n=1$ 时,

$$t_1 \sim \frac{1}{\pi} \frac{1}{1+x^2}$$

这是 Cauchy 分布, 期望不存在.

3. 当 $n=2$ 时,

$$t_2 \sim C \left(1 + \frac{x^2}{2}\right)^{-\frac{3}{2}}$$

期望存在并且为 0, 但是方差不存在.

4. 当 $n \rightarrow \infty$ 时,

$$\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \rightarrow e^{-\frac{x^2}{n} \cdot \frac{n+1}{2}} \rightarrow e^{-\frac{1}{2}x^2}$$

$$\Gamma(n+1) \sim n! \sim \sqrt{2n\pi} \left(\frac{n}{e}\right)^n$$

所以 t_n 分布的极限是标准正态分布。

5. $\xi \sim t_n$, 则

$$\mathbb{E}\xi = 0, n \geq 2$$

$$\text{Var}\xi = \frac{n}{n-2}, n \geq 3$$

计算方法: 三角换元。

推论 1.3.1

总体 $X \sim N(a, \sigma^2)$, 样本 $X_1, \dots, X_n$, 则

$$\frac{\sqrt{n}(\bar{X} - a)}{S_X} \sim t_{n-1}$$

在第三章区间估计与第四章参数假设检验中, 我们将看到这个推论的重要性: 可用于在 μ 和 σ^2 均未知的情况下, 求 μ 的区间估计以及检验样本均值 $\bar{X}$ 和假设均值 μ 之间是否存在显著区别。

证明 $\bar{X} \sim N(a, \frac{\sigma^2}{n})$, 标准化:

$$\frac{\bar{X} - a}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$S_X^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 \Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

且 S_X^2 和 $\bar{X}$ 独立, 那么

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

推论 1.3.2

设 $X_1, \dots, X_m$ i.i.d. $\sim N(\mu_1, \sigma^2)$, $Y_1, \dots, Y_n$ i.i.d. $\sim N(\mu_2, \sigma^2)$, 且样本 $X_1, \dots, X_m$ 和 $Y_1, \dots, Y_n$ 相互独立, 则

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w} \cdot \sqrt{\frac{mn}{m+n}} \sim t_{m+n-2}$$

其中 $(m+n-2)S_w^2 = (m-1)S_X^2 + (n-1)S_Y^2$.

1.3.3 F 分布

定义 1.3.3

$X \sim \chi_m^2$, $Y \sim \chi_n^2$, 且相互独立, 则

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

服从自由度为 m, n 的 F 分布, 记作 $\xi \sim F_{m,n}$.

基本性质:

1. p.d.f. 为

$$f_{m,n}(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m}{2}-1}}{(1+\frac{m}{n}x)^{\frac{m+n}{2}}} I_{(0,+\infty)}$$

2. 若 $\xi \sim F_{m,n}$, 则 $\xi^{-1} \sim F_{n,m}$.

3. 若 $T \sim t_n$, 则 $T \sim F_{1,n}$.

4. 若 $\xi \sim F_{m,n}$, 则 $E\xi = \frac{n}{n-2}, n > 2$.

5. 若 $\xi \sim F_{m,n}$, 则

$$\frac{\frac{m}{n}\xi}{1+\frac{m}{n}\xi} \sim \text{Beta}\left(\frac{m}{2}, \frac{n}{2}\right)$$

6. 若 $\xi \sim F_{m,n}$, 记 $P(\xi > F_{m,n}(\alpha)) = \alpha$, 则 $F_{m,n}(1-\alpha) = \frac{1}{F_{m,n}(\alpha)}$.

推论 1.3.3

设 $X_1, \dots, X_m$ i.i.d. $\sim N(\mu_1, \sigma^2)$, $Y_1, \dots, Y_n$ i.i.d. $\sim N(\mu_2, \sigma^2)$, 且样本 $X_1, \dots, X_m$ 和 $Y_1, \dots, Y_n$ 相互独立, 则

$$\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F_{m-1, n-1}$$



例 1.3.1.

$X \sim N(0, 1)$, $Y \sim \chi_n^2$, 那么考虑

$$t_n^2 = \frac{X^2}{Y/n}$$

它实际上满足 F 分布。

1.4 次序统计量

次序统计量: 设 $X_1, \dots, X_n$ iid. $X_i \sim f(x)$, $F(x)$ 为 d.f., 按照从小到大来排列:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

那么每个 $X_{(i)}$ 都是统计量, $(X_{(1)}, \dots, X_{(n)})$ 被称为次序统计量。

例 1.4.1.

利用次序统计量可以构造一些其他的统计量:

1. 样本中位数:

$$m = \begin{cases} X_{(\frac{n+1}{2})} & , n \text{ 是奇数} \\ \frac{1}{2} (X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}) & , n \text{ 是偶数} \end{cases}$$

2. 样本 p 分位数:

$$X_{[(n+1)p]}, p \in (0, 1)$$

3. 样本极差:

$$R = X_{(n)} - X_{(1)}$$

4. 样本中程数:

$$mR = \frac{1}{2} (X_{(1)} + X_{(n)})$$

例 1.4.2. 经验分布函数 *

$$F_n(x) = \frac{1}{n} \# \{k : X_k \leq x\} = \frac{1}{n} \sum_{k=1}^n I_{\{X_{(k)} \leq x\}} = \frac{1}{n} \sum_{k=1}^n I_{\{X_k \leq x\}}$$

经验分布函数的基本性质:

1. 具有累积分布函数的基本要素: 有界、单调递增、右连续。
2. $nF_n(x) \sim \text{Binomial}(n, F(x))$.
3. 对与任意给定的 x , $F_n(x) \xrightarrow{P}$ 分布函数 $F(x) = P(X \leq x)$.
4. Glivenko-Cantelli 定理:

$$\sup_x |F_n(x) - F(x)| \xrightarrow{P} 0$$

1.4.1 联合分布

考察 $(X_{(1)}, \dots, X_{(n)})$, 设

$$\begin{aligned} (X_{(1)}, \dots, X_{(n)}) &\sim f(x_1, \dots, x_n) I_{\{x_1 \leq \dots \leq x_n\}} dx_1 \cdots dx_n \\ &\approx P(x_i \leq X_{(i)} \leq x_i + dx_i, i = 1, 2, \dots, n) \\ &= n! f(x_1) dx_1 \cdots f(x_n) dx_n I_{\{x_1 \leq \dots \leq x_n\}} \end{aligned}$$

所以

$$f(x_1, \dots, x_n) = n! \prod_{i=1}^n f(x_i) I_{\{x_1 \leq \dots \leq x_n\}}$$

1.4.2 单个量的分布

考察 $X_{(i)}$,

$$X_{(i)} \sim f(x) dx$$

即: $X_{(i)}$ 落到 $(x, x+dx)$ 中, $i-1$ 个落在 x 之前, $n-i$ 个落到 $x+dx$ 之后,

$$\frac{n!}{(i-1)!(n-i)!} F(x)^{i-1} f(x) dx (1-F(x))^{n-i}$$

结论为:

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} [1-F(x)]^{n-i} f(x)$$

1.4.3 两个量的联合分布

考察 $(X_{(i)}, X_{(j)})$,

$$(X_{(i)}, X_{(j)}) \sim f(x, y) dx dy I_{\{x < y\}}$$

$X_{(i)}$ 落到 $(x, x+dx)$, $X_{(j)}$ 落到 $(y, y+dy)$, 于是有 $i-1$ 个落在 x 之前, $j-i-1$ 个落在 $(x+dx, y)$, $n-j$ 个落在 $y+dy$ 之后,

$$\frac{n!}{(i-1)!(j-i-1)!(n-j)!} [F(x)]^{i-1} f(x) dx [F(y) - F(x)]^{j-i-1} f(y) dy [1-F(y)]^{n-j} I_{\{x < y\}}$$

结论为:

$$f_{ij}(x, y) = n! \frac{[F(x)]^{i-1} [F(y) - F(x)]^{j-i-1} [1-F(y)]^{n-j}}{(i-1)!(j-i-1)!(n-j)!} f(x) f(y) I_{\{x < y\}}$$

1.4.4 两个量之差的分布

考察 $X_{(j)} - X_{(i)}$, 令 $Z = X_{(i)}$, $V = X_{(j)} - X_{(i)}$,

$$\begin{pmatrix} Z \\ V \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} X_{(i)} \\ X_{(j)} \end{pmatrix}$$

于是 $(Z, V) \sim g(\zeta, \zeta + v) \Rightarrow V \sim \int g(\zeta, \zeta + v) d\zeta$. 这个积分并不容易, 我们只考虑特殊情况: $X_i \sim U(0, 1)$. 那么 m.f 为

$$f(x) = I_{(0,1)}(x)$$

于是

$$(X_{(i)}, X_{(j)}) \sim g(x, y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} x^{i-1} (y-x)^{j-i-1} (1-y)^{n-j} I_{\{0 < x < y < 1\}}$$

推导过程中要利用到 Beta 函数, 较为复杂, 这里略过, 结论为:

$$V \sim h(v) = \frac{n!}{(j-i-1)!(n-j+i)!} v^{j-i-1} (1-v)^{n-j+i} I_{\{0 < v < 1\}}$$

例 1.4.3.

一些常用的特殊情况:

$$X_{(1)} \sim \frac{n!}{(n-1)!} f(x) [1 - F(x)]^{n-1} = n f(x) [1 - F(x)]^{n-1} \quad \text{最小值}$$

$X_{(1)}$ 的 d.f. 为 $1 - [1 - F(x)]^n$.

$$X_{(n)} \sim \frac{n!}{(n-1)!} [F(x)]^{n-1} f(x) = n [F(x)]^{n-1} f(x) \quad \text{最大值}$$

$X_{(n)}$ 的 d.f. 为 $F(x)^n$.

$$R = X_{(n)} - X_{(1)} \sim \frac{n!}{(n-2)!} x^{n-2} (1-x) I_{(0,1)} = n(n-1) x^{n-2} (1-x) I_{(0,1)} \quad U(0,1) \text{ 分布下的极差}$$

1.4.5 离散型分布

定理 1.4.1

设 $X_1, \dots, X_n$ i.i.d. 取自概率质量函数 $P(X = x_i) = p_i$ 的离散型总体 X , 其中 $x_i, i \in \mathbb{N}$ 是 X 的所有可能的取值。定义:

$$P_0 = 0, P_1 = p_1, P_2 = p_1 + p_2, \dots, P_i = p_1 + \dots + p_i, \dots$$

则

$$P(X_{(j)} \leq x_i) = \sum_{k=j}^n \binom{n}{k} P_i^k (1 - P_i)^{n-k}$$

进而

$$P(X_{(j)} = x_i) = \sum_{k=j}^n \binom{n}{k} [P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1}^{n-k})]$$

1.5 指数族

1.5.1 定义与例子

定义 1.5.1

若样本 $X = (X_1, \dots, X_n)$ 的样本空间为

$$\mathcal{F} = \{f(\mathbf{x}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$$

且

$$f(\mathbf{x}, \boldsymbol{\theta}) = c(\boldsymbol{\theta}) \exp\{Q(\boldsymbol{\theta})'T(\mathbf{x})\}h(\mathbf{x})$$

其中 $Q(\boldsymbol{\theta}) = Q(\theta_1, \dots, \theta_k) = (Q_1(\boldsymbol{\theta}), \dots, Q_k(\boldsymbol{\theta}))$, $T(\mathbf{x}) = (T_1(\mathbf{x}), \dots, T_k(\mathbf{x}))$, $h(\mathbf{x}) > 0$, 则称 $\mathcal{F}$ 为指数型分布族, 简称指数族。

$X \sim f(\mathbf{x})$, $\{\mathbf{x} : f(\mathbf{x}) > 0\}$ 称为随机向量的支撑, 指数分布族有共同的支撑, $f(\mathbf{x}, \boldsymbol{\theta}) > 0 \Leftrightarrow h(\mathbf{x}) > 0$, 所以支撑集与 $\boldsymbol{\theta}$ 无关。如果有关则说明不是指数族。

例 1.5.1. 正态总体

总体 $\sim N(a, \sigma^2)$, 样本 $X = (X_1, \dots, X_n)$ 密度函数为

$$\begin{aligned} f(\mathbf{x}, a, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum (x_i - a)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{na^2}{2\sigma^2}\right\} \exp\left\{\frac{a}{\sigma^2} \sum x_i - \frac{1}{2\sigma^2} \sum x_i^2\right\} \end{aligned}$$

令

$$\begin{aligned} c(a, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{na^2}{2\sigma^2}\right\} \\ Q(a, \sigma^2) &= \left(\frac{a}{\sigma^2}, -\frac{1}{2\sigma^2}\right)^T \\ T(\mathbf{x}) &= \left(\sum x_i, \sum x_i^2\right)^T \\ h(\mathbf{x}) &= 1 \end{aligned}$$

例 1.5.2. Gamma 总体

总体 $\sim \Gamma(\nu, \lambda)$,

$$\Gamma(\nu, \lambda) \sim \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x} I_{(0, \infty)}(x)$$

于是样本 $X = (X_1, \dots, X_n)$ 服从分布:

$$\begin{aligned} f(\mathbf{x}, \nu, \lambda) &= \prod_{i=1}^n \frac{\lambda^\nu}{\Gamma(\nu)} x_i^{\nu-1} e^{-\lambda x_i} I_{(0, \infty)}(x_i) \\ &= \frac{\lambda^{n\nu}}{[\Gamma(\nu)]^n} \exp\left\{-\lambda \sum_{i=1}^n x_i + (\nu-1) \sum_{i=1}^n \ln x_i\right\} \prod_{i=1}^n I_{(0, \infty)}(x_i) \end{aligned}$$

令

$$\begin{aligned}c(\nu, \lambda) &= \frac{\lambda^{n\nu}}{[\Gamma(\nu)]^n} \\Q(\nu, \lambda) &= (-\lambda, \nu - 1)' \\T(\mathbf{x}) &= \left(\sum x_i, \sum \ln x_i\right)' \\h(\mathbf{x}) &= \prod_{i=1}^n I_{(0, \infty)}(x_i)\end{aligned}$$

总体

注 我们也可以写成：

$$f(\mathbf{x}, \nu, \lambda) = \frac{\lambda^{n\nu}}{[\Gamma(\nu)]^n} \exp \left\{ -\lambda \sum_{i=1}^n x_i + \nu \sum_{i=1}^n \ln x_i \right\} \prod_{i=1}^n x_i^{-1} I_{(0, \infty)}(x_i)$$

这说明构造是不唯一的。

例 1.5.3. 二项分布总体

$n = 1$ ，二项分布族 $B(N, p)$ ，

$$\begin{aligned}P_p(X = x) &= \binom{N}{x} p^x (1-p)^{N-x} \\&= (1-p)^N \left(\frac{p}{1-p}\right)^x \binom{N}{x} \\&= (1-p)^N \exp \left\{ x \ln \frac{p}{1-p} \right\} \binom{N}{x}\end{aligned}$$

令

$$c(p) = (1-p)^N, Q(p) = \ln \frac{p}{1-p}, T(x) = x, h(x) = \binom{N}{x}$$

即可。多元的情况类似。

例 1.5.4. Poisson 总体

总体 $\sim P(\lambda)$ ， $n = 1$ ，

$$P_\lambda(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} e^{x \ln \lambda} \frac{1}{x!}, x = 0, 1, 2, \dots$$

例 1.5.5. 非指数族的例子

总体 $\sim U(0, \theta)$ ， $\theta > 0$

$$X = (X_1, \dots, X_n) \sim f(\mathbf{x}, \theta) = \frac{1}{\theta^n} \prod_{i=1}^n I_{(0, \theta)}(x_i)$$

支撑集与 θ 有关，不是指数族。同理，如果总体 $\sim$ 双参数指数分布，即

$$f(\mathbf{x}, \mu, \sigma) = \frac{1}{\sigma} \exp \left\{ -\frac{x - \mu}{\sigma} \right\} I_{\{x > \mu\}}$$

不是指数族。若 μ 固定，则为指数族。
非指数族的例子

1.5.2 指数族的自然形式和自然参数空间

定义 1.5.2

若样本分布族 $\mathcal{F}$ 中的元 $f(\mathbf{x}, \boldsymbol{\theta})$ 可以写成

$$\begin{aligned} f(\mathbf{x}, \boldsymbol{\theta}) &= c^*(\boldsymbol{\theta}) \exp\{\boldsymbol{\theta}' T(\mathbf{x})\} h(\mathbf{x}) \\ &= c^*(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^n \theta_i T_i(\mathbf{x})\right\} h(\mathbf{x}) \end{aligned}$$

则该分布族称为指数族的自然形式，此时集合

$$\left\{ \boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^T : 0 < \int_X \exp\left\{\sum_{i=1}^n \theta_i T_i(\mathbf{x})\right\} h(\mathbf{x}) d\mathbf{x} < \infty \right\}$$

被称为自然参数空间。

对于一般的指数族

$$f(\mathbf{x}, \boldsymbol{\theta}) = c(\boldsymbol{\theta}) \exp\{Q(\boldsymbol{\theta})' T(\mathbf{x})\} h(\mathbf{x})$$

令 $\varphi_i = Q_i(\theta_1, \dots, \theta_k)$ ，若 $(\varphi_1, \dots, \varphi_k)$ 与 $(\theta_1, \dots, \theta_k)$ 一一对应，反解出 $\theta_i = \psi_i(\varphi_1, \dots, \varphi_k)$ ，则得到

$$f(\mathbf{x}, \boldsymbol{\varphi}) = c^*(\boldsymbol{\varphi}) \exp\{\boldsymbol{\varphi}' T(\mathbf{x})\} h(\mathbf{x})$$

这说明一般的指数族可以转化为自然指数族。

例 1.5.6. 正态总体

总体 $\sim N(a, \sigma^2)$,

$$(X_1, \dots, X_n) \sim f(\mathbf{x}, a, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{na^2}{2\sigma^2}\right\} \exp\left\{\frac{a}{\sigma^2} \sum x_i - \frac{1}{2\sigma^2} \sum x_i^2\right\}$$

设 $\varphi_1 = \frac{a}{\sigma^2}$, $\varphi_2 = -\frac{1}{2\sigma^2}$, 于是 $a = -\frac{\varphi_1}{2\varphi_2}$, $\sigma^2 = -\frac{1}{2\varphi_2}$,

$$(X_1, \dots, X_n) \sim \left(-\frac{\varphi_2}{\pi}\right)^{\frac{n}{2}} \exp\left\{\frac{n\varphi_1^2}{4\varphi_2}\right\} \exp\{\varphi_1 \sum x_i + \varphi_2 \sum x_i^2\}$$

参数空间:

$$\{(\varphi_1, \varphi_2) : \varphi_1 \in \mathbb{R}, \varphi_2 < 0\}$$

例 1.5.7. 二项分布总体

二项分布族

$$b(N, P, x) = (1-p)^N \exp \left\{ x \ln \frac{p}{1-p} \right\} \binom{N}{x}$$

$$\varphi = \ln \frac{p}{1-p} \Rightarrow p = \frac{e^\varphi}{1+e^\varphi}$$

$$\Rightarrow b(N, P, x) = (1+e^\varphi)^{-N} \exp \{x\varphi\} \binom{N}{x}$$

参数空间:

$$\{\varphi : \varphi \in \mathbb{R}\}$$

定理 1.5.1

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

$$G(\eta) = \int g(x) \exp \left\{ \sum_{i=1}^k \eta_i T_i(\mathbf{x}) \right\} h(\mathbf{x}) d\mathbf{x}$$

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

$$\frac{\partial^m}{\partial \eta_1^{m_1} \dots \partial \eta_k^{m_k}} G(\eta) = \int \frac{\partial^m}{\partial \eta_1^{m_1} \dots \partial \eta_k^{m_k}} g(x) \exp \left\{ \sum_{i=1}^k \eta_i T_i(\mathbf{x}) \right\} h(\mathbf{x}) d\mathbf{x}$$

即积分和求偏导顺序可交换。

定理 1.5.2 (*)

如果样本分布有指数族的自然形式, 则参数空间为凸集^a。

^a凸集: 集合内任意两点连线上的点仍在集合内。

证明 利用不等式:

$$\mathbb{E}(U^\alpha V^{1-\alpha}) \leq (\mathbb{E}U)^\alpha (\mathbb{E}V)^{1-\alpha}$$

其中 U, V 是非负随机变量。

兰老师没有提及该定理。

1.6 充分统计量

充分性原理: 如果 $T(\mathbf{x})$ 是 θ 的一个充分统计量, 则 θ 的任意依赖于样本 $\mathbf{x}$ 的推断都可以由 $T(\mathbf{x})$ 的值完成。即, 如果 $T(\mathbf{x}) = T(\mathbf{y})$, 则无论观测到的样本是 $X = \mathbf{x}$ 还是 $X = \mathbf{y}$, 关于 θ 的推断都完全相同。

1.6.1 定义与例子

定义 1.6.1

设样本分布族

$$\mathcal{F} = \{f(\mathbf{x}|\theta) : \theta \in \Theta\}$$

$T = T(X_1, \dots, X_n)$ 是统计量, 若在已知 T 的条件下, $(X_1, \dots, X_n)$ 的条件分布与 θ 无关, 则称 T 为充分统计量。

θ 的充分统计量 $T(\mathbf{x})$ 提炼了样本中有关 θ 的全部信息, 即除充分统计量的值之外, 样本中其余数据不能再提供关于 θ 的任何更多的信息。

例 1.6.1. 二项分布总体

总体 $X \sim B(1, \theta)$, 样本 $\mathbf{X} = (X_1, \dots, X_n)$, 证明 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 是 θ 的充分统计量。

例 1.6.2. 正态分布总体

总体 $X \sim N(\theta, \sigma^2)$, 样本 $\mathbf{X} = (X_1, \dots, X_n)$, 证明 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 是 θ 的充分统计量。

例 1.6.3. 指数分布总体

总体 $X \sim \exp(\theta)$, 样本 $\mathbf{X} = (X_1, \dots, X_n)$, 证明 $T(\mathbf{X}) = \sum_{i=1}^n X_i$ 是 θ 的充分统计量。

定理 1.6.1 (因子分解定理)

设样本 $(X_1, \dots, X_n)$ 的概率函数为 $f_\theta(x_1, \dots, x_n)$, T 为统计量, 则 T 为充分统计量 $\Leftrightarrow$

$$f_\theta(x_1, \dots, x_n) = g_\theta(T(x_1, \dots, x_n))h(x_1, \dots, x_n)$$

$h(x_1, \dots, x_n)$ 和 θ 无关。

证明 设 $T = (T_1, \dots, T_k)$, 补充 $W = (W_1, \dots, W_{n-k})$ 使得 X 和 (T, W) 一一对应, 则

$$\frac{\partial(X_1, \dots, X_n)}{\partial(T_1, \dots, T_k, W_1, \dots, W_{n-k})} = \left(\frac{\partial(t, w)}{\partial x} \right)^{-1} := J^{-1}$$

与 θ 无关。

充分性: 设 $X \sim f_\theta(x) = g_\theta(t(x))h(x)$, 于是

$$(T, W) \sim g_\theta(t) \cdot h(x(t, w)) \cdot |J|^{-1} := g_\theta(t) \cdot H(t, w)$$

所以

$$W|T \sim \frac{g_\theta(t)H(t, w)}{\int g_\theta(t)H(t, w)dw} = \frac{H(t, w)}{\int H(t, w)dw}$$

与 θ 无关, 所以 $X|T$ 的分布也与 θ 无关, T 是充分统计量。

必要性: 设 T 是充分统计量, 则 $X|T$ 条件分布与 θ 无关, 则

$$W|T \sim G(t, w)$$

与 θ 无关,

$$(T, W) \sim g_\theta(t) \cdot G(t, w)$$

X 与 (T, W) 一一对应, 所以

$$X \sim f_\theta(x_1, \dots, x_n) = g_\theta(t(x_1, \dots, x_n))G(t(x_1, \dots, x_n), w(x_1, \dots, x_n)) \cdot |J| := g_\theta(t(x)) \cdot h(x)$$

推论 1.6.1

设 $T = T(\mathbf{X})$ 是 θ 的充分统计量, 若存在一个可测函数 φ 和另一个统计量 S 使得 $T = \varphi(S)$, 则 S 也是 θ 的充分统计量。

推论 1.6.2

样本 $X = (X_1, \dots, X_k)$ 为服从概率密度函数

$$f(\mathbf{x}|\boldsymbol{\theta}) = c(\boldsymbol{\theta})\exp\{\mathbf{Q}(\boldsymbol{\theta})^T T(\mathbf{x})\}h(\mathbf{x})$$

的指数族中抽取的简单样本, 则 $T(X)$ 为充分统计量。



例 1.6.4.

$X_1, \dots, X_n \text{ iid. } \sim N(\theta, 1)$, 证明 $\bar{X}$ 是充分统计量。

证明 $(X_1, \dots, X_n) \rightarrow (T, W_1, \dots, W_{n-1})$, 令

$$\begin{pmatrix} T \\ W_1 \\ \vdots \\ W_{n-1} \end{pmatrix} = A \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

取 $A = \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix}$, $A^{-1} = A'$, 且令 $\mathbf{a}_1 = \frac{1}{\sqrt{n}}\mathbf{1} = \frac{1}{\sqrt{n}}(1, \dots, 1)'$.

则

$$T = \sqrt{n}\bar{X} \sim \sqrt{n}N(\theta, \frac{1}{n})$$

$$W_1, \dots, W_{n-1} \text{ iid. } \sim N(0, 1)$$

且 T 与 $W_1, \dots, W_{n-1}$ 是独立的, 于是 $W|T$ 的分布就是 W 的分布, 与 θ 无关, 所以 $\bar{X} = \frac{1}{\sqrt{n}}T$ 是充分统计量。

注 如果本题中取 $T_1 = X_1$, 不是充分统计量, 因为 X_2 等与 X_1 独立, 即使给定 X_1 , $X_2 \sim N(\theta, 1)$, 与 θ 有关。

例 1.6.5.

$X_1, \dots, X_n \text{ iid. } \sim N(a, \sigma^2)$, 参数 $\theta = (a, \sigma^2)$, 证明 $(\bar{X}, S^2)$ 是充分统计量。

引理 1.6.1

设 $T = T(X_1, \dots, X_n)$ 为充分统计量, $S = \varphi(T(X_1, \dots, X_n))$, S 与 T 一一对应 (例如 φ 严格单调), 则 S 也是充分统计量。

证明

$$\{X : T(x) = t_0\} = \{X : \varphi(T) = \varphi(t_0)\}$$

$X|T = t_0$ 与 θ 无关, 于是

$$\{S = \varphi(t_0)\} = \{T = t_0\}$$

所以 $X|S = \varphi(t_0)$ 与 θ 无关。



证明 记 $T = (T_1, T_2) = (\bar{X}, S^2)$, $\theta = (\theta_1, \theta_2) = (a, \sigma^2)$,

$$(X_1, \dots, X_n) \sim (2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i - a)^2\right\}$$

因为

$$\begin{aligned} (x_i - a)^2 &= (x_i - \bar{x} + \bar{x} - a)^2 \\ &= (x_i - \bar{x})^2 + (\bar{x} - a)^2 + 2(\bar{x} - a)(x_i - \bar{x}) \end{aligned}$$

所以

$$\begin{aligned}\sum_{i=1}^n (x_i - a)^2 &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2 + 2(\bar{x} - a) \sum_{i=1}^n (x_i - \bar{x}) \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2 + 2(\bar{x} - a)(n\bar{x} - n\bar{x}) \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2\end{aligned}$$

从而

$$(X_1, \dots, X_n) \sim (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2 \right] \right\}$$

所以 $(\sum (x_i - \bar{x})^2, \bar{x})$ 为充分统计量，方差多一个 $\frac{1}{n-1}$ ，根据引理这并不影响。

例 1.6.6.

$X_1, \dots, X_n$ iid. $\sim \exp\{\lambda\}$ ，即

$$X_i \sim \lambda e^{-\lambda x} I_{(0, \infty)}(x)$$

于是

$$(X_1, \dots, X_n) \sim \lambda^n e^{-\lambda \sum x_i} \prod_{i=1}^n I_{(0, \infty)}(x_i)$$

于是 $\sum X_i$ 为充分统计量。

例 1.6.7.

$X_1, \dots, X_n$ iid. $\sim U(0, \theta)$ ，则

$$(X_1, \dots, X_n) \sim \frac{1}{\theta^n} \prod_{i=1}^n I_{(0, \theta)}(x_i) = \frac{1}{\theta^n} I_{\{0 < X_{(n)} < \theta\}}$$

所以， $X_{(n)}$ 是一个充分统计量。

例 1.6.8.

$X_1, \dots, X_n$ iid. $\sim U(a, b)$ ，则

$$(X_1, \dots, X_n) \sim (b-a)^{-n} \prod_{i=1}^n I_{(a, b)}(x_i) = (b-a)^{-n} I_{\{a < X_{(1)} \leq X_{(n)} < b\}}$$

所以， $(X_{(1)}, X_{(n)})$ 是一个充分统计量。

例 1.6.9.

$X_1, \dots, X_n$ iid. $\sim U(-\frac{1}{2} + \theta, \frac{1}{2} + \theta)$ ，

$$(X_1, \dots, X_n) \sim \prod_{i=1}^n I_{(-\frac{1}{2} + \theta, \frac{1}{2} + \theta)}(x_i) = I_{\{-\frac{1}{2} + \theta < X_{(1)} \leq X_{(n)} < \frac{1}{2} + \theta\}}$$

设 $n = 2$ ， $\bar{X} \sim \varphi(y) I_{(-\frac{1}{2} + \theta, \frac{1}{2} + \theta)}(y)$ ，不是充分统计量。

1.6.2 极小充分统计量 *

统计量是对数据的一种压缩,充分统计量是包含了所有信息的统计量,那么所谓极小充分统计量,就是指在包含所有信息的前提下尽可能压缩得到的统计量。兰老师的课程不对本节内容作要求。

定义 1.6.2

若 T 与 S 为两个充分统计量,且分别生成的 σ 代数满足 $\sigma(T) \subset \sigma(S)$, 即 T 为 S 的可测函数 $T = q(S)$, 称 T 比 S 更简化。

定义 1.6.3

设 T 为样本分布族 $\mathcal{F}$ 的一个充分统计量,若对任意 $\mathcal{F}$ 中的充分统计量 S , 存在可测函数 q 使得 $T(x) = q(S(x))$, 则称 T 为 $\mathcal{F}$ 的极小充分统计量。

例 1.6.10.

总体均值 a , $T = \sum_{i=1}^n X_i$ 是充分统计量, 而

$$S = (S_1, S_2) = \left(\sum_{i=1}^m X_i, \sum_{i=m+1}^n X_i \right)$$

也是充分统计量, T 是其简化。此外, 样本本身也是充分统计量, 是最细化的充分统计量, 没有进行压缩。

定理 1.6.2

设 $f(\mathbf{x}|\theta)$ 是样本 $\mathbf{X}$ 的概率密度/质量函数, 如果存在函数 $T(\mathbf{x})$ 使得任取两个样本点 $\mathbf{x}, \mathbf{y}$, 比值

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$$

只与 θ 有关当且仅当 $T(\mathbf{x}) = T(\mathbf{y})$, 则 $T(\mathbf{x})$ 是 θ 的极小充分统计量。

1.7 完全统计量

1.7.1 定义与例子

定义 1.7.1

设样本分布族

$$\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$$

$T = T(x)$ 为统计量, 若任何满足

$$\mathbb{E}_\theta[\varphi(T(X))] = 0, \forall \theta \in \Theta$$

T 的可测函数 φ , 都有 $\varphi = 0$ a.e. 则称 $T(X)$ 为完全统计量。等价定义:

$$\mathbb{E}_\theta[\varphi(T(X))] = a, \forall \theta \in \Theta \Rightarrow \varphi = a \text{ a.e.}$$

此时取 $\varphi_1 = \varphi - a$ 即可。

从另一个角度看完全性: 设 $T \sim g_\theta(t)$,

$$\mathbb{E}_\theta[\varphi(T(X))] = 0 \Leftrightarrow \int \varphi(t) g_\theta(t) dt = 0, \forall \theta \in \Theta$$

对于每个 θ , $\varphi(t)$ 和 $g_\theta(t)$ 正交, 等价于 $\varphi(t) = 0$, 这说明 $\{g_\theta(t) : \theta \in \Theta\}$ 是完备的。

推论 1.7.1

1. 如果 T 是完全的, 则对于任意可测函数 φ , $\varphi(T)$ 仍是完全的。
2. 对于任意统计量 T , 如果存在非平凡可测函数 V 使得 $V(T)$ 是辅助统计量^a, 则 T 必定不是完全统计量。

^a如果一个统计量 $V(\mathbf{X})$ 的分布不依赖于 θ , 则称其为辅助统计量。



例 1.7.1.

总体 $\sim B(1, \theta)$, 样本 $X = (X_1, \dots, X_n)$, $T(X) = \sum X_i$, 则 T 是完全统计量。

证明 由定义 $T \sim B(n, \theta)$,

$$P_\theta(T = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

设 $\mathbb{E}_\theta[\varphi(T)] = 0$, 则

$$\sum_{k=0}^n \varphi(k) \binom{n}{k} \theta^k (1 - \theta)^{n-k} = 0, \forall \theta \in (0, 1)$$

即

$$\sum_{k=0}^n \varphi(k) \binom{n}{k} \left(\frac{\theta}{1-\theta}\right)^k = 0, \forall \theta \in (0, 1)$$

这是一个关于 $t = \frac{\theta}{1-\theta} \in (0, \infty)$ 的多项式, 只能是其系数 $\varphi(k)$ 全部为零, 这就证明了 T 是完全统计量。

例 1.7.2.

总体 $\sim U(0, \theta)$,

$$T(X) = X_{(n)} \sim \frac{nt^{n-1}}{\theta^n} I_{(0, \theta)}(t)$$

则 T 是完全统计量。

证明

$$\begin{aligned} \mathbb{E}_\theta[\varphi(T(X))] &= \int_0^\theta \frac{n}{\theta^n} \varphi(t) t^{n-1} dt = 0, \forall \theta > 0 \\ \Leftrightarrow \int_0^\theta \varphi(t) t^{n-1} dt &= 0 \\ \Rightarrow \varphi(\theta) \theta^{n-1} &= 0, \theta > 0 \\ \Rightarrow \varphi(\theta) &= 0, \theta > 0 \end{aligned}$$

例 1.7.3.

总体 $\sim N(\theta, 1)$,

$$T(X) = \bar{X} \sim N\left(\theta, \frac{1}{n}\right)$$

则 T 是完全统计量。

证明

$$\begin{aligned}\mathbb{E}_\theta[\varphi(T(X))] &= \int \varphi(t) \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(t-\theta)^2\right\} dt = 0, \forall \theta \\ \Rightarrow \int \varphi(t) e^{-\frac{nt^2}{2}} e^{t\theta} dt &= 0, \forall \theta\end{aligned}$$

由 Laplace 变换或者 Fourier 变换可知,

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

进而 $\varphi(t) = 0$.**例 1.7.4. 充分统计量不一定是完全统计量**总体 $\sim U(-\frac{1}{2} + \theta, \frac{1}{2} + \theta)$, $X = (X_1, \dots, X_n)$,

$$X \sim \prod_{i=1}^n I_{(-\frac{1}{2} + \theta, \frac{1}{2} + \theta)}(x_i) = I_{\{-\frac{1}{2} + \theta < X_{(1)} \leq X_{(n)} < \frac{1}{2} + \theta\}}$$

所以 $(X_{(1)}, X_{(n)})$ 是充分统计量。令 $Y_i = X_i - \theta + \frac{1}{2} \sim U(0, 1)$, 与 θ 无关, 所以 $Z = X_{(n)} - X_{(1)} = Y_{(n)} - Y_{(1)}$ 的分布与 θ 无关。于是

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

定义

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

则 $\mathbb{E}_\theta[\varphi(Z)] = 0$, 但是 $\varphi \neq 0$.**1.7.2 指数族中统计量的完全性****定理 1.7.1**

设

$$X = (X_1, \dots, X_n) \sim c(\theta) \exp\left\{\sum_{i=1}^k \theta_i T_i(x)\right\} h(x), \theta \in \Theta^*$$

即有指数族的自然形式, 若 $\Theta^* \subset \mathbb{R}^k$ 且 Θ^* 有内点, 则 $T = (T_1, \dots, T_k)$ 为完全统计量。

证明比较啰嗦, 我们直接承认它。

例 1.7.5.条件同例 1.7.1, 用上述定理证明 T 为完全统计量。

证明

$$X \sim \theta^{T(x)} (1 - \theta)^{n - T(x)} = (1 - \theta)^n \exp\left\{T(x) \ln \frac{\theta}{1 - \theta}\right\}$$

令 $\varphi = \ln \frac{\theta}{1 - \theta} \Rightarrow \theta = \frac{e^\varphi}{e^\varphi + 1}$, 于是参数空间为

$$\Theta^* = \{\varphi : \varphi \in \mathbb{R}\}$$

这是一条直线, 一定有内点。

例 1.7.6.

总体 $\sim N(a, \sigma^2)$,

$$\Theta = \{\theta = (a, \sigma^2) : a \in \mathbb{R}, \sigma^2 > 0\}$$

则 $(\bar{X}, S^2)$ 为充分完全统计量。

证明 在例 1.5.6 中我们已经说明了：

$$X = (X_1, \dots, X_n) \sim c^*(\varphi) \exp\{\varphi_1 T_1(x) + \varphi_2 T_2(x)\}$$

这里 $\varphi_1 = \frac{a}{\sigma^2}, \varphi_2 = -\frac{1}{2\sigma^2}, T_1 = \sum x_i, T_2 = \sum x_i^2$. 于是

$$\Theta^* = \{(\varphi_1, \varphi_2) : \varphi_1 \in \mathbb{R}, \varphi_2 < 0\}$$

这是下半平面，有内点，于是 (T_1, T_2) 为充分完全统计量，进而 $(\bar{X}, S^2)$ 为充分完全统计量。

1.7.3 有界完全统计量

定义 1.7.2

设 φ 为任一 (a.e.) 有界的可测函数，若对于每个 $\theta \in \Theta$ ，都有

$$\mathbb{E}_\theta[\varphi(T)] = 0 \Rightarrow \varphi(T) = 0 \text{ a.e.}$$

则称 $T(X)$ 为有界完全统计量。

注 实际上有界完全统计量是把条件放宽了：原来是所有满足条件的 φ 都要几乎处处为零，现在只要求有界的 φ ，无界的 φ 不需要几乎处处为零了。所以完全统计量一定是有界完全统计量，反过来不成立。

定理 1.7.2 (Basu)

设样本分布族 $\mathcal{F} = \{f(x, \theta) | \theta \in \Theta\}$ ，简单随机样本 $X = (X_1, \dots, X_n)$ ， $T(X)$ 为有界充分完全统计量，若统计量 $V(X)$ 的分布与 θ 无关，则对于 $\forall \theta \in \Theta$ ， $V(X)$ 与 $T(X)$ 相互独立。

证明 对于 $\forall A \in \mathcal{B}$ ，即 Borel 集，记 $P(V \in A) = P_A$ ，与 θ 无关。记 $\eta_A(t) = P(V \in A | T = t)$ ，

$$P(V \in A) = \mathbb{E}I_{\{V \in A\}}$$

$$\eta_A(t) = \mathbb{E}[I_{\{V \in A\}} | T = t]$$

由条件期望平滑公式¹：

$$\mathbb{E}\eta_A(t) = \mathbb{E}_T[\mathbb{E}[I_{\{V \in A\}} | T = t]] = \mathbb{E}I_{\{V \in A\}} | T = t = p_A$$

与 θ 无关，又因为 $\zeta_A(T)$ 为 T 的有界可测函数，

$$\varphi(t) = \eta_A(t) - p_A$$

仍为有界可测函数，由于 T 为有界充分完全统计量，所以

$$P(V \in A | T = t) = p_A$$

是与 t 无关的，因此 V 与 T 相互独立。

例 1.7.7.

总体 $\sim N(\theta, 1)$ ， $X = (X_1, \dots, X_n)$ ， $R(X) = X_{(n)} - X_{(1)}$ ，证明 $R(X)$ 和 $\bar{X}$ 相互独立。

¹就是重期望公式

证明 已经证明 $T(X)$ 是充分完全统计量, 只要证 R 的分布与 θ 无关即可, 而 $Y_i = X_i - \theta \sim N(0, 1)$ 与 θ 无关, $R = X_{(n)} - X_{(1)} = Y_{(n)} - Y_{(1)}$ 就与 θ 无关。

例 1.7.8.

总体 $\sim N(0, \sigma^2)$, $\sigma^2 > 0$,

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

则 $T(X)$ 和 $V(X)$ 相互独立。

证明 已经证明 $T(X)$ 是充分完全统计量, 只要证明 $V(X)$ 分布与 θ , 即 σ^2 无关。令 $Y_i = \frac{X_i}{\sigma} \sim N(0, 1)$, 与 σ 无关, 而

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

与 θ 无关。

1.8 统计量的极限分布 *

兰老师没有提及本节内容。

定义 1.8.1

样本 $(X_1, \dots, X_n)$,

1. n 固定, 统计量 $T = T(X_1, \dots, X_n)$ 分布的性质称为小样本性质, 例如 t_n .
2. $n \rightarrow \infty$, 统计量 $T = T(X_1, \dots, X_n)$ 分布的性质称为大样本性质。例如大数定律和中心极限定理:

$$\begin{aligned} \bar{X} &\xrightarrow{P} \mathbb{E}X \\ \frac{\bar{X} - a}{\sigma\sqrt{n}} &\xrightarrow[L(D)]{L} N(0, 1) \end{aligned}$$

林德伯格条件下的中心极限定理:

$$b_n(X_n - a_n) \xrightarrow{L} N(0, 1)$$

$$a_n = \mathbb{E}X_n + o(b_n^{-1}), \quad b_n \rightarrow \infty,$$

$$b_n^{-1} = (\text{Var } X_n)^{\frac{1}{2}} c_n, \quad c_n = 1 + \alpha_n$$

例 1.8.1. $X_{(m)}$ 的极限分布

$(X_1, \dots, X_n) \Rightarrow (X_{(1)}, \dots, X_{(n)})$, 假定 $\forall p \in (0, 1)$,

$$\frac{m}{n} - p = o\left(\frac{1}{\sqrt{n}}\right)$$

设 ξ_p 是 X_i 分布的 p 分位数, 即 $\int_{-\infty}^{\xi_p} dF(x) = p$, 设 $f(x)$ 在 ξ_p 处连续且不等于零, 这是为了保证 ξ_p 的唯一性, 在这些条件下,

$$\sqrt{n}f(\xi_p)(X_{(m)} - \xi_p)/\sqrt{p(1-p)} \xrightarrow{L} N(0, 1)$$

证明

$$X_{(m)} \sim f(x) = \frac{n!}{(m-1)!(n-m)!} [F(x)]^{m-1} [1-F(x)]^{n-m} f_1(x)$$

经验分布函数 $F_n(x) = \frac{1}{n} \sum I_{X_i \leq x}$, $X_{(m)} = F_n^{-1}(\frac{m}{n})$, 反函数断点怎么定义?

$$\mathbb{E}X_{(m)} = \mathbb{E}F_n^{-1}(\frac{m}{n}) \sim \mathbb{E}F^{-1}(p) = \xi_p$$

设 $a_n(X_{(m)} - \xi_p) \xrightarrow{L} N(0, 1)$, a_n 待定,

$$\begin{aligned} Y &= a_n(X_{(m)} - \xi_p) \sim g_m(x) \\ &= f(\xi_p + \frac{x}{a_n}) \frac{1}{a_n} \end{aligned}$$

记 $t = \xi_p + \frac{x}{a_n}$,

$$\begin{aligned} Y \sim g_m(x) &= m \binom{n}{m} [F(t)]^{m-1} [1 - F(t)]^{n-m} f_1(t) \frac{1}{a_n} \\ &= n \binom{n}{m} [F(t)]^m [1 - F(t)]^{n-m} \frac{m}{n \cdot F(t)} \frac{f_1(t)}{f_1(\xi_p)} f_1(\xi_p) \\ &\approx \frac{n \cdot f_1(\xi_p)}{a_n} \binom{n}{m} [F(t)]^m [1 - F(t)]^{n-m} \end{aligned}$$

我们知道若 $Y \sim B(n, p)$, 则

$$\begin{aligned} P(Y = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{1}{\sqrt{2\pi np(1-p)}} \exp \left\{ -\frac{1}{2} \left(\frac{k - np}{\sqrt{np(1-p)}} \right)^2 \right\} \end{aligned}$$

取 $p = F(t)$,

$$\begin{aligned} g_m(x) &\sim \frac{n f_1(\xi_p)}{a_n} \frac{1}{\sqrt{2\pi n F(t) [1 - F(t)]}} \exp \left\{ -\frac{1}{2} \left(\frac{m - n F(t)}{\sqrt{n F(t) [1 - F(t)]}} \right)^2 \right\} \\ F(t) &= F(\xi_p + \frac{x}{a_n}) = F(\xi_p) + f(\xi_p) \frac{x}{a_n} + o(\frac{1}{a_n}) \\ \Rightarrow m - n F(t) &= m - np - \frac{n f_1(\xi_p) x}{a_n} \approx -\frac{n f_1(\xi_p) x}{a_n} \end{aligned}$$

(这里假定 $m - np = o(\frac{n}{a_n})$)

$$\Rightarrow \left| \frac{m - n F(t)}{\sqrt{n F(t) [1 - F(t)]}} \right| \approx \frac{n f_1(\xi_p) x}{\sqrt{np(1-p)}}$$

希望右边 $\rightarrow 1$, 于是取

$$a_n = \frac{\sqrt{n} f_1(\xi_p)}{\sqrt{p(1-p)}}$$

所以

$$g_m(x) \approx \frac{n f_1(\xi_p)}{\sqrt{\frac{n}{p(1-p)}} f_1(\xi_p)} \cdot \frac{1}{\sqrt{2\pi np(1-p)}} \exp \left\{ -\frac{x^2}{2} \right\} = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2} \right\}$$

要看懂这个证明, 有两个要点: 泰勒展开和局部极限定理 (用在近似二项分布的那里)。

第 2 章 点估计

估计是统计推断之一，是指对参数 θ 或者参数的函数 $g(\theta)$ 进行估计。

总体 $\sim f(x, \theta)$, $X = (X_1, \dots, X_n)$, $\theta = (\theta_1, \dots, \theta_k) \in \Theta$, 一般来说 θ 未知, 要由样本对 θ 作一个估计, 或者对 $g(\theta)$ 做一个估计, 这种估计被称为点估计。通常用 $\hat{g}(X)$ 记为 $g(\theta)$ 的一个点估计。

本章我们将介绍点估计的两种方法: 矩估计和极大似然估计, 这两种方法有着不同的统计思想。除此之外还将介绍如何评判点估计的优劣, 以及 UMVUE, 尽管它一般情况下很难求出来。

2.1 矩估计

定义 2.1.1

设 $X = (X_1, \dots, X_n)$ 为总体的一个简单随机样本, $k \in \mathbb{N}$, 分别称

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

为 k 阶样本原点矩和 k 阶样本中心矩, 记

$$\alpha_k = \mathbb{E}X^k, \quad \mu_k = \mathbb{E}(X - \alpha_1)^k$$

分别成为 k 阶总体原点矩和 k 阶总体中心矩。由大数定理, 样本矩依概率收敛到总体矩:

$$a_{nk} \xrightarrow{P} \alpha_k, \quad m_{nk} \xrightarrow{P} \mu_k$$

在实际应用中, 用样本矩代替总体矩, 得到的估计被称为矩估计:

$$g(\theta) = G(\alpha_1, \dots, \alpha_k, \mu_2, \dots, \mu_l)$$

则 $\hat{g}(X) = G(a_{n1}, \dots, a_{nk}, m_{n2}, \dots, m_{nl})$ 称为 $g(\theta)$ 的一个矩估计。矩估计一般用 M 表示。广义矩方法: 不一定非要用 X 的矩, 可以用 $h(X)$ 的矩来估计。

但样本数较少时要进行修正:

$$\mathbb{E}m_{n2} = \frac{1}{n} \mathbb{E} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) = (\mu_2 + \alpha_1^2) - (\text{Var}(\bar{X}) + \alpha_1^2) = \left(1 - \frac{1}{n}\right) \mu_2$$

于是令 $\hat{\mu}_2 = \frac{n}{n-1} m_{n2} = S^2$, 详见点估计的优良准则。

由此可见, 使用矩估计的前提是要用总体矩表示出参数或者参数的函数。

例 2.1.1.

总体均值 $a = \mathbb{E}X$ 的估计 $\hat{a}_M = \bar{X}$, 总体的方差 $\sigma^2 = \text{Var}(X)$ 的估计 $\hat{\sigma}_M^2$, $\tilde{\sigma}_M^2 = S^2$.

例 2.1.2.

总体 $\sim P(\lambda)$, $\mathbb{E}X = \lambda$, $\text{Var}(X) = \lambda$, 所以有两种矩估计:

$$\hat{\lambda}_{M1} = \bar{X}, \hat{\lambda}_{M2} = S^2$$

例 2.1.3.

总体 $\sim U(\theta_1, \theta_2)$, $\mathbb{E}X = \frac{1}{2}(\theta_1 + \theta_2)$, $\text{Var}(X) = \frac{1}{12}(\theta_2 - \theta_1)^2$, 用 $\bar{X}$ 代替 $\mathbb{E}X$, S^2 代替 $\text{Var}(X)$,

$$\begin{cases} \frac{1}{2}(\hat{\theta}_{M1} + \hat{\theta}_{M2}) = \bar{X} \\ \frac{1}{12}(\hat{\theta}_{M2} - \hat{\theta}_{M1})^2 = S^2 \end{cases}$$

解此方程即可。

例 2.1.4.

总体 $\sim N(a, \sigma^2)$, 求 $P(X > 1)$ 的矩估计。

$$\begin{aligned} P(X > 1) &= P\left(\frac{X - a}{\sigma} > \frac{1 - a}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{1 - a}{\sigma}\right) = \Phi\left(\frac{a - 1}{\sigma}\right) \\ \Rightarrow \hat{P}_M(X > 1) &= \Phi\left(\frac{\bar{X} - 1}{S}\right) \end{aligned}$$

例 2.1.5.

总体 X 有如下分布:

X	0	1	3
p	θ	2θ	$1 - 3\theta$

求 $\hat{\theta}_M$, 设 $n = 10$, 样本为 $(0, 0, 0, 1, 1, 1, 3, 3, 3, 3)$.

$$\mathbb{E}X = 3 - 7\theta$$

故 $\bar{X} = 3 - 7\hat{\theta}_M = \frac{3}{2} \Rightarrow \hat{\theta}_M = 0.2143$.

例 2.1.6.

总体 $\sim f(x, \theta)$,

$$f(x, \theta) = \frac{\theta_2}{\Gamma\left(\frac{1+\theta_1}{\theta_2}\right)} x^{\theta_1} \exp\{-x^{\theta_2}\} I_{(0, \infty)}(x)$$

$\theta = (\theta_1, \theta_2)$, $\theta_1 > -1$, $\theta_2 > 0$, 求 $\hat{\theta}_{M1}$, $\hat{\theta}_{M2}$.

$$\mathbb{E}X = \frac{\theta_2}{\Gamma\left(\frac{1+\theta_1}{\theta_2}\right)} \int_0^\infty x^{\theta_1+1} e^{-x^{\theta_2}} dx$$

令 $x^{\theta_2} = t$, $x^{\theta_1+1} = t^{\frac{\theta_1+1}{\theta_2}}$,

$$\mathbb{E}X = \frac{1}{\Gamma\left(\frac{1+\theta_1}{\theta_2}\right)} \Gamma\left(\frac{\theta_1+2}{\theta_2}\right)$$

$$\mathbb{E}X^2 = \frac{\theta_2}{\Gamma\left(\frac{1+\theta_1}{\theta_2}\right)} \int_0^\infty x^{\theta_1+2} e^{-x^{\theta_2}} dx = \frac{1}{\Gamma\left(\frac{1+\theta_1}{\theta_2}\right)} \Gamma\left(\frac{\theta_1+3}{\theta_2}\right)$$

从中解出 θ_1 和 θ_2 , 用 $\bar{X}$ 代替 $\mathbb{E}X$, a_{n2} 代替 $\mathbb{E}X^2$ 即可 (只能得到数值解)。

例 2.1.7.

总体 $\sim \frac{1}{\pi(1+(x-\theta)^2)}$, 期望不存在, 无法用矩估计。

例 2.1.8.

$$\beta_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{\mathbb{E}\left(\frac{X-\alpha_1}{\sigma}\right)^3}{\left(\mathbb{E}\left(\frac{X-\alpha_1}{\sigma}\right)^2\right)^{\frac{3}{2}}}$$

被称为偏度系数。

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

被称为峰度系数。标准正态分布的峰度系数为 3。

$$\nu = \frac{\sqrt{\mu_2}}{\alpha_1}$$

被称为变异系数。这三个例子都是矩的函数, 有了样本也可以进行矩估计。

例 2.1.9. 协方差及相关系数的矩估计

(X, Y) ,

$$\text{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$$

$$\hat{\text{Cov}}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

$$\hat{\rho}_{XY} = r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{m_{11}}{\sqrt{m_{20}m_{02}}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

这里 $m_{kl} = \frac{1}{n} \sum (x_i - \bar{x})^k (y_i - \bar{y})^l$.

2.2 极大似然估计

极大似然估计思想：样本被观测到之后，希望给出一个概率结构，使得它在样本处的概率达到最大。即估计 $\hat{\theta}$ ，使得 $f(X, \hat{\theta}) = \max$.

例 2.2.1.

N 条鱼，捕捉到 1000 条，尾巴涂成红色。过一段时间后，又捕捉到 1000 条，红尾鱼有 100 条，估计池塘中鱼的数量。

解答： X 表示第二次捕到鱼中红尾鱼的条数，则 X 服从超几何分布：

$$P_N(X = k) = \frac{\binom{1000}{k} \binom{N-1000}{1000-k}}{\binom{N}{1000}}, k = 0, 1, \dots, 1000$$

极大似然估计的思想：求 N ，使得上述概率在 $k = 100$ 处最大。记 $p_N = P_N(X = k)$ ，求 k 使得

$$\frac{p_N}{p_{N+1}} \geq 1, \frac{p_N}{p_{N-1}} \geq 1$$

也就是

$$\frac{p_N}{p_{N+1}} = \frac{(N+1-2000+k)(N+1)}{(N+1-2000)^2} \geq 1$$

$$\frac{p_N}{p_{N-1}} = \frac{(N-1000)^2}{N(N-2000+k)} \geq 1$$

令 $k = 100$ ，得到 $10^4 - 1 \leq N \leq 10^4$ ，取 $\hat{N}_L = 10^4$ ，或者 $10^4 - 1$ 均可。

定义 2.2.1 (似然函数)

设样本 $X = (X_1, \dots, X_n) \sim f(\mathbf{x}, \boldsymbol{\theta})$ ，如果固定 $\boldsymbol{\theta} \in \Theta$ ，作为 $\mathbf{x}$ 的函数，就称为样本的密度函数（概率函数）；固定 $x \in \mathcal{X}$ ，作为 $\boldsymbol{\theta}$ 的函数，就称为似然函数，记作 $L(\mathbf{x}, \boldsymbol{\theta})$ 。似然函数表达了给定样本下概率结构的变化：

$$f(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} \approx P(x_i \leq X_i \leq x_i + dx_i, 1 \leq i \leq n)$$

定义 2.2.2 (极大似然估计)

若 $\hat{\theta}(X)$ 为一统计量, 满足

$$L(\hat{\theta}(X)) = \sup_{\theta \in \Theta} L(\theta)$$

则称 $\hat{\theta}(X)$ 是 θ 的极大似然估计。若要估计 $g(\theta)$, 则称 $g(\hat{\theta}(X))$ 是 $g(\theta)$ 的极大似然估计。极大似然估计 MLE 一般用 L 表示。

通常样本 $X = (X_1, \dots, X_n)$ 是简单随机样本, 总体 $\sim f(x, \theta)$,

$$X = (X_1, \dots, X_n) \sim \prod_{i=1}^n f(x_i, \theta)$$

记作 $L(x_1, \dots, x_n, \theta) = L(x, \theta)$,

$$\ln L(x, \theta) = \sum_{i=1}^n \ln f(x_i, \theta)$$

因为 $\ln x$ 增, 所以实际上

$$\hat{\theta}_L = \arg \sup_{\theta \in \Theta} \ln L(x, \theta)$$

至此, 我们把问题转化为了求似然函数的对数的极大值点。一般情况下, 令一阶导数为零 (验证二阶导数 Hessian 矩阵负定) 即可。 $\ln L(x, \theta)$ 对 $x_1, \dots, x_n$ 作 Taylor 展开:

$$f(x, \theta) = f(x_0, \theta) + \left(\frac{\partial f}{\partial \theta} \right)' \Big|_{\theta=\theta_0} (\theta - \theta_0) + \frac{1}{2} (\theta - \theta_0)' H (\theta - \theta_0) + o(\|\theta - \theta_0\|^2)$$

其中

$$\frac{\partial f}{\partial \theta} = \left(\frac{\partial f}{\partial \theta_1}, \dots, \frac{\partial f}{\partial \theta_k} \right)'$$

$$\theta = (\theta_1, \dots, \theta_k)', \theta_0 = (\theta_{10}, \dots, \theta_{k0})'$$

$$H = \left(\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \right)$$

Hessian 矩阵

θ_0 是驻点的情况下, 若 $H < 0$, 则 θ_0 为极大值点。

推论 2.2.1

若 $\prod f(x_i, \theta)$ 有指数族的自然形式, $\ln L(x, \theta)$ 的驻点在 Θ^* 内部, 则该驻点就是 $\hat{\theta}_L$ 。

证明 这是一个上凸函数, 驻点一定是极大值点, 此时无需验证 Hessian 矩阵负定! 太棒了。

例 2.2.2.

总体 $\sim N(a, \sigma^2)$, 样本 $X = (X_1, \dots, X_n)$ 求 $\hat{a}_L$ 和 $\hat{\sigma}_L^2$ 。

解答: 似然函数

$$L(x, \theta) = f(x, a, \sigma^2) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum (x_i - a)^2 \right\}$$

其中 $\theta = (\theta_1, \theta_2) = (a, \sigma^2)$,

$$\ln L(x, \theta) = c - \frac{n}{2} \ln \theta_2 - \frac{1}{2\theta_2} \sum (x_i - \theta_1)^2$$

$$\frac{\partial \ln L}{\partial \theta_1} = -\frac{1}{2\theta_2} \cdot 2 \sum (x_i - \theta_1)(-1) = 0 \Leftrightarrow \theta_1 = \frac{1}{n} \sum x_i \Rightarrow \hat{\theta}_{L1} = \bar{X}$$

$$\frac{\partial \ln L}{\partial \theta_2} = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \sum (x_i - \theta_1)^2 = 0 \Leftrightarrow \theta_2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \Rightarrow \hat{\theta}_{L2} = m_{n2}$$

和矩估计的结论是一样的。

例 2.2.3.

总体 $X \sim \exp\{\lambda\}$,

$$f(x, \lambda) = \lambda e^{-\lambda x} I_{(0, \infty)}(x)$$

$$L(x, \lambda) = \lambda^n e^{-\lambda \sum x_i} \prod I_{(0, \infty)}(x_i)$$

$$\ln L(x, \lambda) = n \ln \lambda - \lambda \sum x_i$$

$$\frac{d \ln L}{d \lambda} = \frac{n}{\lambda} - \sum x_i = 0 \Rightarrow \hat{\lambda}_L = \frac{1}{\bar{X}}$$

例 2.2.4.

总体 $\sim f(x, \theta) = \frac{1}{\pi(1+(x-\theta)^2)}$,

$$f(x, \theta) = \prod_{i=1}^n \frac{1}{\pi(1+(x_i-\theta)^2)}$$

$$\ln L(x, \theta) = -\sum_{i=1}^n \ln(1+(x_i-\theta)^2) + C$$

$$\frac{d \ln L}{d \theta} = \sum_{i=1}^n \frac{2(x_i-\theta)}{1+(x_i-\theta)^2} = 0$$

这个方程交给计算机来解吧。但是因为柯西分布是对称的，所以可以用样本中位数来估计 θ ，这既不是矩估计，也不是极大似然估计。

例 2.2.5.

总体 X 有如下分布：

X	0	1	3
p	θ	2θ	$1-3\theta$

样本 $X = (0, 0, 0, 1, 1, 1, 3, 3, 3, 3)$ ，求 θ 的极大似然估计。

解答：

$$L(x, \theta) = \theta^3 (2\theta)^3 (1-3\theta)^4$$

$$\ln L(x, \theta) = C + 6 \ln \theta + 4 \ln(1-3\theta)$$

$$\frac{d \ln L}{d \theta} = \frac{6}{\theta} + \frac{-12}{1-3\theta} = 0 \Rightarrow \hat{\theta}_l = 0.2$$

例 2.2.6.

总体 $\sim U(0, \theta)$ ，求 θ 的极大似然估计。

解答:

$$L(x, \theta) = \prod_{i=1}^n \frac{1}{\theta} I_{(0, \infty)}(x) = \theta^{-n} I_{(0 < x_{(n)} < \theta)}$$

$$\Rightarrow \hat{\theta}_l = X_{(n)}$$

例 2.2.7.

元件的寿命一般假定为指数分布 $\exp(\lambda)$ 。元件寿命试验有两种方法:

1. 定数截尾: n 个元件同时开始运行, 一段时间后, 恰观测到 r 个失效元件时, 试验停止。设 n 个样本的寿命分别为 $X_1, \dots, X_n$, 我们观测到的是 r 个失效元件的寿命 $X_{(1)}, \dots, X_{(r)}$,

$$(X_{(1)}, \dots, X_{(r)}) \sim \frac{n!}{(n-r)!} f(t_1) \cdots f(t_r) (1-F(t_r))^{n-r} I_{(0 < t_1 \leq t_2 \leq \dots \leq t_r)}$$

2. 定时截尾: n 个元件同时开始运行一段时间 T 后试验停止。设 Y_i 为第 i 个元件的寿命, 令

$$X_i = \begin{cases} Y_i, & Y_i \leq T \\ T, & Y_i > T \end{cases}$$

即 $X_i = Y_i I_{\{Y_i \leq T\}} + T I_{\{Y_i > T\}}$,

$$\begin{aligned} f(t, \lambda) &= f_y(t, \lambda) I_{\{t \leq T\}} + P(Y_i > T) I_{\{t=T\}} \\ &= \lambda e^{-\lambda t} I_{\{t \leq T\}} + e^{-\lambda T} I_{\{t=T\}} \end{aligned}$$

设 n 个元件中有 r 个在 T 之前失效, 不失一般性, 设为前 r 个元件, 则后 $n-r$ 个元件寿命不低于 T ,

$$L(t_1, \dots, t_n, \lambda) = \lambda^r e^{-\lambda(t_1 + \dots + t_r)} e^{-\lambda(n-r)T} = \lambda^r \exp\{-\lambda S\}$$

这里 $S = t_1 + \dots + t_r + (n-r)T$, 马上可推出

$$\hat{\lambda}_l = \frac{r}{S} = \frac{r}{t_1 + \dots + t_r + (n-r)T}$$

例 2.2.8.

事件 $A_1, \dots, A_k$ 构成完备事件群, 即是整个样本空间的一个分割。设 $P(A_i) = p_i$, 做 n 次 Benoulli 试验 ($n > k$), A_i 发生了 n_i 次, 求所有 p_i 的极大似然估计。

解答: 设 x_i 是 n 次试验中事件 A_i 发生的次数,

$$P(X_i = x_i, i = 1, 2, \dots, k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

$$\ln L(x, p_1, \dots, p_k) = C + x_1 \ln p_1 + \dots + x_k \ln p_k$$

这里要利用拉格朗日乘子法, 最终得到驻点

$$\hat{p}_{iL} = \frac{x_i}{n} = \frac{n_i}{n}$$

那怎么证明这是一个极大值点? 要看驻点处的 Hessian 矩阵。

$$\frac{\partial \ln L}{\partial p_i} = \frac{x_i}{p_i}$$

$$\frac{\partial^2 \ln L}{\partial p_i^2} = -\frac{x_i}{p_i^2}, \frac{\partial^2 \ln L}{\partial p_i \partial p_j} = 0$$

所以 $H = -\text{diag}(\frac{x_i}{p_i^2}) < 0$.

2.3 点估计的优良性准则

2.3.1 无偏性

定义 2.3.1

设 $\hat{g}(X)$ 为 $g(\theta)$ 的一个估计, 若 $\mathbb{E}_\theta \hat{g}(X) = g(\theta)$, 对一切 $\theta \in \Theta$, 则称 $\hat{g}(X)$ 为 $g(\theta)$ 的一个无偏估计。
含义: 没有系统误差。

总体均值为 a , 总体方差为 σ^2 ,

$$\hat{a} = \bar{X}, \hat{\sigma}^2 = m_{n2}$$

那么

$$\mathbb{E}\bar{X} = a, \mathbb{E}m_{n2} = \frac{n-1}{n}\sigma^2$$

所以后者系统偏小, 要修正为

$$\hat{\sigma}_1^2 = S^2$$

另外,

$$a_{nk} = \frac{1}{n} \sum_{i=1}^n X_i^k \Rightarrow \mathbb{E}a_{nk} = \alpha_k$$

但是

$$\mathbb{E}m_{nk} = \mathbb{E}\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \neq \mu_k \text{ in usually case}$$

我们来尝试推导 $k=3$ 的情况, 先对样本进行标准化: $Y_i = X_i - a$, 那么 Y_i 期望为零且独立同分布。

$$\begin{aligned} \mathbb{E}m_{n3} &= \frac{1}{n} \sum_{i=1}^n ((X_i - a) - (\bar{X} - a))^3 = \frac{1}{n} \mathbb{E} \sum_{i=1}^n (Y_i - \bar{Y})^3 \\ &= \frac{1}{n} \mathbb{E} \sum_{i=1}^n (Y_i^3 - 3Y_i^2\bar{Y} + 3Y_i\bar{Y}^2 - \bar{Y}^3) \\ &= \frac{1}{n} (n \cdot \mathbb{E}[Y_1^3] - 3n \cdot \mathbb{E}[Y_1^2 \frac{Y_1 + \dots + Y_n}{n}] + 3n \cdot \mathbb{E}[Y_1 \frac{(Y_1 + \dots + Y_n)^2}{n^2}] - n \cdot \mathbb{E}[\frac{(Y_1 + \dots + Y_n)^3}{n^3}]) \\ &= \mu_3 - 3 \cdot \mathbb{E}[Y_1^3/n] + 3 \cdot \mathbb{E}[Y_1^3/n^2] - \mathbb{E}[(Y_1^3 + \dots + Y_n^3)/n^3] \\ &= \mu_3 - \frac{3}{n}\mu_3 + \frac{2}{n^2}\mu_3 = \frac{(n-1)(n-2)}{n^2}\mu_3 \end{aligned}$$

系统偏小, 应当修正为

$$\hat{\mu}_3 = \frac{n^2}{(n-1)(n-2)} m_{n3}$$

其余情况的推导同理。结论是:

$$X_i \text{ iid. } \mathbb{E}X_i = 0 \Rightarrow \mathbb{E}(X_1 + \dots + X_n)^k \sim n^{\frac{k}{2}}$$

例 2.3.1.

总体 $\sim P(\lambda)$, 我们之前给出过估计:

$$\hat{\lambda}_1 = \bar{X}, \hat{\lambda}_2 = S^2$$

两个都是无偏估计: $\mathbb{E}\bar{X} = \mathbb{E}S^2 = \lambda$.

例 2.3.2.

总体 $\sim \exp\{\lambda\}$, 矩估计 $\hat{\lambda}_M$ 是否无偏?

解答: 我们已经计算过

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

求它的期望, 要利用到 Γ 分布:

$$f(x, \lambda, \alpha) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} I_{(0, \infty)}(x)$$

所以

$$\sum_{i=1}^n X_i \sim \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)} I_{(0, \infty)}(x)$$

于是

$$\mathbb{E}\hat{\lambda}_M = n \int_0^\infty x^{-1} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)} dx = \frac{n}{n-1} \lambda$$

所以它是系统偏大的, 修正为

$$\hat{\lambda}_1 = \frac{n-1}{n\bar{X}}$$

例 2.3.3.

总体 $\sim B(1, \theta)$, $g(\theta) = \theta(1-\theta)$ 求 $\hat{\theta}_M$ 和 $\hat{g}_M$, 是否是无偏估计?

解答: $\hat{\theta}_M = \bar{X}$, 无偏; $\hat{g}_M = \bar{X}(1-\bar{X})$,

$$\begin{aligned} \mathbb{E}\hat{g}_M &= \theta - \mathbb{E}\bar{X}^2 \\ &= \theta - \frac{1}{n^2} \mathbb{E}(X_1 + \cdots + X_n)^2 \\ &= \theta - \frac{1}{n^2} (n\theta + n(n-1)\theta) = \frac{n-1}{n} \theta(1-\theta) \end{aligned}$$

系统偏小。

例 2.3.4.

总体 X 有如下分布:

X	0	1	3
p	θ	2θ	$1-3\theta$

设取了 n 次后, 取到 0, 1, 3 的次数分别为 n_1, n_2, n_3 , 求 $\hat{\theta}_l$, 是否是无偏估计?

解答:

$$\begin{aligned} L(x, \theta) &= \theta^{n_1} (2\theta)^{n_2} (1-3\theta)^{n_3} \\ \ln L(x, \theta) &= C + (n_1 + n_2) \ln \theta + n_3 \ln (1-3\theta) \\ \frac{d}{d\theta} \ln L(x, \theta) &= \frac{n_1 + n_2}{\theta} - \frac{3n_3}{1-3\theta} = 0 \end{aligned}$$

可得

$$\hat{\theta}_l = \frac{n_1 + n_2}{3n} = \frac{n - n_3}{3n} = \frac{1}{3} - \frac{n_3}{3n}$$

其期望很容易求得为 $\mathbb{E}\hat{\theta}_l = \frac{1}{3} - \frac{n(1-3\theta)}{3n} = \theta$, 所以是无偏估计。

2.3.2 有效性

定义 2.3.2

设 $\hat{g}_1(X), \hat{g}_2(X)$ 都是 $g(\theta)$ 的无偏估计, 若

$$\text{Var}(\hat{g}_1(X)) \leq \text{Var}(\hat{g}_2(X)), \forall \theta \in \Theta$$

其至少对于某个 $\theta \in \Theta$, 上式严格小于号成立, 则称 $\hat{g}_1(X)$ 更有效。

例 2.3.5.

总体 $\sim P(\lambda)$, 我们有两个无偏估计

$$\hat{\lambda}_1 = \bar{X}, \hat{\lambda}_2 = S^2$$

两个都是无偏估计,

$$\text{Var}(\bar{X}) = \frac{\lambda}{n}$$

$$\text{Var}(S^2) = \frac{\lambda}{n} + \frac{2\lambda^2}{n}(\lambda+1)(\lambda+2) + o(1/n)$$

所以 $\hat{\lambda}_1$ 更有效。一般来说, 低阶矩估计的方差要更小。

例 2.3.6.

总体 $\sim f(x, \theta)$,

$$f(x, \theta) = 2\sqrt{\frac{\theta}{\pi}} e^{-\theta x^2} I_{(0, \infty)}(x)$$

利用公式

$$\int_0^\infty x^\alpha e^{-a^2 x^2} dx = \frac{\Gamma(\frac{\alpha+1}{2})}{2a^{\alpha+1}}$$

可以求出:

$$\mathbb{E}X = \frac{1}{\sqrt{\pi\theta}}, \mathbb{E}X^2 = \frac{1}{2\theta}, \mathbb{E}X^3 = \frac{1}{\sqrt{\pi\theta\theta}}, \mathbb{E}X^4 = \frac{3}{4\theta^2}$$

于是得到两个估计：

$$\hat{g}_1(X) = \left(\frac{\hat{1}}{\hat{\theta}}\right)_1 = \pi \bar{X}^2$$

$$\hat{g}_2(X) = \left(\frac{\hat{1}}{\hat{\theta}}\right)_2 = \frac{2}{n} \sum X_i^2$$

但是

$$\mathbb{E}\hat{g}_1(X) = \frac{1}{\theta} \left(1 + \left(\frac{\pi}{2} - 1\right) \frac{1}{n}\right)$$

不是无偏估计，所以修正为

$$\tilde{g}_1(X) = \frac{\pi}{1 + \left(\frac{\pi}{2} - 1\right) \frac{1}{n}} \bar{X}^2$$

容易验证 $\hat{g}_2(X)$ 是无偏估计。然后我们来比较其方差，太麻烦了！就不详细写了。只需要注意到

$$\mathbb{E}\bar{X}^4 = n^{-4} \left(\mathbb{E} \sum X_i^4 + 4\mathbb{E} \sum X_i^3 X_j + 3\mathbb{E} \sum X_i^2 X_j^2 + 6\mathbb{E} \sum X_i^2 X_j X_k + \mathbb{E} \sum X_i X_j X_k X_l \right)$$

就能计算了。我们直接给出结论：

$$\text{Var}\tilde{g}_1(X) = \frac{3(\pi-2)}{\pi^2\theta^2} \cdot \frac{1}{n} + o\left(\frac{1}{n}\right)$$

$$\text{Var}\hat{g}_2(X) = \frac{2}{\theta^2} \cdot \frac{1}{n}$$

前者要更有效。

2.3.3 大样本性质 *

当 $n \rightarrow \infty$ 时 $\hat{g}(X)$ 的性质。

2.3.3.1 渐近无偏估计

定义 2.3.3

设 $\hat{g}(X)$ 为 $g(\theta)$ 的一个点估计，记 $\hat{g}(X)$ 为 $\hat{g}_n(X)$ ，如果

$$\lim_{n \rightarrow \infty} \mathbb{E}_\theta \hat{g}_n(X) = g(\theta), \forall \theta \in \Theta$$

则称 $\hat{g}_n(X)$ 为渐近无偏估计。

例 2.3.7.

$$a_{nk} = \frac{1}{n} \sum_{i=1}^n X_i^k$$

如果是无偏估计，则一定是渐近无偏估计。

$$m_{nk} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

如 $\mathbb{E}m_{n2} = \frac{n-1}{n}\sigma^2 \rightarrow \sigma^2$, 是渐近无偏估计。 k 一般的情况也可以, 简单说一下:

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k \binom{k}{j} X_i^j \bar{X}^{k-j} (-1)^{k-j} \\ &= \sum_{j=1}^k (-1)^{k-j} \frac{1}{n} \sum_{i=1}^n X_i^j \bar{X}^{k-j}\end{aligned}$$

由 Kormogor 强大数定律,

$$\bar{X} \rightarrow \alpha_1 \text{ a.s.}$$

$$a_{nk} \rightarrow \alpha_k \text{ a.s.}$$

于是

$$\mathbb{E}m_{nk} \rightarrow \sum_{j=1}^k (-1)^{k-j} \alpha_k^j \alpha_1^{k-j} = \mathbb{E}(X - \alpha_1)^k = \mu_k$$

2.3.3.2 相合性

或者叫一致性。

定义 2.3.4

如果 $\hat{g}_n(X)$ 为 $g(\theta)$ 的一个点估计, 若

$$\hat{g}_n(X) \xrightarrow{p} g(\theta), \forall \theta \in \Theta$$

则称 $\hat{g}_n(X)$ 为 $g(\theta)$ 的一个弱相合估计; 若

$$\hat{g}_n(X) \xrightarrow{\text{a.s.}} g(\theta), \forall \theta \in \Theta$$

则称 $\hat{g}_n(X)$ 为 $g(\theta)$ 的一个强相合估计。

2.3.3.3 渐近正态性

定义 2.3.5

如果 $\hat{g}_n(X)$ 为 $g(\theta)$ 的一个弱相合估计, 存在 $A_n(\theta)$ 和 $B_n(\theta) > 0$ 使得

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

则称 $\hat{g}_n(X)$ 为 $g(\theta)$ 的一个相合渐近正态估计。

常见的情况是取

$$B_n(\theta) = \frac{1}{\sqrt{n}} D_n(\theta), a < D_n(\theta) < b$$

2.3.3.4 其他讨论

完全没听懂这节课在讲什么。好像是在讲矩估计和极大似然估计有着怎样的大样本性质，

$$\begin{aligned} a_{nk} - \alpha_k &= \frac{1}{n} \sum_{i=1}^n (X_i^k - \alpha_k) \\ \frac{a_{nk} - \alpha_k}{\sqrt{\frac{\alpha_{2k}}{n}}} &= \frac{1}{\sqrt{\alpha_{2k}n}} \sum_{i=1}^n (X_i^k - \alpha_k) \\ &\Rightarrow \sqrt{n}(a_{nk} - \alpha_k) \xrightarrow{L} N(0, \sigma^2) \end{aligned}$$

如果设 $g(\theta) = G(\alpha_1, \dots, \alpha_k, \mu_2, \dots, \mu_l) = G(\eta)$ ，矩估计 $\hat{g}_n(\theta) = G(a_{n1}, \dots, a_{nk}, m_{n2}, \dots, m_{nl}) = G(\hat{\eta})$ ，那么

$$G(\hat{\eta}) - G(\eta) = \left(\frac{\partial G}{\partial \eta} \right)' (\hat{\eta} - \eta) + o_p(1) := \beta'(\hat{\eta} - \eta)$$

定义矩阵的期望为每个元素的期望构成的矩阵，向量亦然，则 $E(AXB) = A \cdot (EX) \cdot B$ 。设 $X = (X_1, \dots, X_k)'$ ， $\text{Var}(X) = E(X - EX)(X - EX)' = (\text{Cov}(X_i, X_j))_{k \times k}$ ，一般是个非负定的矩阵。则

$$\text{Var}(\beta'X) = \beta' \text{Var}(X) \beta$$

这说明矩估计是？（突然又说）MLE 在一定条件下是相合渐近正态的。

$$\begin{aligned} 0 &= \frac{\partial \ln L(x, \hat{\theta})}{\partial \theta} = \sum_{i=1}^n \frac{\partial \ln L(x_i, \hat{\theta})}{\partial \theta} \\ &= \sum \frac{\partial \ln L(x_i, \theta_0)}{\partial \theta} + \sum \frac{\partial^2 \ln L(x_i, \theta_0)}{\partial \theta^2} (\hat{\theta} - \theta_0) + o_p(\|\hat{\theta} - \theta_0\|) \end{aligned}$$

再用个中心极限定理。

2.4 一致最小方差无偏估计

UMVU 估计，best 估计！

2.4.1 定义

定义 2.4.1 (均方误差)

设 $\hat{g}_n(X)$ 为 $g(\theta)$ 的一个估计，

$$\text{MSE}_\theta(\hat{g}_n(X)) \stackrel{\text{def}}{=} E(\hat{g}_n(X) - g(\theta))^2$$

称为均方误差 (MSE, mean square error)。对于 $g(\theta)$ 的估计 $\hat{g}_1(X)$ 和 $\hat{g}_2(X)$ ，如果

$$E_\theta(\hat{g}_1(X) - g(\theta))^2 \leq E_\theta(\hat{g}_2(X) - g(\theta))^2, \forall \theta \in \Theta$$

且至少对一个 $\theta \in \Theta$ ，严格小于号成立，则称在 MSE 准则下 $\hat{g}_1$ 优于 $\hat{g}_2$ ，若存在 $\hat{g}^*(X)$ 优于任一 $g(\theta)$ 的估计，则称 $\hat{g}^*(X)$ 为一致最小均方误差估计。

考虑

$$\begin{aligned} E(\hat{g}(X) - g(\theta))^2 &= E[(\hat{g}(X) - E\hat{g}(X)) + (E\hat{g}(X) - g(\theta))]^2 \\ &= \text{Var}\hat{g}(X) + (E\hat{g}(X) - g(\theta))^2 + 2E(\hat{g}(X) - E\hat{g}(X))(E\hat{g}(X) - g(\theta)) \\ &= \text{Var}\hat{g}(X) + (E\hat{g}(X) - g(\theta))^2 \end{aligned}$$

记 $\text{Bias}_\theta \hat{g}(X) = (E\hat{g}(X) - g(\theta))^2$ ，称为估计 $\hat{g}(X)$ 的偏差。所有无偏估计的偏差都是 0，那么 MSE 的大小只取决于方差。

定义 2.4.2

若存在无偏 $\hat{g}^*(X)$, 则对于任一无偏 $\hat{g}(X)$, 都有

$$\text{Var}_\theta \hat{g}^*(X) \leq \text{Var}_\theta \hat{g}(X), \forall \theta \in \Theta$$

则称 $\hat{g}^*(X)$ 为 $g(\theta)$ 的一致最小方差无偏估计 (Uniformly minimum variance unbiased estimate), 简记为 UMVUE.

2.4.2 利用条件期望寻找 UMVUE

引理 2.4.1 (Rao-Blackwell Theorem, R-B 引理)

设 X, Y 为两个 r.v. $\varphi(x)$ 为可测函数, $\mathbb{E}Y^2, \mathbb{E}X^2, \mathbb{E}\varphi^2(X) < \infty$, 则当 $\varphi(X) = \mathbb{E}(Y|X)$ 时, $\mathbb{E}(Y - \varphi(X))^2$ 取得最小值。

证明

$$\begin{aligned} \mathbb{E}(Y - \varphi(X))^2 &= \mathbb{E}[Y - \mathbb{E}(Y|X) + \mathbb{E}(Y|X) - \varphi(X)]^2 \\ &= \mathbb{E}(Y - \mathbb{E}(Y|X))^2 + \mathbb{E}(\mathbb{E}(Y|X) - \varphi(X))^2 + 2\mathbb{E}(Y - \mathbb{E}(Y|X))(\mathbb{E}(Y|X) - \varphi(X)) \\ &= \mathbb{E}(Y - \mathbb{E}(Y|X))^2 + \mathbb{E}(\mathbb{E}(Y|X) - \varphi(X))^2 + 2\mathbb{E}_x(Y - \mathbb{E}(Y|X))(\mathbb{E}(Y|X) - \varphi(X)|X=x) \end{aligned}$$

$\varphi(X) = \mathbb{E}(Y|X)$ 时上式取最小值,

本引理的直观理解: 我们知道

$$L = \{\text{r.v. } X, \mathbb{E}X = 0, \mathbb{E}X^2 < \infty\}$$

是一个线性空间, 定义内积:

$$(X, Y) = \mathbb{E}(XY) = \text{Cov}(X, Y)$$

于是两个向量的夹角为:

$$\cos(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \rho_{XY}$$

这是相关系数的几何意义。 $\forall \text{r.v. } X$,

$$L(X) = \{\varphi(X) | \varphi \text{ 可测}, \mathbb{E}\varphi(X) = 0, \mathbb{E}\varphi^2(X) < \infty\}$$

也是一个线性空间, $L(X) \subset L$, 给定 $X \in L$, 定义

$$L_1(X) = \{aX | a \in \mathbb{R}\}$$

则 $L_1(X) \subset L(X) \subset L$, Y 在 X 上的线性投影, 记作 $\hat{\mathbb{E}}(Y|X) = aX$. 其中 $a = \text{Cov}(Y, X)\text{Var}^{-1}(X)$.

现在我们考虑一般情况, 即 $\mathbb{E}X, \mathbb{E}Y \neq 0$, 中心化之后得到

$$\hat{\mathbb{E}}(Y|X) = \mu_Y + \text{Cov}(Y, X)\text{Var}^{-1}(X)(X - \mu_X)$$

可以把 X, Y 推广到 p 维和 q 维向量, Y 在 $L(X)$ 上的投影为 $\mathbb{E}(Y|X) := (\mathbb{E}(Y_1|X), \dots, \mathbb{E}(Y_q|X))'$, Y 在 $L_1(X)$ 上的线性投影为

$$\hat{\mathbb{E}}(Y|X) = \mu_Y + \text{Cov}(Y, X)\text{Var}^{-1}(X)(X - \mu_X)$$

这些内容在回归分析中也有过介绍。

言归正传, 由引理 2.4.1 得到一个直接推论:

引理 2.4.2

设 $T(X)$ 为充分统计量, 且

$$\mathbb{E}_\theta \hat{g}(X) = g(\theta), \theta \in \Theta$$

取 T 的函数 $h(T) = \mathbb{E}(\hat{g}(X)|T)$, 于是

$$\mathbb{E}_{\theta} h(T(X)) = g(\theta), \theta \in \Theta$$

且

$$\text{Var}_{\theta} h(T(X)) \leq \text{Var}(\hat{g}(X))$$



证明 $T(X)$ 为充分统计量, $X|T$ 的分布与 θ 无关, $h(T) = \mathbb{E}(\hat{g}(X)|T)$, 表达式中无 θ , $h(T(X))$ 为统计量,

$$\mathbb{E}_{\theta} h(T(X)) = g(\theta), \theta \in \Theta$$

由引理 2.4.1,

$$\text{Var}_{\theta} h(T) \leq \text{Var}_{\theta} \hat{g}(X), \theta \in \Theta$$

至此, 我们得到两个结论:

1. $\forall g(\theta)$ 的无偏估计 $\hat{g}(X)$, 若 T 为充分统计量, 则 $\mathbb{E}(\hat{g}|T) := h(T)$ 仍为 $g(\theta)$ 的无偏估计。
2. T 为充分统计量, 若 $\hat{g}(X)$ 不是 T 的函数, 则可找到 T 的函数 $h(T)$, 使 $h(T)$ 为 $g(\theta)$ 的无偏估计, 且方差不超过 $\text{Var} \hat{g}(X)$.

故找 UMVUE 只需在 T 的函数中寻找。

例 2.4.1.

但不是所有的 $g(\theta)$ 都存在无偏估计: 设样本 $X \sim B(N, p)$, $n = 1$, $g(p) = \frac{1}{p}$, 则其无偏估计不存在。

证明 (反证) 假设存在无偏估计 $\hat{g}(X)$, X 的取值只有 $N+1$ 个, 即 $\{0, 1, \dots, N\}$, 从而 $\hat{g}(X)$ 的取值最多只有 $N+1$ 个, 记作 $\{a_0, \dots, a_N\}$,

$$\mathbb{E}_p \hat{g}(X) = \sum_{i=0}^N a_i \binom{N}{i} p^i (1-p)^{N-i} = \frac{1}{p}, \forall p \in (0, 1)$$

等价于

$$\sum_{i=0}^N a_i \binom{N}{i} p^{i+1} (1-p)^{N-i} - 1 = 0$$

左边是 p 的 $N+1$ 次非零多项式, 最多只有 $N+1$ 个零点, 无法对于任何 $p \in (0, 1)$ 都成立。

下面我们要在 $g(\theta)$ 存在无偏估计的前提下讨论问题。

定理 2.4.1 (Lehmann-Scheff)

设总体 $\sim f(x, \theta)$, 简单随机样本 $X = (X_1, \dots, X_n)$, $g(\theta)$ 为可估函数^a, $T(X)$ 为充分完全统计量, 若 $\hat{g}(T)$ 为 $g(\theta)$ 的一个无偏估计, 则 $\hat{g}(T)$ 为 $g(\theta)$ 的唯一的 UMVUE。

^a即存在无偏估计。



证明 唯一性: 设 $\mathbb{E}_{\theta} \hat{g}_1(T) = g(\theta), \theta \in \Theta$, $\delta(T) = \hat{g}_1(T) - g(T)$, 由于 T 是完全统计量, $\mathbb{E}_{\theta} \delta(T) = 0$, 进而 $\delta(T) = 0$, 即 $\hat{g}_1(T) = g(T)$ 。

存在性: 任取统计量 $\varphi(X)$, 令 $h(T) = \mathbb{E}(\varphi(X)|T)$, 因为 T 是充分统计量, 所以 $h(T)$ 不含 θ , 为统计量, 由引理 2.4.2 可知

$$\text{Var}_{\theta} h(T) \leq \text{Var}_{\theta} \varphi(X)$$

由唯一性, $h(T)$ 就是 $g(T)$ 。

总结一下, 寻找 UMVUE 的步骤为:

1. 找一个充分完全统计量 T 。
2. 找一个 T 的函数 $g(T)$ 使得 $\mathbb{E}_{\theta} g(T) = g(\theta)$ 。

3. 或者找一个统计量 $\varphi(X)$ 使得 $\mathbb{E}_\theta \varphi(X) = g(\theta)$. (无偏估计)

4. 令 $h(T) = \mathbb{E}(\varphi(X)|T)$, 即为 UMVUE.

步骤 2 和 3 进行其一即可。

推论 2.4.1

样本 $X = (X_1, \dots, X_n)$ 有指数族的自然形式, 即

$$X \sim c^*(\theta) \exp\{\theta' T(x)\} h(x)$$

$\theta \in \Theta^*$, 且 $\Theta^* \subset \mathbb{R}^k$ 有内点, 若

$$\mathbb{E}_\theta h(T) = g(\theta), \theta \in \Theta$$

则 $h(T)$ 为 $g(\theta)$ 的 UMVUE.



例 2.4.2.

证明: 若总体 $\sim B(1, p)$, 样本 $X = (X_1, \dots, X_n)$, 则 $\hat{p} = \bar{X}$ 为 p 的 UMVUE.

证明

$$X \sim p^{\sum x_i} (1-p)^{n-\sum x_i}$$

前面已经证明过, 样本的分布有指数族的自然形式, 且参数空间有内点, $T = \sum X_i$ 为充分完全统计量, 而 $h(T) = T/n$, $\mathbb{E}_p h(T) = p, \forall p \in (0, 1)$, 所以 $h(T) = \bar{X}$ 为 UMVUE.

例 2.4.3.

上例中, 改为 $T = \sum X_i \sim B(n, p)$, 求 $g(p) = p(1-p)$ 的 UMVUE.

证明 因为 T 为充分完全统计量,

$$\varphi(X_1, \dots, X_n) = I_{\{X_1=1\}} I_{\{X_2=0\}}$$

$$\mathbb{E}_p \varphi(X) = P(X_1 = 1)P(X_2 = 0) = p(1-p), \forall p \in (0, 1)$$

于是令

$$\begin{aligned} h(t) &= \mathbb{E}(\varphi(X)|T=t) \\ &= P(X_1=1, X_2=0 | \sum X_i=t) \\ &= \frac{P(X_1=1, X_2=0, \sum X_i=t)}{P(T=t)} \\ &= \frac{P(X_1=1, X_2=0, \sum_{i=3}^n X_i=t-1)}{P(T=t)} \\ &= \frac{p(1-p) \binom{n-2}{t-1} p^{t-1} (1-p)^{n-2-(t-1)}}{\binom{n}{t} p^t (1-p)^{n-t}} \\ &= \frac{t(n-t)}{n(n-1)} \end{aligned}$$

即

$$h(T) = \frac{T(n-T)}{n(n-1)}$$

2.4.3 利用零无偏估计寻找 UMVUE

所谓零无偏估计,是指对 0 的无偏估计,也就是期望为 0 的统计量¹。

定理 2.4.2

设 $\mathbb{E}_\theta \hat{g}(X) = g(\theta), \theta \in \Theta$, 且 $\text{Var}_\theta \hat{g}(X) < \infty, \theta \in \Theta$, 若对于任一满足期望为 0 的统计量 $l(X)$:

$$\mathbb{E}_\theta l(X) = 0, \forall \theta \in \Theta$$

都有 $\text{Cov}_\theta(\hat{g}(X), l(X)) = 0, \forall \theta \in \Theta$, 则 $\hat{g}(X)$ 的 $g(X)$ 为 UMVUE.

推论 2.4.2

设 $T = T(X)$ 是 θ 的充分统计量, $h(T)$ 是 $g(\theta)$ 的一个无偏估计, 且 $\text{var}_\theta(h(T)) < \infty, \forall \theta \in \Theta$, 则 $h(T)$ 是 $g(\theta)$ 的 UMVUE 当且仅当任一零无偏估计 $U(T)$, 也就是满足期望为零的 T 的函数 $U(T)$, 都有 $\text{Cov}_\theta(h(T), U(T)) = 0, \forall \theta \in \Theta$.

例 2.4.4. 相似习题: Ex3.40

总体 $\sim \exp\{\lambda\}$, 求 $1/\lambda$ 的 UMVUE.

解答:

$$(X_1, \dots, X_n) \sim \lambda^n \exp\{-\lambda \sum_{i=1}^n x_i\} \prod_{i=1}^n I_{(0, \infty)}(x_i)$$

$T = \sum X_i$ 为充分统计量,

$$T \sim \Gamma(n, \lambda) \sim f(t, \lambda) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} I_{(0, \infty)}(t)$$

于是 $\mathbb{E}T = n/\lambda \Rightarrow 1/\lambda = \mathbb{E}(T/n)$, T 为 $1/\lambda$ 的无偏估计, 设

$$\begin{aligned} \mathbb{E}_\lambda \delta(T) &= \int_0^\infty \delta(t) \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} dt = 0 \\ \Rightarrow g(\lambda) &= \int_0^\infty \delta(t) t^{n-1} e^{-\lambda t} dt = 0, \lambda > 0 \\ \Rightarrow -\frac{dg(\lambda)}{\lambda} &= \int_0^\infty \delta(t) t^n e^{-\lambda t} dt = 0, \lambda > 0 \\ \Rightarrow \text{Cov}(\delta(T), T) &= 0, \lambda > 0 \end{aligned}$$

由定理 2.4.2 可得 T/n 为 $1/\lambda$ 的 UMVUE.

例 2.4.5. 相似习题: Ex3.30

$X = (X_1, \dots, X_n)$, $X_i \sim N(a, \sigma^2)$, 用零无偏估计法求 $\theta = (\theta_1, \theta_2) = (a, \sigma^2)$ 的 UMVUE.

解答: 已经证明过

$$T = (T_1, T_2) = (\bar{X}, \sum (X_i - \bar{X})^2)$$

是充分统计量,

$$T_1 \sim N(a, \frac{\sigma^2}{n}), \sigma^{-2} T_2 \sim \chi_{n-1}^2$$

¹这个“对零进行估计”莫名戳我笑点。

且相互独立,

$$\begin{aligned} T \sim f_{\theta}(t_1, t_2) &= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{n}{2\sigma^2}(t_1 - a)^2\right\} \frac{\sigma^{n-1}}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} t_2^{\frac{n-1}{2}-1} \exp\left\{-\frac{\sigma^2}{2}t_2\right\} I_{(0,\infty)}(t_2) \\ &= C(n, \sigma) t_2^{\frac{n-1}{2}-1} \exp\left\{-\frac{n}{2\sigma^2}(t_1 - a)^2\right\} \exp\left\{-\frac{\sigma^2}{2}t_2\right\} I_{(0,\infty)}(t_2) \end{aligned}$$

设 $\delta(T) = \delta(T_1, T_2)$ 为任一零无偏估计,

$$\mathbb{E}_{\theta}\delta(T) = \iint \delta(t_1, t_2) f_{\theta}(t_1, t_2) dt_1 dt_2 = 0, a \in \mathbb{R}, \sigma^2 > 0 \quad (2.4.1)$$

式 (2.4.1) 两边对 σ 求导, 具有以下形式:

$$\begin{aligned} \frac{\partial}{\partial \sigma} \mathbb{E}_{\theta}\delta(T) &= \frac{\partial C}{\partial \sigma} \frac{1}{C} \mathbb{E}_{\theta}\delta(T) + C_1(n, \theta) \iint \delta(t_1, t_2) (n(t_1 - a)^2 + \sigma^4 t_2) f_{\theta}(t_1, t_2) dt_1 dt_2 \\ &= C_1(n, \theta) \iint \delta(t_1, t_2) (n(t_1 - a)^2 + \sigma^4 t_2) f_{\theta}(t_1, t_2) dt_1 dt_2 = 0 \end{aligned}$$

即

$$\text{Cov}_{\theta}(\delta(T), n(T_1 - a)^2 + \sigma^4 T_2) = 0 \quad (2.4.2)$$

式 (2.4.1) 两边对 a 求两次导, 具有以下形式:

$$\begin{aligned} \frac{\partial}{\partial a} \mathbb{E}_{\theta}\delta(T) &= C_2(n, \sigma) \iint \delta(t_1, t_2) (t_1 - a) f_{\theta}(t_1, t_2) dt_1 dt_2 = 0 \\ \frac{\partial^2}{\partial a^2} \mathbb{E}_{\theta}\delta(T) &= C_2(n, \sigma) \iint \delta(t_1, t_2) (C_2(n, \sigma) (t_1 - a)^2 - 1) f_{\theta}(t_1, t_2) dt_1 dt_2 \\ &= C_2(n, \sigma)^2 \iint \delta(t_1, t_2) (t_1 - a)^2 f_{\theta}(t_1, t_2) dt_1 dt_2 \end{aligned}$$

即

$$\text{Cov}_{\theta}(\delta(T), T_1 - a) = 0 \quad (2.4.3)$$

$$\text{Cov}_{\theta}(\delta(T), (T_1 - a)^2) = 0 \quad (2.4.4)$$

由 (2.4.3) 可知 $\text{Cov}_{\theta}(\delta(T), T_1) = 0$, 又因为 $T_1 = \bar{X}$ 是 a 的无偏估计, 所以 $\bar{X}$ 为 a 的 UMVUE; 结合 (2.4.2) 和 (2.4.4) 可知 $\text{Cov}_{\theta}(\delta(T), \frac{1}{n-1}T_2) = 0$, 又因为 $\frac{1}{n-1}T_2 = S^2$ 是 σ^2 的无偏估计, 所以 $\frac{1}{n-1}T_2 = S^2$ 为 σ^2 的 UMVUE.

例 2.4.6.

总体 $\sim P(\lambda)$, 求:

1. $g_1(\lambda) = \lambda$
2. $g_2(\lambda) = \lambda^r, r \in \mathbb{N}$
3. $g_3(\lambda) = P_{\lambda}(X = x)$

的 UMVUE.

解答:

$$X = (X_1, \dots, X_n) \sim e^{-n\lambda} \lambda^{\sum x_i} \prod_{i=1}^n (x_i!)^{-1}$$

所以 $T = \sum X_i$ 为充分完全统计量,

1. $\hat{g}_1(X) = T/n = \bar{X}$, $\mathbb{E}_{\lambda}\bar{X} = \lambda$, 故 λ 的 UMVUE 为 $\bar{X}$.
2. $r.v.X$ 的生成函数为

$$h(t) = \mathbb{E}t^X = \sum_{k=0}^{\infty} t^k p_k$$

于是

$$\begin{aligned} h'(t)|_{t=1} &= \mathbb{E}X \\ h''(t)|_{t=1} &= \mathbb{E}X(X-1) \\ &\vdots \\ h^{(n)}(t)|_{t=1} &= \mathbb{E}X(X-1)\cdots(X-r+1) \end{aligned}$$

对于 Poisson 分布而言, $h(t) = e^{\lambda(t-1)}$, $h^{(r)}(1) = \lambda^r$, 因为 $T = \sum X_i \sim P(n\lambda)$, 于是

$$\mathbb{E}T(T-1)\cdots(T-r+1) = (n\lambda)^r$$

于是 $n^{-r}T(T-1)\cdots(T-r+1)$ 是 λ^r 的无偏估计, 由 L-S 定理可知其为 UMVUE.

3. $\hat{g}_3(x) = I_{\{X_1=x\}}$,

$$\mathbb{E}\hat{g}_3(X) = P_\lambda(X_1 = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

令

$$\begin{aligned} h(T) &= \mathbb{E}(\hat{g}_3(X)|T=t) = P(X_1=x|\sum X_i=t) \\ &= \frac{P(X_1=x, \sum X_i=t)}{\sum X_i=t} I_{\{x \leq t\}} \\ &= \frac{P(X_1=x, \sum_{i=2}^n X_i=t-x)}{P(\sum X_i=t)} \\ &= e^{-\lambda} \frac{\lambda^x}{x!} \frac{e^{-(n-1)\lambda} ((n-1)\lambda)^{t-x}}{(t-x)!} \frac{t!}{e^{-n\lambda} (n\lambda)^t} \\ &= \binom{t}{x} \left(\frac{n-1}{n}\right)^{t-x} \left(\frac{1}{n}\right)^x \end{aligned}$$

所以

$$\hat{g}_3(X) = \binom{T}{x} \left(\frac{n-1}{n}\right)^{T-x} \left(\frac{1}{n}\right)^x$$

为 $P_\lambda(X=x)$ 的 UMVUE.

例 2.4.7. 相似习题: Ex3.36

总体 $\sim \exp(\lambda)$, 样本 $X = (X_1, \dots, X_n) \sim f(x_1, \dots, x_n)$,

$$f(x_1, \dots, x_n) = \lambda^n e^{-\lambda \sum x_i} \prod_{i=1}^n I_{(0, \infty)}(x_i)$$

$T = \sum X_i$ 为充分完全统计量, 求 $g_1(\lambda) = \lambda$ 和 $g_2(\lambda) = 1 - e^{-\lambda x_0}$ 的 UMVUE.

解答: 我们已经说明过, $\hat{g}_1(T) = (n-1)/T$ 为 λ 的无偏估计, 由 L-S 定理知是 λ 的 UMVUE.

然后设

$$\hat{g}_3(X) = P(X_1 \leq x_0) = 1 - e^{-\lambda x_0}$$

$$\begin{aligned} h(T) &= \mathbb{E}(\hat{g}_3(X)|T=t) = P(X_1 \leq x_0|T=t) \\ &= \int_0^{x_0} f_{X_1|T}(x|t) dx \\ &= \int_0^{x_0} \frac{f(x_1, \sum x_i=t)}{f_2(\sum X_i=t)} dx \\ &= \int_0^{x_0} \frac{f(X_1=x_1, \sum_{i=2}^n X_i=t-x_1)}{f(\sum X_i=t)} dx \end{aligned}$$

$\sum X_i \sim \Gamma(x, \lambda, n)$,

$$\sum_{i=2}^n X_i \sim \Gamma(y, \lambda, n-1) = \frac{f_1(x_1)f_1(t-x, \lambda, n-1)}{f_2(t, \lambda, n)} = \frac{\lambda e^{-\lambda x_1} \lambda^{n-1} (t-x_1)^{n-2} e^{-\lambda(t-x_1)} \cdot (n-1)!}{(n-1)! \lambda^n t^{n-1} e^{-\lambda t}} I_{x_1 \leq t}$$

所以

$$h(t) = \int_0^{x_0} \frac{(n-1)(t-x_1)^{n-2}}{t^{n-1}} dx_1 I_{\{x_0 \leq t\}} = [1 - (1 - \frac{x_0}{t})^{n-1}] I_{\{x_0 \leq t\}}$$

所以 $h(T) = [1 - (1 - x_0/T)^{n-1}] I_{\{x_0 \leq T\}}$ 为 $g_2(\lambda)$ 的 UMVUE.

这段写得也太混乱了, 抽空参考作业里的 Ex3.36 完善一下。

例 2.4.8.

总体 $\sim N(a, \sigma^2)$, $\theta = (\theta_1, \theta_2) = (a, \sigma^2)$, 求

1. a, σ^2

2. $g_1(\theta) = \sigma^r = \theta_2^{\frac{r}{2}}, r > 0$

3. $g_2(\theta) = \frac{a}{\sigma^2} = \frac{\theta_1}{\theta_2}$

的 UMVUE.

解答: 样本 $X = (X_1, \dots, X_n) \sim f(x_1, \dots, x_n)$,

$$T = (T_1, T_2) = (\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)$$

是充分完全统计量。

1. $\mathbb{E}_\theta \bar{X} = \mathbb{E}_\theta T_1 = \theta_1$, $\mathbb{E}_\theta \frac{1}{n-1} T_2 = \mathbb{E}_\theta S^2 = \sigma^2 = \theta_2$, 故 $\bar{X}$ 和 S^2 为 θ_1 和 θ_2 的 UMVUE.

2. 注意到 $\sum (X_i - \bar{X})^2 \sim \sigma^2 \chi_{n-1}^2$, 记 $Y = T_2/\theta_2 \sim \chi_{n-1}^2 \sim \Gamma(x, n/2, 1/2)$,

$$\Gamma(x, \frac{n}{2}, \frac{1}{2}) = \frac{x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} I_{(0, \infty)}(x)$$

于是

$$\begin{aligned} \mathbb{E}(\frac{T_2}{\sigma^2})^r &= \int_0^\infty y^{\frac{r}{2}} \frac{y^{\frac{n-1}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} dy \\ &\stackrel{t=y/2}{=} \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} \int_0^\infty 2^{\frac{n+r-3}{2}} t^{\frac{n+r-1}{2}-1} e^{-t} dt \\ &= \frac{2^{\frac{r}{2}} \Gamma(\frac{n+r-1}{2})}{\Gamma(\frac{n-1}{2})} := \frac{1}{K_{n,r}} \end{aligned}$$

于是

$$\mathbb{E}_\theta (T_2^r \cdot K_{n,r}) = \sigma^r$$

所以 $K_{n,r} T_2^r$ 为 σ^r 的 UMVUE, 特别地取 $r=1$,

$$\hat{\sigma}_{\text{UMVUE}} = \sqrt{\frac{n-1}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} S$$

取 $n=9$, $\hat{\sigma} \approx 1.032S$.

3. a 用 $\bar{X}$, $1/\sigma^2$ 用 $K_{n,-2} T_2^{-2}$,

$$\mathbb{E} \hat{g}_2(T) = \mathbb{E} T_1 \cdot K_{n,-2} T_2^{-2} = \mathbb{E} T_1 \cdot \mathbb{E} K_{n,-2} T_2^{-2} = \frac{a}{\sigma^2}$$

为 UMVUE.

例 2.4.9.

总体 $X \sim N(\theta, 1)$, 求 $P_\theta(X > 1)$ 的 UMVUE.

解答：令 $\hat{g}(X) = I_{\{X_1 > 1\}}$,

$$\mathbb{E}\hat{g}(X) = P(X_1 > 1) = P\left(\frac{X_1 - \theta}{1} > \frac{1 - \theta}{1}\right) = \Phi(\theta - 1)$$

$T = \sum X_i$ 为充分完全统计量, 则

$$\mathbb{E}(\hat{g}(X)|T=t) = P(X_1 > 1 | \sum X_i = t) = \int_1^\infty f_{X_1|T}(x_1|t)dx_1$$

直接按条件期望定义展开计算很麻烦, 我们用之前的结论:

$$\mathbb{E}(X_1|T=t) = \mu_1 + \text{Cov}(X_1, \sum X_i) \text{Var}^{-1}(\sum X_i)(t - n\theta) = \frac{t}{n}$$

方差:

$$\text{Var}(X_1) - \text{Cov}(X_1, \sum X_i) \text{Var}^{-1}(\sum X_i) \text{Cov}(X_1, \sum X_i) = 1 - \frac{1}{n}$$

2.5 Cramer-Rao 不等式

对参数 $g(\theta)$ 有一个估计 $\hat{g}(X)$, 考虑其有效能到什么程度, 也就是方差能有多小, 最小总不能是零吧, 那下界是哪里呢? Cramer-Rao 不等式给出了在满足 C-R 正则条件下方差的下界。达到方差下界的估计称为有效估计, 无偏的有效估计就是 UMVUE, 这也是一种寻找 (判断?) UMVUE 的办法。但是反过来, UMVUE 不一定就是达到方差下界的估计。

2.5.1 单参数 C-R 不等式

回想似然函数: 假设求偏导和求期望 (积分) 可交换,

$$\begin{aligned} S_1(X_i, \theta) &= \ln f(X_i, \theta) \\ \frac{\partial}{\partial \theta} \mathbb{E} S_1(X_i, \theta) &= \frac{\partial}{\partial \theta} \mathbb{E}(\ln f(X_i, \theta)) \\ &= \mathbb{E} \frac{\partial}{\partial \theta} \ln f(X_i, \theta) \\ &= \int \frac{\partial f(x_i, \theta)}{\partial \theta} dx_i \\ &= \frac{\partial}{\partial \theta} \int f(x_i, \theta) dx_i = 0 \end{aligned}$$

令 $S(X, \theta) = \frac{\partial \ln f(X, \theta)}{\partial \theta}$, 则

$$\mathbb{E} S(X, \theta) = \mathbb{E} \sum_{i=1}^n \frac{\partial \ln f(X_i, \theta)}{\partial \theta} = 0$$

设 $\mathbb{E}_\theta \hat{g}(X) = g(\theta)$,

$$\begin{aligned} \text{Cov}(\hat{g}(X), S(X, \theta)) &= \mathbb{E} \hat{g}(X) S(X, \theta) = \mathbb{E} \hat{g}(X) \frac{\partial \ln f(X, \theta)}{\partial \theta} \\ &= \mathbb{E} \hat{g}(X) \frac{1}{f(X, \theta)} \frac{\partial f(X, \theta)}{\partial \theta} \\ &= \int \hat{g}(X) \frac{\partial f(X, \theta)}{\partial \theta} dx \\ &= \frac{\partial}{\partial \theta} \int \hat{g}(X) f(x, \theta) dx = g'(\theta) \end{aligned}$$

再看 $S(X, \theta)$ 的方差:

$$\begin{aligned}\text{Var}S(X, \theta) &= \mathbb{E}S^2(X, \theta) \\ &= \mathbb{E}\left(\sum_{i=1}^n \frac{\partial \ln f(X_i, \theta)}{\partial \theta}\right)^2 \\ &= n\mathbb{E}\left(\frac{\partial \ln f(X_1, \theta)}{\partial \theta}\right)^2 := nI(\theta)\end{aligned}$$

$I(\theta)$ 被称为总体信息函数, $nI(\theta)$ 为样本信息函数。

定义 2.5.1 (单参数 C-R 正则条件)

若单参数概率函数族 $\mathcal{F} = \{f(x, \theta), \theta \in \Theta\}$ 满足 C-R 正则条件:

1. Θ 是 $\mathbb{R}$ 上的开区间;
2. $\forall \theta \in \Theta, \{f(x, \theta) > 0\}$ 相同, 即分布族具有共同的支撑;
3. $\forall x \in X, \theta \in \Theta, \frac{\partial f(x, \theta)}{\partial \theta}$ 存在。
4. $f(x, \theta)$ 的积分与微分运算可交换, 即

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

离散情况下, 上述条件改为无穷级数和微分运算可交换;

5. $I(\theta)$ 存在且非负:

$$0 < I(\theta) = \mathbb{E}_{\theta} \left(\frac{\partial \ln f(X, \theta)}{\partial \theta} \right)^2 < \infty$$

则该分布族称为 C-R 正则分布族。

定理 2.5.1

设 $\mathcal{F} = \{f(x, \theta) | \theta \in \Theta\}$ 是 C-R 正则分布族, $g(\theta)$ 可微, 样本 $X = (X_1, \dots, X_n)$, $\hat{g}(X)$ 是对 $g(\theta)$ 的无偏估计, 则

$$\text{Var}_{\theta} \hat{g}(X) \geq \frac{g'(\theta)^2}{nI(\theta)}$$

其中

$$I(\theta) = \mathbb{E} \left(\frac{\partial \ln f(X_1, \theta)}{\partial \theta} \right)^2$$

证明 记

$$S(X, \theta) = \frac{\partial \ln f(X, \theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \ln f(X_i, \theta)}{\partial \theta}$$

已经推导过:

$$\mathbb{E}_{\theta} S(X, \theta) = 0$$

于是

$$\begin{aligned}\text{Cov}_{\theta}(\hat{g}, S(X, \theta)) &= \mathbb{E} \hat{g}(X) S(X, \theta) \\ &= \int \hat{g}(x) \frac{\partial f(x, \theta)}{\partial \theta} \frac{1}{f(x, \theta)} f(x, \theta) dx \\ &= \frac{\partial}{\partial \theta} \int \hat{g}(x) f(x, \theta) dx \\ &= \frac{\partial}{\partial \theta} \mathbb{E} \hat{g}(\theta) = \frac{\partial}{\partial \theta} g(\theta) = g'(\theta)\end{aligned}$$

$$\begin{aligned}\text{Var}_\theta S(X, \theta) &= \text{Var}_\theta \left(\sum_{i=1}^n \frac{\partial \ln f(X_i, \theta)}{\partial \theta} \right) \\ &= n \text{Var}_\theta \left(\frac{\partial \ln f(X, \theta)}{\partial \theta} \right) = \mathbb{E} \left(\frac{\partial \ln f(X, \theta)}{\partial \theta} \right)^2 = -\mathbb{E} \left(\frac{\partial^2 \ln f(X, \theta)}{\partial \theta^2} \right)\end{aligned}$$

最后一个等号是分部积分。我们已经讲过：

$$\mathbb{E}(\hat{g}(X)|S(X, \theta)) = g(\theta) + \text{Cov}(\hat{g}(X), S(X, \theta)) \text{Var}^{-1}(S(X, \theta)) S(X, \theta)$$

$$Y = AX \Rightarrow \text{Var}(Y) = \mathbb{E}(AX - A\mathbb{E}X)(AX - A\mathbb{E}X)' = A \text{Var}(X) A',$$

$$\begin{aligned}\text{Var}_\theta(\mathbb{E}(\hat{g}|S)) &= \text{Cov}(\hat{g}, S) \text{Var}^{-1}(S) \text{Var}(S) \text{Var}^{-1}(S) \text{Cov}(S, \hat{g}) \\ &= \text{Cov}(\hat{g}, S) \text{Var}^{-1}(S) \text{Cov}(S, \hat{g}) = \frac{g'(\theta)^2}{nI(\theta)}\end{aligned}$$

由投影知

$$\text{Var}_\theta \hat{g}(X) \geq \text{Var}_\theta(\mathbb{E}(\hat{g}|S)) = \frac{g'(\theta)^2}{nI(\theta)}$$

例 2.5.1.

总体 $\sim B(1, p)$ ，用 C-R 下界证明 $\bar{X}$ 为 p 的 UMVUE.

证明

$$X = (X_1, \dots, X_n) \sim p^{n\bar{X}}(1-p)^{n-n\bar{X}}$$

指数族，满足定理条件。总体 $\sim f(x, p) = p^x(1-p)^{1-x}, x = 0, 1$

$$\begin{aligned}\ln f(x, p) &= x \ln p + (1-x) \ln(1-p) \\ \frac{\partial}{\partial p} \ln f(x, p) &= \frac{x}{p} - \frac{1-x}{1-p} = \frac{x-p}{p(1-p)} \\ \mathbb{E} \left(\frac{\partial \ln f(X, p)}{\partial p} \right)^2 &= \mathbb{E} \left(\frac{X-p}{p(1-p)} \right)^2 = \frac{1}{p(1-p)}\end{aligned}$$

$g(p) = p$, $g'(p) = 1$, C-R 下界为 $p(1-p)/n$, 而 $\text{Var} \bar{X} = p(1-p)/n$, 达到此下界, 故 $\bar{X}$ 为 p 的 UMVUE.

2.5.2 多参数 C-R 不等式 *

$\theta = (\theta_1, \dots, \theta_k)'$, 总体 $\sim f(x, \theta_1, \dots, \theta_k) = f(x, \theta)$, 样本 $X = (X_1, \dots, X_n) \sim \prod f(x_i, \theta)$, 要估计 $g(\theta) = g(\theta_1, \dots, \theta_k)$, 定义

$$S(X, \theta) = \frac{\partial \ln f(X, \theta)}{\partial \theta} = \left(\frac{\partial \ln f(X, \theta)}{\partial \theta_1}, \dots, \frac{\partial \ln f(X, \theta)}{\partial \theta_k} \right)'$$

同样可证 $\mathbb{E}_\theta S(X, \theta) = 0$, $\text{Cov}_\theta(\hat{g}(X), S(X, \theta)) = \left(\frac{\partial g(\theta)}{\partial \theta} \right)'$, $\text{Cov}_\theta(S(X, \theta), \hat{g}(X)) = \frac{\partial g(\theta)}{\partial \theta}$,

$$\text{Var}_\theta(S(X, \theta)) = \mathbb{E}[S(X, \theta) \cdot S(X, \theta)'] = n \mathbb{E} \left(\frac{\partial \ln f(X, \theta)}{\partial \theta_i} \frac{\partial \ln f(X, \theta)}{\partial \theta_j} \right)_{k \times k}$$

可以证明,

$$\mathbb{E} \left(\frac{\partial \ln f(X, \theta)}{\partial \theta_i} \frac{\partial \ln f(X, \theta)}{\partial \theta_j} \right)_{k \times k} = -\mathbb{E} \left(\frac{\partial^2 \ln f(X, \theta)}{\partial \theta_i \partial \theta_j} \right)_{k \times k}$$

则

$$\text{Var}_\theta \hat{g}(X) \geq \left(\frac{\partial g}{\partial \theta} \right)' (nI(\theta))^{-1} \frac{\partial g}{\partial \theta}$$

效率:

$$e_{\hat{g}(\theta)} = \frac{\left(\frac{\partial g}{\partial \theta}\right)' (nI(\theta))^{-1} \frac{\partial g}{\partial \theta}}{\text{Var}_{\theta} \hat{g}(X)} \in (0, 1)$$

当 $n \rightarrow \infty$, 若 $e_{\hat{g}} \rightarrow 1$, 称 $\hat{g}(X)$ 为 $g(\theta)$ 的渐近有效估计。

例 2.5.2.

总体 $\sim N(a, \sigma^2)$, 求 (a, σ^2) 估计的 C-R 方差下界。

解答: $\theta = (\theta_1, \theta_2) = (a, \sigma^2)$, $\hat{a} = \bar{X}$, $\hat{\sigma}^2 = S^2$, 都是 UMVUE,

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}} \sigma^{-1} \exp\left\{-\frac{1}{2\sigma^2}(x-a)^2\right\} = (2\pi)^{-\frac{1}{2}} \theta_2^{-\frac{1}{2}}$$

$$\ln f(x, \theta) = C - \frac{1}{2} \ln \theta_2 - \frac{1}{2\theta_2} (x - \theta_1)^2$$

$$\frac{\partial \ln f(x, \theta)}{\partial \theta} = \left(\frac{\partial \ln f(x, \theta)}{\partial \theta_1}, \frac{\partial \ln f(x, \theta)}{\partial \theta_2} \right) = \left(\frac{x - \theta_1}{\theta_2}, -\frac{1}{2\theta_2} + \frac{(x - \theta_1)^2}{2\theta_2^2} \right)$$

求二阶导

$$\frac{\partial^2 \ln f(x, \theta)}{\partial \theta_1^2} = -\frac{1}{\theta_2}, \quad \frac{\partial^2 \ln f(x, \theta)}{\partial \theta_1 \partial \theta_2} = \frac{\theta_1 - x}{\theta_2^2}, \quad \frac{\partial^2 \ln f(x, \theta)}{\partial \theta_2^2} = \frac{1}{2\theta_2^2} - \frac{(x - \theta_1)^2}{\theta_2^3}$$

上式中每个 x 换成 X , 再求期望, 得到信息矩阵:

$$I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

对于 $\hat{g}_2(X) = S^2$, 其 C-R 下界为

$$\frac{1}{n} \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{2\sigma^4}{n}$$

而 $\text{Var} S^2 = 2\sigma^4/(n-1)$, 虽然是 UMVUE, 仍没有达到 C-R 下界。

对于 $\hat{g}_1(X) = \bar{X}$, 其 C-R 下界为

$$\frac{1}{n} \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\sigma^2}{n}$$

$\text{Var} \bar{X} = \sigma^2/n$, 达到了下界。

例 2.5.3.

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

$$X = (X_1, \dots, X_n) \sim e^{-\sum (x_i - a)} \prod I_{(a, \infty)}(x_i) = e^{na} e^{-\sum x_i} I_{(a, \infty)}(x_{(1)})$$

由因子分解定理, $X_{(1)}$ 为充分统计量,

$$X_{(1)} \sim f(t, a) = ne^{-n(t-a)} I_{(a, \infty)}(t)$$

$$\mathbb{E} X_{(1)} = a + \frac{1}{n}$$

由 L-S 定理, $X_{(1)} - \frac{1}{n}$ 为 a 的 UMVUE, 但不可以讨论其是否为有效估计, 因为不满足正则条件。

$$\text{Var}_a \left(X_{(1)} - \frac{1}{n} \right) = \frac{2a}{n} + \frac{1}{n^2}$$

第 3 章 区间估计

3.1 基本概念

定义 3.1.1

用样本构造一个统计量 $\hat{g}(X)$, 来估计 $g(\theta)$, 这叫做点估计。用样本构造一个区间, 不妨设 $g(\theta) = \theta$, 取 $(\hat{\theta}_1(X), \hat{\theta}_2(X))$, 把 θ 估计在这个区间中, 这被称为区间估计。

$$\begin{array}{ll} E(\hat{\theta}_2 - \hat{\theta}_1) & \text{精度} \\ P_\theta(\theta \in (\hat{\theta}_1, \hat{\theta}_2)) & \text{置信水平} \\ \inf_{\theta \in \Theta} P_\theta(\theta \in (\hat{\theta}_1, \hat{\theta}_2)) & \text{置信系数} \end{array}$$

θ 为未知常数, $\hat{\theta}_1, \hat{\theta}_2$ 为 r.v.

Neyman: 保证置信水平下使得精度最高。这是我们最常采用的观点。

定义 3.1.2

给定一个很小的 α , 若 $P_\theta(\theta \in [\hat{\theta}_1, \hat{\theta}_2]) \geq 1 - \alpha$, 称 $[\hat{\theta}_1, \hat{\theta}_2]$ 为置信水平为 $1 - \alpha$ 的置信区间, 简称置信区间。

常常取 $\alpha = 0.05$, 也就是 95% 的置信水平。

定义 3.1.3

如果 $P_\theta(\theta \leq \hat{\theta}_U(X)) \geq 1 - \alpha, \forall \theta \in \Theta$, 则 $\hat{\theta}_U(X)$ 称为 θ 的置信系数为 $1 - \alpha$ 的置信上限, 简称置信上限。同样地, $P_\theta(\theta \geq \hat{\theta}_L(X)) \geq 1 - \alpha, \forall \theta \in \Theta$, 则 $\hat{\theta}_L(X)$ 称为置信下限。

定义 3.1.4

当 $\theta = (\theta_1, \dots, \theta_k)' \in \Theta \subset \mathbb{R}^k$, 若统计量 $S(X)$ 满足

1. $\forall x \in \mathbb{R}^n, S(X) \subset \Theta$.
2. 给定 $\alpha \in (0, 1)$, $P_\theta(\theta \in S(X)) \geq 1 - \alpha, \theta \in \Theta$.

则称 $S(X)$ 为 θ 的置信水平为 $1 - \alpha$ 的置信域。

如果满足

$$\inf_{\theta \in \Theta} P_{\theta}(\theta \in S(X)) \geq 1 - \alpha$$

则 $S(X)$ 称为置信系数为 $1 - \alpha$ 的置信域。

如何构造置信区间？

1. Neyman 置信区间：通过枢轴变量（频率派）；
2. Fisher 信仰推断：容忍区间和容忍限；
3. Bayes 方法。

我们本章只介绍第一种方法。

3.2 正态总体参数的置信区间

在点估计的基础上，构造一个置信区间。

定义 3.2.1

$S(T, \theta)$ 满足

1. 包含一个优良点估计 T ，同时是 θ 的函数。
2. $S(T, \theta)$ 分布与 θ 无关。
3. 对任意的常数 a, b ，不等式 $a \leq S(T, \theta) \leq b$ 可以改写为 $A \leq \theta \leq B$ ，其中 A, B 与 θ 无关，是统计量。则称之为枢轴变量。

如何作区间估计？先找 a, b ，使得

$$P(a \leq S(T, \theta) \leq b) \geq 1 - \alpha$$

转化为

$$P(A \leq \theta \leq B) \geq 1 - \alpha$$

那么 $[A, B]$ 就是 θ 的置信水平为 $1 - \alpha$ 的置信区间。

3.2.1 单个正态总体参数的置信区间

总体 $\sim N(\mu, \sigma^2)$,

μ 的置信区间：

$$\mu \in \bar{X} \pm d := [\bar{X} - d, \bar{X} + d]$$

1. σ^2 已知，要作 μ 的区间估计。

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

那么

$$a \leq \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq b \Leftrightarrow \mu \in [\bar{X} - \frac{\sigma}{\sqrt{n}}b, \bar{X} - \frac{\sigma}{\sqrt{n}}a]$$

2. σ^2 未知，利用

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$$

其中 d 只有三种形式:

$$d = \begin{cases} \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}} & , \sigma^2 \text{ 已知} \\ \frac{S}{\sqrt{n}} t_{n-1}(\frac{\alpha}{2}) & , \sigma^2 \text{ 未知} \\ \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}} & , \sigma^2 \text{ 未知}, n > 30 \end{cases}$$

因为 t 分布大样本下和正态分布相差不多。我们的原则是:在保证置信水平的前提下精度越高越好,而精度越高置信水平越低,所以我们直接令置信水平等于 $1 - \alpha$, 然后计算就好。

比较重要的两个数据: $u_{0.05} = 1.645$, $u_{0.025} = 1.96$.

σ^2 的置信区间:

1. 如果 μ 已知, 利用

$$S_1^2 := \frac{1}{n} \sum (X_i - \mu)^2, \frac{nS_1^2}{\sigma^2} \sim \chi_n^2$$

那么

$$P(a \leq \frac{nS_1^2}{\sigma^2} \leq b) = 1 - \alpha$$

常用的做法是:取 a 使得 $P(\frac{nS_1^2}{\sigma^2} \leq a) = \frac{\alpha}{2}$, 取 b 使得 $P(\frac{nS_1^2}{\sigma^2} \geq b) = \frac{\alpha}{2}$, 这么做虽然不能保证一定精度最高,但是很简便。最后转化为

$$P(\frac{nS_1^2}{\chi_n^2(\frac{\alpha}{2})} \leq \sigma^2 \leq \frac{nS_1^2}{\chi_n^2(1 - \frac{\alpha}{2})}) \leq 1 - \alpha$$

2. 如果 μ 未知, 利用

$$S^2 := \frac{1}{n-1} \sum (X_i - \bar{X})^2, \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

其余步骤相同。

一般估计出来的范围形如 $\sigma^2 \in [\alpha S^2, \beta S^2], \alpha < 1 < \beta$.

3.2.2 两个正态总体参数的置信区间

$G_1 \sim N(\mu_1, \sigma_1^2)$, 样本 $(X_1, \dots, X_n)$; $G_2 \sim N(\mu_2, \sigma_2^2)$, 样本 $(Y_1, \dots, Y_n)$.

第 4 章 参数假设检验

总体 $F(x, \theta)$, 类型已知, 某些参数未知, 根据样本对参数作出假设, 并进行检验, 称为参数假设检验 (Parametric Hypothesis Testing); 类型未知, 对总体作有关的检验, 称为非参数假设检验 (Nonparametric Hypothesis Testing)。本章主要介绍参数假设检验。

4.1 基本概念

4.1.1 问题的提出

例 4.1.1.

食盐包装线, 规定每袋 350 克, 抽取 10 袋, 平均重量 348 克, 包装线工作是否正常?

解答: 给出假设:

$$H_0: a = 350 \longleftrightarrow H_1: a \neq 350$$

或者

$$H'_0: a = 350 \longleftrightarrow H'_1: a > 350$$

$$H''_0: a = 350 \longleftrightarrow H''_1: a < 350$$

一般, 设 $\Theta = \Theta_1 + \Theta_0$,

$$H_0: \theta \in \Theta \longleftrightarrow H_1: \theta \in \Theta_1$$

如果 Θ_0 只有一个点, 则称为简单假设; 若含有 2 个及以上, 则称为复合假设。

4.1.2 否定域、检验函数和检验统计量

定义 4.1.1

设样本空间为 $\mathfrak{X}$, 对 $\mathfrak{X}$ 作一个划分:

$$\mathfrak{X} = D \cup \overline{D}, D \cap \overline{D} = \emptyset$$

当样本 $X \in D$, 拒绝 H_0 , 接受 H_1 , D 称为否定域或拒绝域; 当 $x \in \overline{D}$, 拒绝 H_1 , 接受 H_0 , $\overline{D}$ 称为接受

域。

通常情况下 D 可用统计量 $T(X)$ 落在某个区域来表达, 即

$$D = \{X = (X_1, \dots, X_n) : T(X) \in S\}$$

例 4.1.2.

总体 $\sim N(a, \sigma^2)$,

$$H_0 : a = a_0 \longleftrightarrow H_1 : a \neq a_0$$

这里 a 是未知常数, a_0 是 (假设给出的) 已知常数。

根据大数定律, $\bar{X} \xrightarrow{P} a$, H_0 成立, 即 $a = a_0$, 那么 $|\bar{X} - a_0| \leq d$ 则接受 H_0 , 否则拒绝。这种检验被称为**双侧检验**。

如果这家企业比较良心, 只能多装不能少装:

$$H_0 : a = a_0 \longleftrightarrow H_1 : a > a_0$$

$\bar{X} \geq a_0 + d$ 则接受 H_1 。这种检验被称为**单侧检验**。

示性函数可以很方便地表示检验函数:

$$\varphi(X) = I_D = I_{\{T(X) \in S\}} = \begin{cases} 1 & , T(X) \in S \\ 0 & , T(X) \notin S \end{cases}$$

当 $\varphi(X) = 1$ 时否定原假设。在例 4.1.2 中 $\varphi(X) = I_{\{|\bar{X} - a_0| > d\}}$ 和 $\varphi(X) = I_{\{\bar{X} - a_0 > d\}}$ 。

由此可见, 寻找合适的 $T(X)$ 是关键, 它被称为**检验统计量**。

4.1.3 两类错误和功效函数

定义 4.1.2

由于样本有随机性, 所有的判断不能 100% 正确。

1. I 型错误、“弃真”错误: 原假设 H_0 正确, 但根据样本的检验规则把 H_0 否定了, 即 X 落到了否定域中。犯 I 型错误的概率通常用 α 来表示。

2. II 型错误、“存伪”错误: 原假设 H_0 不正确, 但接受了 H_0 , 即 X 落到了接受域中。犯 II 型错误的概率通常用 β 来表示。

在一个检验中, 不可能同时犯两种错误。

为了研究原假设与对立假设地位是否相等, 引入功效函数。

定义 4.1.3

设 $H_0 : \theta \in \Theta_0 \longleftrightarrow H_1 : \theta \in \Theta_1$ 的一个检验函数 $\varphi(X)$, 拒绝域为 D , 则定义

$$\beta_\varphi(\theta) = P_\theta(\text{检验 } \varphi \text{ 否定 } H_0)$$

称为**功效函数**。

由定义可以得到功效函数 $\beta_\varphi(\theta)$ 与犯 I 型、II 型错误概率的关系:

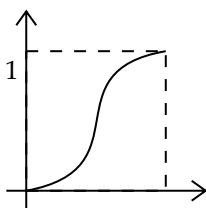
$$\beta_\varphi(\theta) = P_\theta(\text{检验 } \varphi \text{ 否定 } H_0) = P_\theta(X \in D) = E_\theta \varphi(X) = \begin{cases} \text{犯 I 型错误的概率} & , \theta \in \Theta_0 \\ 1 - \text{犯 II 型错误的概率} & , \theta \in \Theta_1 \end{cases}$$

也就是：

$$\alpha_{\varphi}^*(\theta) = \begin{cases} \beta_{\varphi}(\theta) & , \theta \in \Theta_0 \\ 0 & , \theta \in \Theta_1 \end{cases}$$

$$\beta_{\varphi}^*(\theta) = \begin{cases} 0 & , \theta \in \Theta_0 \\ 1 - \beta_{\varphi}(\theta) & , \theta \in \Theta_1 \end{cases}$$

这两个分别是犯 I、II 型错误的概率，我们都希望越小越好。假设样本空间是 $(0, 1)$ ， $\Theta_0 = (0, \frac{1}{2})$ ， $\Theta_1 = (\frac{1}{2}, 1)$ ，那么 $\beta_{\varphi}(\theta)$ - θ 的图像应当类似于：



4.1.4 检验水平和控制犯第一类错误原则

控制犯第一类错误原则：限制 I 型错误的概率不超过 α ，常常取 0.05 或 0.01（分别为宽标准和严标准）。在此基础上寻找检验 φ 使得 II 型错误的概率 β 越小越好。

也就是：记

$$S_{\alpha} = \{ \text{检验 } \varphi : \alpha_{\varphi}^*(\theta) \leq \alpha \} = \{ \text{检验 } \varphi : \sup_{\theta \in \Theta_0} \beta_{\varphi}(\theta) \leq \alpha \}$$

从 S_{α} 中寻找 φ 使得 $\beta_{\varphi}^*(\theta)$ 越小越好。

该原则由 Neyman-Pearson 提出，体现了保护原假设的立场。因为原假设来源于历史中的经验、实际运行的规律等等，并非凭空做出的假设。

定义 4.1.4

设 φ 是 $H_0 : \theta \in \Theta_0 \longleftrightarrow H_1 : \theta \in \Theta_1$ 的一个检验， $\alpha \in [0, 1]$ 。若 φ 满足： $\beta_{\varphi}(\theta) \leq \alpha, \theta \in \Theta_0$ ，即

$$\sup_{\theta \in \Theta_0} \beta_{\varphi}(\theta) \leq \alpha$$

称 α 是检验 φ 的一个水平， φ 称为显著性水平 α 的一个检验，简称水平 α 的检验。

当

$$\sup_{\theta \in \Theta_0} \beta_{\varphi}(\theta) = \alpha$$

时，称 α 为检验 φ 的真实水平：

4.1.5 求解假设检验

Step1: 提出 H_0 和 H_1 . 原则：

1. 把久经考验的事实放在 H_0 .
2. 把你希望得到的结论放在 H_1 ，希望通过检验 H_0 来得到你的结论。¹

Step2: 确定检验统计量。一般方法：

1. 由枢轴变量转化来的统计量，在 4.2 进行介绍；
2. 似然比统计量。

¹为什么要站在保护原假设 H_0 的立场，而不是倾向于自己希望的结论？这是因为，我都充分保护原假设了，最后还是通过检验将其推翻，这就说明我有充分的把握说我的主张是错误的。该原则是为了提高检验的可信度。

Step3: 具体问题具体分析, 选择相应的 α .

Step4: 检验并得出结论。

Step5: 解释得到的结论。

注

1. H_0 和 H_1 是否平等? 是不平等的, 我们一般要站在保护原假设的立场。
2. 如何设立 H_0 和 H_1 ? 要把久经考验的结论设定为 H_0 。
3. 如何根据样本来检验? 寻找检验统计量。
4. 检验方法是否唯一? 如果不唯一, 哪个方法更好? 我们采用“控制犯第一类错误原则”, 在限制 I 型错误的概率不超过 α 的基础上, 检验 φ 使得 II 型错误的概率 $\beta_\varphi^*(\theta)$ 越小则越好, 进而可知功效函数 $\beta_\varphi(\theta)$ 是衡量检验有效性的标准, 越大越有效。

4.2 正态总体参数的假设检验——枢轴变量

考虑检验问题 $H_0: \theta \in \Theta_0 \longleftrightarrow H_1: \theta \in \Theta_1$, 所谓枢轴变量, 是指包含了参数 θ 的良好的点估计、无未知参数的变量, 那么枢轴变量应当落在理论值的附近, 例如正态总体 $N(a, \sigma^2)$ 的样本均值 $\bar{X}$ 应当落在 a 附近。分析枢轴变量的分布, 就可以得到否定域。这种思路与求区间估计十分相似。

4.2.1 单个正态总体的检验

考虑检验问题:

$$H_0: a = a_0 \longleftrightarrow H_1: a \neq a_0 \quad (4.2.1)$$

利用枢轴变量:

$$\frac{\sqrt{n}(\bar{X} - a_0)}{\sigma} \sim N(0, 1)$$

设拒绝域形如 $\{|\bar{X} - a_0| > d\}$, 首先控制犯第一类错误的概率:

$$P(|\bar{X} - a_0| > d | a = a_0) \leq \alpha$$

其中 d 待定。这等价于:

$$P\left(\frac{\sqrt{n}|\bar{X} - a_0|}{\sigma} > \frac{\sqrt{n}d}{\sigma} | a = a_0\right) \leq \alpha$$

利用 $\frac{\sqrt{n}(\bar{X} - a_0)}{\sigma} \sim N(0, 1)$, 找到标准正态分布的 $\alpha/2$ 分位点 $u_{\frac{\alpha}{2}}$, 可以计算出 d 的范围:

$$d \geq \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}} := d_0$$

在此基础上考虑犯第二类错误的概率, 也就是 $a \neq a_0$ 时得出 $|\bar{X} - a_0| \leq d$ 的概率 $P(|\bar{X} - a_0| \leq d | a \neq a_0)$, 我们希望它越小越好:

$$\begin{aligned} P(|\bar{X} - a_0| \leq d) &= P\left(\frac{-d + a_0 - a}{\sigma} \leq \frac{\bar{X} - a}{\sigma} \leq \frac{d + a_0 - a}{\sigma}\right) \\ &= P\left(\frac{d}{\sigma} \leq N - k \leq \frac{d}{\sigma}\right) \end{aligned}$$

其中 $N \sim N(0, 1)$, $k = (a_0 - a)/\sigma$ 为其原点偏移量, 是一个固定值, 可以发现 d 越小时犯 II 型错误的概率越小, 所以取 $d = d_0$. 故可得问题 4.2.1 的拒绝域:

$$D = \{|\bar{X} - a_0| > \frac{\sigma}{\sqrt{n}} u_{\alpha/2}\}$$

写成检验函数的形式为:

$$\varphi(X) = I_{\{|\bar{X} - a_0| > \frac{\sigma}{\sqrt{n}} u_{\alpha/2}\}}$$

拒绝域形如 $(-\infty, \mu_0 - d_0) \cup (\mu_0 + d_0, +\infty)$, 这是一个双侧检验。如果 σ 未知, 用 S 代替 σ , 可以利用:

$$\frac{\sqrt{n}(\bar{X} - a_0)}{S} \sim t_{n-1}$$

推导过程类似, 最终可得拒绝域为:

$$D = \{|\bar{X} - a_0| > \frac{S}{\sqrt{n}} t_{n-1}(\frac{\alpha}{2})\}$$

考虑检验问题:

$$H_0: a = a_0 \longleftrightarrow H_1: a > a_0 \quad (4.2.2)$$

那么它的拒绝域形如 $\{\bar{X} - a_0 > d\}$, 只有一侧, 因此这是一个单侧检验。控制犯第一类错误的概率:

$$P(\bar{X} - a_0 > d | a = a_0) \leq \alpha$$

推导过程完全类似, 最后得到

$$d \geq \frac{\sigma}{\sqrt{n}} u_\alpha := d_0$$

在此基础上控制犯第二类错误的概率, 功效函数

$$\beta_\varphi(a) = P_a(\bar{X} - a_0 > d) = P_a\left(\frac{\bar{X} - a}{\sigma/\sqrt{n}} > \frac{d + a_0 - a}{\sigma/\sqrt{n}}\right) := P_a\left(N > \frac{d - (a - a_0)}{\sigma/\sqrt{n}}\right)$$

越大越好, 故 d 应尽可能小, 因此仍然取 $d = d_0$. 于是问题4.2.2的拒绝域为:

$$D = \{\bar{X} > a_0 + \frac{\sigma}{\sqrt{n}} u_\alpha\}$$

考虑检验问题:

$$H_0: a \leq a_0 \longleftrightarrow H_1: a > a_0 \quad (4.2.3)$$

与前两个例子不同的是, 这里即使原假设成立, 也没办法对枢轴变量给出一个确定的分布, 当 $a \leq a_0$ 时,

$$\beta_\varphi(a) = P(\bar{X} - a_0 > d) \leq \beta_\varphi(a_0) = P(\bar{X} - a_0 > d) = \alpha$$

也就是

$$\sup_{a \in \Theta_0} \beta_\varphi(a) \leq \alpha$$

所以 $\varphi(X)$ 是上面假设检验中水平为 α 的一个检验, 也就是最好的检验, 和单侧检验完全一致。其他方向的情况较为相似, 这里省略。

原假设 H_0	对立假设 H_1	枢轴变量 $T(X)$	拒绝域
$a = a_0$	$a \neq a_0$	σ 已知, $\sqrt{n}(\bar{X} - a_0)/\sigma$ σ 未知, $\sqrt{n}(\bar{X} - a_0)/S$	$\{T(X) : T(X) > u_{\alpha/2}\}$ $\{T(X) : T(X) > t_{n-1}(\alpha/2)\}$
$a = a_0$ 或 $a \leq a_0$	$a > a_0$	σ 已知, $\sqrt{n}(\bar{X} - a_0)/\sigma$ σ 未知, $\sqrt{n}(\bar{X} - a_0)/S$	$\{T(X) : T(X) > u_\alpha\}$ $\{T(X) : T(X) > t_{n-1}(\alpha)\}$
$a = a_0$ 或 $a \geq a_0$	$a < a_0$	σ 已知, $\sqrt{n}(\bar{X} - a_0)/\sigma$ σ 未知, $\sqrt{n}(\bar{X} - a_0)/S$	$\{T(X) : T(X) < u_\alpha\}$ $\{T(X) : T(X) < t_{n-1}(\alpha)\}$

表 4.1: 均值检验

考虑检验问题:

$$H_0: \sigma^2 = \sigma_0^2 \longleftrightarrow H_1: \sigma^2 \neq \sigma_0^2 \quad (4.2.4)$$

如果 a 已知, 利用

$$\frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - a)^2 \sim \chi_n^2$$

如果 a 未知, 利用

$$\frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi_{n-1}^2$$

推导过程完全类似, 这里省略, 直接给出结论:

原假设 H_0	对立假设 H_1	枢组变量 $T(X)$	拒绝域
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	a 已知, $\sum_{i=1}^n (X_i - a)^2 / \sigma_0^2$ a 未知, $\sum_{i=1}^n (X_i - \bar{X})^2 / \sigma_0^2$	$\{T(X) > \chi_n(\frac{\alpha}{2}) \text{ or } T(X) < \chi_n(1 - \frac{\alpha}{2})\}$ $\{T(X) > \chi_{n-1}(\frac{\alpha}{2}) \text{ or } T(X) < \chi_{n-1}(1 - \frac{\alpha}{2})\}$
$\sigma^2 = \sigma_0^2$ 或 $\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	a 已知, $\sum_{i=1}^n (X_i - a)^2 / \sigma_0^2$ a 未知, $\sum_{i=1}^n (X_i - \bar{X})^2 / \sigma_0^2$	$\{T(X) : T(X) > \chi_n(\alpha)^2\}$ $\{T(X) : T(X) > \chi_{n-1}(\alpha)\}$
$\sigma^2 = \sigma_0^2$ 或 $\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	a 已知, $\sum_{i=1}^n (X_i - a)^2 / \sigma_0^2$ a 未知, $\sum_{i=1}^n (X_i - \bar{X})^2 / \sigma_0^2$	$\{T(X) : T(X) < \chi_n(1 - \alpha)^2\}$ $\{T(X) : T(X) < \chi_{n-1}(1 - \alpha)\}$

表 4.2: 方差检验

4.2.2 两个正态总体的检验

4.3 似然比检验

4.3.1 定义与例子

定义 4.3.1

总体 $\sim f(x, \theta)$, 似然函数

$$L(\theta|x) = \prod_{i=1}^n f(x_i, \theta)$$

检验:

$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1 = \Theta_0^c \quad (1)$$

给定样本 $x = X$, 若原假设成立, 则 $L(\theta|X)$ 在 $\theta \in \Theta_0$ 中的取值较大, 因此, 取^a

$$\lambda(X) = \frac{\sup_{\theta \in \Theta} L(\theta|X)}{\sup_{\theta \in \Theta_0} L(\theta|X)} := \frac{L_{\Theta}(X)}{L_{\Theta_0}(X)}$$

如果该统计量数值较大, 则将否定原假设, 该统计量被称为似然比检验统计量 (Likelihood ratio test), 简称 LRT 统计量。我们将由以下检验函数

$$\varphi(X) = \begin{cases} 1 & , \lambda(X) > c \\ r & , \lambda(X) = c \\ 0 & , \lambda(X) < c \end{cases}$$

检验问题 (1), 称之为似然比检验^b。

^a 兰老师 ppt 里的定义是倒过来的, 即分母上取全空间, 不影响后续结论, 注意一下即可。

^b 有点反直觉: $\varphi(X) = 1$ 时否定原假设, 这么看来兰 ppt 上的定义也有一定道理。

实际上直接分析 λ 较为困难, 如果能找到统计量 $T(X)$ 和严格单调的函数 g 使得 $\lambda(X) = g(T(X))$, 则检验函数可以写成:

$$\varphi(X) = \begin{cases} 1 & , T(X) > c' \\ r & , T(X) = c' \\ 0 & , T(X) < c' \end{cases}$$

c', r 待定, 其使得 $\beta_{\varphi}(\theta) = \alpha, \theta \in \Theta$.

例 4.3.1.

总体 $\sim N(\mu, \sigma^2)$, 求水平 α 的似然比检验:

$$H_0: \mu = \mu_0 \longleftrightarrow H_1: \mu \neq \mu_0$$

解答: 设 $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+$, 似然函数

$$L(\theta|x) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}$$

已经求解过极大似然估计:

$$\hat{\mu}_L = \bar{X}, \quad \hat{\sigma}_L = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

那么

$$\sup_{\theta \in \Theta} L(\theta|x) = \left(\frac{2\pi}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{n}{2}\right\} = \left(\frac{2\pi e}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{-\frac{n}{2}}$$

对于 $\theta \in \Theta_0 = \{\mu_0\} \times \mathbb{R}_+$, 直接代入

$$\sup_{\theta \in \Theta_0} L(\theta|x) = \left(\frac{2\pi e}{n} \sum_{i=1}^n (x_i - \mu_0)^2\right)^{-\frac{n}{2}}$$

因此

$$\lambda(X) = \left(\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right)^{\frac{n}{2}} = \left(1 + \frac{n(\bar{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right)^{\frac{n}{2}}$$

令

$$T(X) = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}$$

那么

$$\lambda(X) = \left(1 + \frac{1}{n-1} T(X)^2\right)^{\frac{n}{2}}$$

该函数关于 $|T(X)|$ 单调递增, 所以

$$\varphi(X) = \begin{cases} 1 & , |T(X)| > c' \\ 0 & , |T(X)| < c' \end{cases}$$

实际上, $T(X) \sim t_{n-1}$, 因此对于水平 α 的检验, $c' = t_{n-1}(\alpha/2)$.

例 4.3.2. 相似习题: Ex5.30

总体 $\sim N(\mu, \sigma^2)$, 求水平 α 的似然比检验:

$$H_0: \mu \leq \mu_0 \longleftrightarrow H_1: \mu > \mu_0$$

解答: 变化的只有参数空间: $\theta \in \Theta_0 = (-\infty, \mu_0] \times \mathbb{R}_+$,

$$\begin{aligned} \sup_{\theta \in \Theta_0} L(\theta|x) &= \left(\frac{2\pi}{n}\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \\ &= \left(\frac{2\pi}{n}\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2\right\} \exp\left\{-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right\} \end{aligned}$$

分析 $\exp\left\{-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right\}$ 最大值:

1° $\bar{x} > \mu_0$, 最大值在 $\mu = \mu_0$ 处达到;

2° $\bar{x} \leq \mu_0$, 最大值在 $\mu = \bar{x}$ 处达到。

因此

$$\lambda(X) = \begin{cases} 1 & , \bar{X} \leq \mu_0 \\ \left(1 + \frac{1}{n-1}T(X)^2\right)^{-\frac{n}{2}} & , \bar{X} \geq \mu_0 \end{cases}$$

也是关于 $|T(X)|$ 单调递增的, 但是这里要写成

$$\varphi(X) = \begin{cases} 1 & , T(X) > c' \\ 0 & , T(X) \leq c' \end{cases}$$

不能加绝对值, 因为这里是单边检验. $\mu = \mu_0$ 下, $T(X) \sim t_{n-1}$, 于是取 $c' = t_{n-1}(\alpha)$. 当 $\mu \leq \mu_0 \in \Theta_0$, $\beta_\varphi(\mu)$ 在 H_0 下关于 μ 单调递增, $\beta_\varphi(\mu) \leq \beta_\varphi(\mu_0) = \alpha$. 所以此即为所求的水平 α 的似然比检验.

例 4.3.3.

总体 $\sim N(\mu, \sigma^2)$, 求水平 α 的似然比检验:

$$H_0: \sigma^2 = \sigma_0^2 \longleftrightarrow H_1: \sigma^2 \neq \sigma_0^2$$

解答: 同样地, 只有参数空间 Θ_0 不同: $\theta \in \Theta_0 = \mathbb{R} \times \{\sigma_0^2\}$ 时,

$$L(\theta|x) = (2\pi\sigma_0)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2\right\}$$

在 $\mu = \bar{x}$ 时达到最大值, 所以

$$\lambda(X) = \left(\frac{e}{n}\right)^{-\frac{n}{2}} \left[\frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x})^2\right]^{-\frac{n}{2}} \exp\left\{\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x})^2\right\}$$

令

$$\xi = \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x})^2, \quad g(\xi) = \xi^{-\frac{n}{2}} e^{\frac{\xi}{2}}$$

于是

$$\lambda(X) = \left(\frac{e}{n}\right)^{-\frac{n}{2}} g(\xi)$$

$g(\xi)$ 先减后增, 因此

$$\lambda(X) \leq c \Leftrightarrow g(\xi(X)) \leq c \Leftrightarrow \xi(X) \in [k_1, k_2], \quad g(k_1) = g(k_2) = c, \quad P(\xi(X) \in [k_1, k_2]) = 1 - \alpha$$

也就是说我们要找一个区间 $[k_1, k_2]$, 它只与 ξ 的分布有关, 而 $\xi(X) \sim \chi_{n-1}^2$, 所以可取

$$[k_1, k_2] = [\chi_{n-1}^2(1 - \alpha/2), \chi_{n-1}^2(\alpha/2)]$$

例 4.3.4.

总体 $\sim U(0, \theta)$, $\theta > 0$, 求水平 α 的似然比检验:

$$H_0: \theta \leq \theta_0 \longleftrightarrow H_1: \theta > \theta_0$$

解答: 似然函数

$$L(\theta|x) = \frac{1}{\theta^n} I_{0 < x_{(n)} < \theta}$$

极大似然估计 $\hat{\theta}_L = X_{(n)}$, $L_{\Theta} = X_{(n)}^{-n}$. 原假设下, $L_{\Theta_0} = X_{(n)}^{-n} I_{(0, \theta_0)}(X_{(n)})$, 于是

$$\lambda(X) = \begin{cases} 1 & , 0 < X_{(n)} < \theta_0 \\ +\infty & , X_{(n)} > \theta_0 \end{cases}$$

可以求出, $X_{(n)}$ 的 p.d.f 为

$$f_n(t) = \frac{nt^{n-1}}{\theta^n} I_{(0, \theta)}(t)$$

c 应满足 $P_{\Theta_0}(X_{(n)} > c) = \alpha$, 当 $\theta = \theta_0$ 时,

$$\int_c^{\theta_0} \frac{nt^{n-1}}{\theta_0^n} dt = \alpha \Rightarrow c = \theta_0(1 - \alpha)^{\frac{1}{n}}$$

拒绝域为

$$D = \{X_{(n)} > \theta_0(1 - \alpha)^{\frac{1}{n}}\}$$

若 $\theta \leq \theta_0$,

$$\beta_\varphi(\theta) = P_\theta(X_{(n)} > \theta_0(1 - \alpha)^{\frac{1}{n}}) = 1 - (1 - \alpha) \left(\frac{\theta_0}{\theta}\right)^n$$

是增函数, 所以 $\beta_\varphi(\theta) \leq \beta_\varphi(\theta_0), \theta \leq \theta_0$. 所以此即为所求的水平 α 的似然比检验。

例 4.3.5.

总体 $\sim \exp\{\frac{1}{\lambda}\}$,

$$f(x, \lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} I_{(0, +\infty)}(x)$$

求水平 α 的 LR 检验:

$$H_0: \lambda = \lambda_0 \longleftrightarrow H_1: \lambda \neq \lambda_0$$

解答: 似然函数:

$$L(\lambda|x) = \lambda^{-n} \exp\left\{-\frac{n\bar{x}}{\lambda}\right\} I_{(0, +\infty)}(x_{(1)})$$

$\Theta_0 = \{\lambda_0\}$, 所以

$$L_{\Theta}(x) = \sup_{\lambda \in (0, +\infty)} \lambda^{-n} \exp\left\{-\frac{n\bar{x}}{\lambda}\right\} I_{(0, +\infty)}(x_{(1)}) = e^{-n\bar{x}}$$

$$L_{\Theta_0}(x) = \lambda_0^{-n} \exp\left\{-\frac{n\bar{x}}{\lambda_0}\right\}$$

记 $T(X) = n\bar{X}$, 则

$$\lambda(T) = \left(\frac{n\lambda_0}{e}\right)^n \frac{1}{T^n} e^{-\frac{T}{\lambda_0}} := C \cdot g(T)$$

分析 $g(T)$ 可知其先增后减, 所以拒绝域 $\{X: \lambda(X) \leq c\}$ 形如 $\{X: T(X) \in (-\infty, k_1] \cup [k_2, +\infty)\}$, $2T/\lambda_0 \sim \chi_{2n}^2$, 所以可以取

$$(-\infty, k_1] \cup [k_2, +\infty) = (-\infty, \frac{\lambda_0}{2} \chi_{2n}^2(1 - \alpha/2)] \cup [\frac{\lambda_0}{2} \chi_{2n}^2(\alpha/2), +\infty)$$

定理 4.3.1

$T(X)$ 是关于 θ 的一个充分统计量, $\lambda^*(T)$ 和 $\lambda(X)$ 分别是依赖于 T 和 X 的 LRT 统计量, 则对于样本空间内的每一点 X , 均有 $\lambda^*(T(X)) = \lambda(X)$.

这个定理告诉我们, 拒绝域 $D = \{X: \lambda \leq c\}$ 经过简化之后, 一定能化为只与充分统计量有关的形式, 可以简化寻找 LRT 统计量的过程。

例 4.3.6.

总体 $\sim N(\theta, 1)$, 利用充分统计量, 求水平 α 的似然比检验:

$$H_0: \theta = \theta_0 \longleftrightarrow H_1: \theta \neq \theta_0$$

解答: $\bar{X}$ 是 θ 的充分统计量, 且 $\bar{X} \sim N(\theta, \frac{1}{n})$, 因此似然比检验的拒绝域形如 $\{X: |\bar{X} - \theta_0| \geq c\}$.

4.4 一致最优检验

定义 4.4.1

设检验问题为

$$H_0: \theta \in \Theta_0 \longleftrightarrow H_1: \theta \in \Theta_1 \quad (4.4.1)$$

记 Φ_α 为检验问题 4.4.1 的一切水平为 α 的检验 ($\beta_\varphi(\theta) \leq \alpha, \theta \in \Theta_0$), 若存在 $\varphi \in \Phi_\alpha$, 对于任一 $\varphi_1 \in \Phi_\alpha$ 有

$$\beta_\varphi(\theta) \geq \beta_{\varphi_1}(\theta), \theta \in \Theta_1$$

则称 φ 为检验 4.4.1 的水平 α 的一致最优检验 (uniformly most powerful test, UMPT)。

引理 4.4.1 (N-P 引理)

随机样本 $X = (X_1, \dots, X_n) \sim f(x_1, \dots, x_n, \theta) := f(x, \theta), \forall \alpha \in (0, 1)$, 考虑检验问题:

$$H_0: \theta = \theta_0 \longleftrightarrow H_1: \theta = \theta_1 \quad (4.4.2)$$

1. 存在性: 检验问题 4.4.2 存在检验 φ 、非负常数 c 和 $r \in (0, 1)$ 使得

$$\varphi(x) = \begin{cases} 1 & , \frac{f(x, \theta_1)}{f(x, \theta_0)} > c \\ r & , \frac{f(x, \theta_1)}{f(x, \theta_0)} = c \\ 0 & , \frac{f(x, \theta_1)}{f(x, \theta_0)} < c \end{cases} \quad (4.4.3)$$

且

$$E_{\theta_0} \varphi(X) = \alpha \quad (4.4.4)$$

2. 一致最优性: 任何满足式 (4.4.3) 和 (4.4.4) 的 φ 是检验问题 (4.4.2) 的 UMPT.

证明

例 4.4.1.

总体 $\sim N(\mu, 1)$, 求检验问题:

$$H_0: \mu = 0 \longleftrightarrow H_1: \mu = \mu_1$$

的 UMPT.

解答: 样本

$$X = (X_1, \dots, X_n) \sim (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \right\}$$

所以

$$\begin{aligned} f(x, \theta_0) &= (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n x_i^2 \right\} \\ f(x, \theta_1) &= (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \\ \lambda(x) &= \frac{f(x, \theta_0)}{f(x, \theta_1)} = \exp \left\{ -\frac{1}{2} n \mu_1^2 + n \mu_1 \bar{x} \right\} \end{aligned}$$

当 $\mu_1 > 0$ 时, $\lambda(x)$ 为 $\bar{x}$ 的严格单调递增函数, 所以 UMPT 的否定域为 $\{x: \lambda(x) > C'\} = \{x: \sqrt{n}\bar{x} > c\}$, 其中 $\bar{x} \sim N(0, \frac{1}{n})$, 给定 $\alpha \in (0, 1)$, 那么 $C = u_\alpha$. 即 UMPT 为

$$\varphi(X) = I_{\{\bar{X} > \frac{1}{\sqrt{n}} u_\alpha\}}$$

注 注意到上述结果 UMPT 的结果与 μ_1 无关, 所以 $\varphi(X)$ 也是检验问题:

$$H_0: \mu = 0 \longleftrightarrow H_1: \mu > 0 \quad (4.4.5)$$

的 UMPT.

定义 4.4.2

设一元随机变量 $T \sim g(t|\theta), \theta \in \Theta$, 称之关于 θ 有单调似然比 (MLR), 如果:

$$\forall \theta_2 > \theta_1, \frac{g(t|\theta_1)}{g(t|\theta_2)} \text{ 在 } \{t: g(t|\theta_1) > 0 \text{ or } g(t|\theta_2) > 0\} \text{ 上是 } t \text{ 的单调函数}$$

定理 4.4.1 (K-R Thm)

对于检验问题

$$H_0: \theta \leq \theta_0 \longleftrightarrow H_1: \theta > \theta_0 \quad (4.4.6)$$

设 T 是关于 θ 的一个充分统计量, 且有非降的 MLR, 则问题 4.4.6 的 UMPT 的拒绝域形如 $\{T > t_0\}$, 检验水平为 $\alpha = P(T > t_0 | \theta = \theta_0)$.

推论 4.4.1

如果总体的分布属于一维指数族:

$$X \sim c(\theta) \exp \{Q(\theta)T(x)\} h(x)$$

且其中 $Q(\theta)$ 非降, 那么问题 4.4.6 的 UMPT 的拒绝域形如 $\{T > t_0\}$, 检验水平为 $\alpha = P(T > t_0 | \theta = \theta_0)$.

第 5 章 非参数假设检验

参数假设检验只有在总体类型已知时才能使用，本章所介绍的是总体类型未知时可以使用的非参数假设检验。

5.1 符号检验与符号秩和检验

5.1.1 符号检验

一、小样本方法：

例 5.1.1. 符号检验

有甲、乙两种啤酒，希望评估年轻人对于两种啤酒的喜爱程度，于是随机邀请 N 个年轻人品尝两种啤酒，并给出对于两种啤酒的评分，得到样本 (x_i, y_i) ，记 $z_i = y_i - x_i$ ， $n^+ = \#\{z_i > 0\}$ ， $n^- = \#\{z_i < 0\}$ ， $n = n^+ + n^-$ 。

H_0 : 两啤酒口味无差异 $\leftrightarrow H_1$: 有差异

在 H_0 成立的情况下， $n^+ \sim B(n, \frac{1}{2})$ ，否定域 $D = \{n^+ \geq c\} \cup \{n^+ \leq d\}$ ，如果

$$P_{H_0}(\{X \geq c\} \cup \{X \leq d\}) \geq \alpha, \text{接受 } H_0$$

否则拒绝 H_0 ，现在的任务是给出合适的 c, d ，使得

$$2P_{H_0}(X \geq c) < \alpha, d = n - c$$

于是 $n^+ \in (c, d)$ 则接受 H_0 ，否则拒绝 H_0 。

以上是符号检验的一个例子，可以发现它的使用过程中“只考虑数据类型的正负”，从而利用“正数据”服从二项分布的这一特点进行检验。但这会忽略很多信息，例如中虽然某人对甲乙都做出正面评价，但对待两种啤酒的相对态度可能会有很大差别，比如 0.0001 和 1 都算正数据，这并不是十分合理。所以，实际情况中要自行判断是否要使用符号检验。

二、大样本方法：根据中心极限定理，

$$\frac{n^+ - \frac{n}{2}}{\sqrt{\frac{n}{4}}} = \frac{2n_+ - n}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0, 1)$$

此时否定域为

$$D = \left\{ \frac{|2n_+ - n|}{\sqrt{n}} > u_{\alpha/2} \right\}$$

5.1.2 符号秩和检验

定义 5.1.1

设 $X_1, \dots, X_n$ 为两两不相等的一组样本，将其按照大小排列为 $X_{(1)} < \dots, X_{(n)}$ ，若 $X_i = X_{(R_i)}$ ，则称 X_i 在样本 $(X_1, \dots, X_n)$ 中的秩为 R_i ， $R = (R_1, \dots, R_n)$ 称为 $(X_1, \dots, X_n)$ 的秩统计量。基于秩统计量的检验方法称为秩检验。

一、小样本方法：

例 5.1.2. Wilcoxon 双侧符号秩和检验

问题背景同例 5.1.1，假设请了 13 个人品尝啤酒，评分结果见下表：

	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 ($\text{sgn } z_i$)	+	-	-	-	+	-	-	-	-	-	-	-	0
$ z_i = x_i - y_i $	20	5	2.1	4.5	26	2.3	4	1.1	3	1.3	2.8	1.5	0
秩 (R_i)	[11]	10	4	9	[12]	5	8	1	7	2	6	3	不定

表 5.1: 评分与秩

将符号为 + 的秩相加得到的和，称之为符号秩和，即：

$$W^+ \stackrel{\text{def}}{=} \sum_{i=1}^n I_{\{z_i > 0\}} \cdot R_i$$

那么

$$H_0: \text{两啤酒口味无差异} \longleftrightarrow H_1: \text{有差异}$$

在 H_0 成立的情况下，否定域为 $D = \{W^+ \geq c\} \cup \{W^+ \leq d\}$ ，此处的 c, d 应满足

$$P_{H_0}(W^+ \geq c) \leq \alpha/2, \quad d = \frac{n(n+1)}{2} - c$$

在本例中， $W^+ = 23$ ， $n = 12$ ，取 $\alpha = 0.10$ ，查表可得 $c = 61, d = 17$ ，应当接受 H_0 。

以上是符号秩和检验的一个例子，这种检验方法叫做 **Wilcoxon 双侧符号秩和检验**，简称双侧 W^+ 检验。但要注意， W^+ 检验既不忽视数值，也不全看数值，介于符号检验与基于正态假定 t 检验之间。

二、大样本方法：可以证明

$$\mathbb{E}(W^+) = \frac{n(n+1)}{4}, \quad \text{var}(W^+) = \frac{n(n+1)(2n+1)}{24}$$

所以

$$W_*^+ \stackrel{\text{def}}{=} \frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \xrightarrow{\mathcal{L}} N(0, 1)$$

此时否定域为

$$D = \{|W_*| > u_{\alpha/2}\}$$

5.2 Wilcoxon 两样本秩和检验

一、问题背景：两样本检验问题（也称两样本问题）的一般提法如下：设 $X_1, \dots, X_m$ 和 $Y_1, \dots, Y_n$ 分别是分布为 F_1 和 F_2 的一维总体中抽取的简单样本，且假定两组样本相互独立，目的是检验下列假设：

$$H_0: F_1 = F_2 \longleftrightarrow H_1: F_1 \neq F_2$$

考虑以下实际情况：

1. 如果有理由假定 F_1 和 F_2 为具有相同方差的正态分布：

$$F_1 \sim N(a, \sigma^2), F_2 \sim N(b, \sigma^2)$$

其中 a, b, σ^2 未知，这属于参数假设检验问题中的两样本 t 检验。

2. 如果 F_1 和 F_2 完全未知，处理该问题的方法之一为斯米尔诺夫检验，我们将在 5.5 节中讨论。
3. 本小节所关注的是一种中间情况： F_1 和 F_2 分布类型未知，但是我们可以确定二者只相差一个位置参数，也就是说可以设 $F_1 \sim F(x)$ ， $F_2 \sim F(x - \theta)$ ，其中 θ 未知，此时 $H_0: \theta = 0$ 。

定义 5.2.1

设 $X_1, \dots, X_m, Y_1, \dots, Y_n$ 两两不相等，按照大小排列结果为

$$Z_1 < Z_2 < \dots < Z_N, N = n + m$$

若 $Y_i = Z_{R_i}$ ，则称 Y_i 在合样本 $X_1, \dots, X_m, Y_1, \dots, Y_n$ 中的秩为 R_i ， $Y_1, \dots, Y_n$ 的秩和为 $W = R_1 + \dots + R_n$ ，这被称为 Wilcoxon 两样本秩和统计量。

二、小样本方法：如果 H_0 成立， W 所取之值应当集中在平均数 $n(N+1)/2$ 附近，否定域形如 $D = \{W \geq c\} \cup \{W \leq d\}$ 。($R_1, \dots, R_n$) 的联合分布为

$$P(R_1 = r_1, \dots, R_n = r_n) = \frac{1}{N(N-1)\dots(N-n+1)}, r_1, \dots, r_n \leq N \text{ 为互不相同的自然数}$$

由此不难形式地写出 W 的分布，进而由 $\alpha = P(W \leq d \text{ or } W \geq c | H_0)$ 确定出 c 和 d ，见课本附表 12。这里的 c 和 d 满足 $c + d = n(m + n + 1)$ 。

如果检验是单边的，即要检验

$$H'_0: \theta \leq 0 \longleftrightarrow H'_1: \theta > 0$$

则否定域为 $D = \{W \geq c\}$ ，反之亦然。

例 5.2.1.

某种羊毛在进行工艺处理之前和之后，各抽取一个容量为 5 的样本，测得其含脂率分别为：(0.20, 0.24, 0.66, 0.42, 0.12) 和 (0.13, 0.07, 0.21, 0.08, 0.19)。处理后含脂率是否下降 ($\alpha = 0.05$)？

解答：用 X, Y 分别表示处理前后含脂率，分布函数分别为 $F(x)$ 与 $F(x - \theta)$ ，则检验

$$H_0: \theta \geq 0 \longleftrightarrow \theta < 0$$

将两组样本观察值按照从小到大排列：

1	2	3	4	5	6	7	8	9	10
[0.07]	[0.08]	0.12	[0.13]	[0.19]	0.20	[0.21]	0.24	0.42	0.66

表 5.2: 含脂率（括号内为处理后）

本题中 $m = n = 5, \alpha = 0.05, Y$ 的秩和 $W = 19$, 查表可得 $c = 36, d = 19$, 因此否定 H_0 .

三、大样本方法: 可以证明

$$\mathbb{E}(W) = \frac{n(N+1)}{2}, \text{var}(W) = \frac{nm(N+1)}{12}$$

所以

$$W^* \stackrel{\text{def}}{=} \frac{W - n(N+1)/2}{\sqrt{nm(N+1)/12}} \xrightarrow{\mathcal{L}} N(0, 1)$$

此时否定域为

$$D = \{|W^*| > u_{\alpha/2}\}$$

单侧检验则为 $D = \{W^* \geq u_{\alpha}\}$ 或者 $D = \{W^* \leq -u_{\alpha}\}$

注 秩和检验方法假定了样本互不相同, 如果 F 处处连续, 则该事实将以概率 1 成立。但是 F 不连续时可能存在相同的样本, 即所谓“结”的问题, 例如

$$x_1 < x_2 = x_4 < x_5 < x_3$$

此时我们将 x_2 和 x_4 的秩定义为 2 和 3 的平均数 2.5, 其余情况类似。对所有的结处理之后, 按照前面所述方法讨论秩和检验问题。

5.3 拟合优度检验

5.3.1 问题背景

拟合优度检验问题的提法如下: 设有一个一维或者多维的随机变量 X , 样本 $(X_1, \dots, X_n)$, F 是已知的分布函数, 我们将要利用样本 $X_1, \dots, X_n$ 取检验假设:

$$H_0: \text{r.v. } X \text{ 的分布为 } F$$

其中 F 称为理论分布。

导出这种假设检验的想法大致如下: 设法提出一个反映实际数据 $X_1, \dots, X_n$ 与理论分布 F 偏差的量 $D = D(X_1, \dots, X_n; F)$, 如果 D 较大则否定原假设。但一般来说, 理论与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? 对此问题将提供一个介于 0 和 1 的数字作为回答:

$$p(d_0) = P(D \geq d_0 | H_0)$$

为在选定的偏离指标 D 下, 样本与理论分布的拟合优度, 其中 d_0 为实际样本观测值计算出的理论偏差。

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的一个检验如下:

$$p(d_0) < \alpha \text{ 时否定 } H_0$$

这种类型的检验就称为拟合优度检验。问题的关键在于怎样给出一个合适的 D 的定义, 最著名的是 K. Pearson 在 1900 年提出的 χ^2 检验 (称为 Pearson χ^2 检验) 和柯尔莫哥洛夫在 1933 年提出的一种检验。本节讨论前者, 后者将在 5.5 节介绍。

5.3.2 Pearson χ^2 检验: 理论分布完全已知

一、理论分布为离散型分布且只取有限个不同值: 设 $X_1, \dots, X_n$ 为总体 X 抽取的简单样本, 理论分布为:

$$F: \begin{pmatrix} a_1, & a_2, & \dots, & a_r \\ p_1, & p_2, & \dots, & p_r \end{pmatrix}$$

其中的 $p_1, \dots, p_r$ 已知且和为 1, 检验问题:

$$H_0: P(X = a_i) = p_i, i = 1, 2, \dots, r$$

设样本 $X_1, \dots, X_n$ 中等于 a_i 的个数为 v_i , 称为 a_i 的观察频数, 而 np_i 称为 a_i 的理论频数。可见

$$\sum_{i=1}^r c_i (v_i/n - p_i)^2$$

可作为样本与理论分布偏差的一种度量, 取平方是防止抵消, 那么该如何选取 c_i 的值?

定理 5.3.1 (K. Pearson, 1900)

若取 $c_i = n/p_i$, 在 H_0 成立的前提下,

$$K_n = K_n(X_1, \dots, X_n; F) = \sum_{i=1}^r \frac{(v_i - np_i)^2}{np_i}$$

的极限分布为 χ_{r-1}^2 , 即

$$K_n \xrightarrow{\mathcal{L}} \chi_{r-1}^2$$



按照这一定理可以提出如下的检验方法: 当 n 充分大时, 可以近似地认为检验统计量 K_n 的分布就是 χ_{r-1}^2 , 于是得到检验问题的水平近似为 α 的检验:

$$K_n > \chi_{r-1}^2(\alpha) \text{ 时否定 } H_0$$

记 k_0 为对一组具体样本算出的 K_n 之值, 则

$$p(k_0) = P(K_n \geq k_0 | H_0) \approx P(\chi_{r-1}^2 \geq k_0)$$

就是拟合优度, 它代表检验统计量 K_n 落在 $[k_0, +\infty)$ 中的可能性。

二、理论分布为任一确定分布: 这一情形包括离散型但取可列个值, 或者连续分布的情形。

设 $X_1, \dots, X_n$ 是从总体 X 中抽取的简单样本, 要检验

$$H_0: \text{r.v. } X \text{ 的分布为 } F$$

其中 F 为已知分布, 具体步骤如下:

1. 如果 X 为一维, 取 $r-1$ 个常数 $a_1, \dots, a_{r-1}$ 满足

$$a_0 = -\infty < a_1 < a_2 < \dots < a_{r-1} < +\infty = a_r$$

将实数轴 $(-\infty, +\infty)$ 分成 r 个子区间:

$$I_1 = (-\infty, a_1), I_2 = [a_1, a_2), \dots, I_j = [a_{j-1}, a_j), \dots, I_r = [a_{r-1}, +\infty)$$

如果 X 为 m 维, 则将 $\mathbb{R}^m$ 分割成 r 个彼此无公共点的区域 $I_1, \dots, I_r$.

2. 计算 r 个事件在 H_0 成立下的概率

$$p_j = P_F(X \in I_j) = F(a_j) - F(a_{j-1}), \quad j = 1, 2, \dots, r$$

而且 $F(a_0) = 0, F(a_r) = 1$, 则检验问题转化为

$$H_0: P(X \in I_j) = p_j, \quad j = 1, 2, \dots, r$$

3. 令 v_j 维 $X_1, \dots, X_n$ 落入区间 I_j 的观察频数, 计算检验统计量

$$K_n = \sum_{j=1}^r \frac{(v_j - np_j)^2}{np_j}$$

在 H_0 成立时, 定理 5.4.1 仍然成立。

4. 记 k_0 为 K_n 算出的具体值, 计算出拟合优度 $p(k_0) \approx P(\chi_{r-1}^2 \geq k_0)$, 对检验问题作出结论。

注 需要注意的是, 第 1 步中 $a_1, \dots, a_{r-1}$ 的选择应当使得理论频数 np_i 和观察频数 v_i 都不小于 5 为宜, 否则应当将相邻子区间合并, 直到满足上述要求为止; $a_1, \dots, a_{r-1}$ 必须不依赖于样本, 不能根据样本的位置而去选择它们。

5.3.3 Pearson χ^2 检验：理论分布有未知参数

此时要检验的假设是： $r.v.X$ 的分布属于一个确定的分布族 $\{F(x; \theta_1, \dots, \theta_s) : (\theta_1, \dots, \theta_s) \in \Theta\}$ ，其中 Θ 为参数空间。令 $X_1, \dots, X_n$ 为从总体 X 中抽取的简单样本，检验假设

$$H_0 : \exists (\theta_1^0, \dots, \theta_s^0) \in \Theta \text{ s.t. } X \sim F(x; \theta_1^0, \dots, \theta_s^0)$$

求解上述检验方法是上一小节的直接推广，同样取出 $\mathbb{R}$ 的一个分割： $I_1, \dots, I_r$ ，并改为以下形式：

$$p_j(\theta_1, \dots, \theta_s) = P(X \in I_j) = F(a_j; \theta_1, \dots, \theta_s) - F(a_{j-1}; \theta_1, \dots, \theta_s), \quad j = 1, 2, \dots, r, \quad a_0 = -\infty, a_r = +\infty$$

此时检验问题转化为

$$H_0 : \exists (\theta_1^0, \dots, \theta_s^0) \in \Theta \text{ s.t. } P(X \in I_j) = p_j(\theta_1^0, \dots, \theta_s^0), \quad j = 1, 2, \dots, r$$

其中对 $p_j(\theta_1, \dots, \theta_s)$ 作如下假定：

(1). 对于任何 $(\theta_1, \dots, \theta_s) \in \Theta$ ， $p_j = p_j(\theta_1, \dots, \theta_s) > 0$ ， $\sum p_j = 1$ 。

(2). 对于任何 $1 \leq j \leq r$ ， $p_j(\theta_1, \dots, \theta_s)$ 对 $\theta_1, \dots, \theta_s$ 有一阶连续偏导数。

算出样本 $X_1, \dots, X_n$ 落入 I_j 的观察频数 v_j ，同时按照某种估计方法，用样本 $X_1, \dots, X_n$ 对 $\theta_1, \dots, \theta_s$ 做出估计 $\hat{\theta}_1, \dots, \hat{\theta}_s$ ，那么可得检验统计量

$$K_n^* = K_n^*(\hat{\theta}_1, \dots, \hat{\theta}_s) = \sum_{j=1}^r \frac{(v_j - np_j(\hat{\theta}_1, \dots, \hat{\theta}_s))^2}{np_j(\hat{\theta}_1, \dots, \hat{\theta}_s)}$$

于是问题转化为研究 K_n^* 的极限分布，从而作出拟合优度检验。

Pearson 曾错误地认为 K_n^* 的分布仍然收敛于 χ_{r-1}^2 ，后来 Fisher 在 20 世纪 20 年代发现了这一疏忽，并指出其极限分布的自由度不是 $r-1$ ，而是 $r-1-s$ ，下面简述 Fisher 方法及其主要结果。

定理 5.3.2 (R.A. Fisher)

设 $(\theta_1^0, \dots, \theta_s^0)$ 为参数空间 Θ 的内点，使得 X 的分布为 $F(x, \theta_1^0, \dots, \theta_s^0)$ ，通过极大似然方法对 $\theta_1, \dots, \theta_s$ 作一估计：设 v_j 为样本 $X_1, \dots, X_n$ 中落入 I_j 的个数， $j = 1, 2, \dots, r$ ，似然函数为

$$L = \frac{n!}{v_1! \cdots v_r!} (p_1(\theta_1, \dots, \theta_s))^{v_1} \cdots (p_r(\theta_1, \dots, \theta_s))^{v_r}$$

对于 $\ln L$ 关于 θ_i 求导，可得方程组：

$$\sum_{j=1}^r \frac{v_j}{p_j(\theta_1, \dots, \theta_s)} \cdot \frac{\partial p_j(\theta_1, \dots, \theta_s)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, s$$

解得 $\hat{\theta}_i = \hat{\theta}_i(X_1, \dots, X_n)$ ， $i = 1, 2, \dots, s$ ，由此给出

$$K_n^* = K_n^*(\hat{\theta}_1, \dots, \hat{\theta}_s) = \sum_{j=1}^r \frac{(v_j - np_j(\hat{\theta}_1, \dots, \hat{\theta}_s))^2}{np_j(\hat{\theta}_1, \dots, \hat{\theta}_s)}$$

那么在 H_0 成立的情况下，当 $n \rightarrow \infty$ 时有

$$K_n^* \xrightarrow{\mathcal{L}} \chi_{r-1-s}^2$$

然后利用此定理我们就可以给出检验问题的水平近似为 α 的大样本检验如下：

$$K_n^* > \chi_{r-1-s}^2(\alpha) \text{ 时否定 } H_0$$

类似地，可以计算拟合优度：

$$p(k_0^*) = P(K_n^* > k_0^* | H_0) \approx P(\chi_{r-1-s}^2 > k_0^*)$$

注 要解定理 5.3.2 中的方程组并不容易，一个常用的做法为：直接通过样本而不是频数求出 MLE，可以大大简化计算。然而，用样本给出的 MLE 所计算出的检验统计量 $\tilde{k}_0^*$ 并不一定有极限分布 χ_{r-1-s}^2 ，真正的拟合优度值将介于 $P(\chi_{r-1-s}^2 \geq \tilde{k}_0)$ 和 $P(\chi_{r-1}^2 \geq \tilde{k}_0^*)$ 之间。所以这将求出一个近似结果，虽然有差别但差别不大。

例 5.3.1.

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, 各锭子的断头记录见下表:

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

表 5.3: 样本观测值

问锭子的断头数是否服从 Poisson 分布 ($\alpha = 0.05$)?

解答: 设锭子的断头数为 X , 要检验的假设为 H_0 : 存在 λ_0 使得 $X \sim \text{Poisson}$ 分布 $P(\lambda_0)$. 将数据重新分组, 使得每组断头的锭子数不少于 5:

每锭断头数	0	1	2	3-8
锭数 (实测)	263	112	38	30

表 5.4: 样本观测值 (分组后)

然后求未知参数的 MLE, 这里直接给出结果:

$$\hat{\lambda}^* = \bar{X} = \frac{308}{443} = 0.7$$

接下来求 $\tilde{K}_n^*$, 本例中 $n = 443$,

$$n\hat{p}_0^* = nP_{\hat{\lambda}^*}(X=0) = 443 \times \frac{0.7^0 e^{-0.7}}{0!} = 219.99$$

$$n\hat{p}_1^* = 443 \times \frac{0.7^1 e^{-0.7}}{1!} = 153.99$$

$$n\hat{p}_2^* = 443 \times \frac{0.7^2 e^{-0.7}}{2!} = 53.90$$

$$n\hat{p}_3^* = 443 \times P(3 \leq X \leq 8) = 15.13$$

得到 $\tilde{K}_n^* = 39.16$, 最后求拟合优度:

$$p(\tilde{k}_0^*) \approx P(\chi_2^2 > 39.16) < 0.005 < \alpha$$

于是否定 H_0 , 即认为锭子的断头数不服从 Poisson 分布。

例 5.3.2.

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) 分布见下表:

压强区间	190-200	200-210	210-220	220-230	230-240	240-250
频数 v_i	10	26	56	64	30	14

表 5.5: 抗压强度分布

设 $X_1, \dots, X_{200}$ 为从总体 $F(x)$ 中抽取的简单样本, $F(x)$ 为压强分布函数, 问 $F(x)$ 是否为正态分布 ($\alpha = 0.05$)?

解答: 用 X 表示混凝土制件的抗压强度, 记 $\theta = (\mu, \sigma^2)$, 检验问题为

$$H_0: \exists \theta_0 = (\mu_0, \sigma_0^2) \text{ s.t. } X \sim N(\mu_0, \sigma_0^2)$$

取每个区间的中点作为代表值, 则 μ 和 σ^2 的 MLE 为

$$\hat{\mu}_* = 221, \hat{\sigma}_*^2 = 12.33^2$$

检验问题转化为:

$$H_0: X \sim N(221, 12.33^2)$$

为了计算检验统计量 $\tilde{K}_n^*$, 先计算 X 在每个区间中的概率:

$$\hat{p}_i^* = P(a_{i-1} \leq X \leq a_i) = \int_{a_{i-1}}^{a_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

其中 $u_i = (a_i - 221)/12.33$, 最终可以算得 $\tilde{K}_n^*$ 的具体值 $\tilde{k}_0^* = 1.33$, 拟合优度 $p(\tilde{k}_0^*) = P(\chi_3^2 > 1.33) = 0.70 > \alpha$.

5.4 列联表中的独立性和齐一性检验

独立性检验目的是检验同一总体、两种属性之间的独立性, 进行给定总次数的试验, 将问题转化为联合分布与理论分布的 (Pearson χ^2) 拟合优度检验; **齐一性检验**目的是检验若干总体分布是否相同 (齐一性), 对于各总体分别进行给定次数的试验, 如果理论分布未知则根据试验的总结果对理论分布进行估计, 将问题转化为各总体对与理论分布的拟合优度检验, 此时的检验问题不能化为 Pearson χ^2 检验问题, 本节只给出了检验方法, 省略了复杂的论证过程。

5.4.1 独立性检验

一、提出问题: 设总体内的每一个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:

	有 B	无 B
有 A	n_{11}	n_{12}
无 A	n_{21}	n_{22}

表 5.6: 频数表

这种表被称为 2×2 列联表, 现在要根据表中的资料来推断 A 和 B 属性之间是否有关。

将这一想法推广, 得出下面的一般问题: 设总体中每一个个体按照 A, B 属性分类, 属性 A, B 分别有 r, s 个水平, 记作 $A_1, \dots, A_r, B_1, \dots, B_s$. 从总体中抽取容量为 n 的随机样本, 测得第 i 个个体上指标状况为 $(A_{r_i}, B_{s_i}), i = 1, 2, \dots, n$. 要根据这些资料判断 A 和 B 两个属性是否独立。

$X = (X^{(1)}, X^{(2)})$	1	...	j	...	s	Σ
1	n_{11}	...	n_{1j}	...	n_{1s}	$n_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
i	n_{i1}	...	n_{ij}	...	n_{is}	$n_{i\cdot}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
r	n_{r1}	...	n_{rj}	...	n_{rs}	$n_{r\cdot}$
Σ	$n_{\cdot 1}$	...	$n_{\cdot j}$	...	$n_{\cdot s}$	n

表 5.7: 列联表

其中 $X^{(1)}, X^{(2)}$ 分别记同一个个体上的 A, B 指标的水平数, 要检验的假设为:

$$H_0: X^{(1)} \perp\!\!\!\perp X^{(2)}$$

二、检验方法: 记

$$P(X^{(1)} = i, X^{(2)} = j) = p_{ij}, i = 1, 2, \dots, r; j = 1, 2, \dots, s$$

如果假设成立, 则对于一切 (i, j) 有:

$$P(X^{(1)} = i, X^{(2)} = j) = P(X^{(1)} = i)P(X^{(2)} = j) = p_{i.}p_{.j}, \quad i = 1, 2, \dots, r; j = 1, 2, \dots, s$$

将 $p_{i.}, p_{.j}$ 视为参数, 满足

$$\sum_i p_{i.} = 1, \sum_j p_{.j} = 1$$

可知独立的未知参数一共有 $s + r - 2$ 个, 因此检验问题变成 5.3 小节中所讨论的理论分布含有未知参数的情形。

首先要求未知参数的 MLE, 计算方法类似例 2.2.8, 可得

$$\hat{p}_{i.}^* = \frac{n_{i.}}{n}, \hat{p}_{.j}^* = \frac{n_{.j}}{n}$$

可以算出 χ^2 统计量之值:

$$K_n^* = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n\hat{p}_{i.}^*\hat{p}_{.j}^*)^2}{n\hat{p}_{i.}^*\hat{p}_{.j}^*} = n \left(\sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_{i.}n_{.j}} - 1 \right)$$

而且

$$K_n^* \xrightarrow{\mathcal{L}} \chi_{rs-1-(r+s-2)}^2 = \chi_{(r-1)(s-1)}^2$$

转化为水平近似为 α 的检验:

$$K_n^* > \chi_{(r-1)(s-1)}^2(\alpha) \text{ 时否定 } H_0$$

检验拟合优度为

$$p(k_0^*) \approx P(\chi_{(r-1)(s-1)}^2 \geq k_0^*)$$

注 特殊地, $r = s = 2$ 时

$$K_n^* = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}$$

如果指标是连续取值, 就将其划分为合适的若干区间, 然后按照离散型的情形处理。

5.4.2 齐一性检验

一、提出问题: 设有 r 个取到有限个值的离散型总体 $X^{(1)}, \dots, X^{(r)}$, 它们可能的取值都相同: $a_1, \dots, a_s$, 且第 i 个总体 $X^{(i)}$ 取值为 a_j 的概率记作 $p_i(j)$, 要求检验假设: 这 r 个总体具有同样的分布, 即

$$H_0: p_1(j) = \dots = p_r(j), j = 1, 2, \dots, s$$

如果从第 i 个总体中抽取大小为 n_i 的样本, 其中取值为 a_j 的有 n_{ij} 个, 可得列联表:

$X \backslash a$	a_1	$\dots$	a_j	$\dots$	a_s	Σ
1	n_{11}	$\dots$	n_{1j}	$\dots$	n_{1s}	$n_{1.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
i	n_{i1}	$\dots$	n_{ij}	$\dots$	n_{is}	$n_{i.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
r	n_{r1}	$\dots$	n_{rj}	$\dots$	n_{rs}	$n_{r.}$
Σ	$n_{.1}$	$\dots$	$n_{.j}$	$\dots$	$n_{.s}$	n

表 5.8: 列联表

注意这个表中的 $n_{i.}$ 为事先选定的数, 没有随机性。

二、检验方法: 若分布完全已知, 即

$$p_1(j) = \dots = p_r(j) = p_j^0, j = 1, 2, \dots, s$$

则此时

$$K_n = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{i \cdot} p_j^0)^2}{n_{i \cdot} p_j^0} \xrightarrow{\mathcal{L}} \chi_{(s-1)r}^2$$

若分布未知, 则先对 p_j 求出其 MLE:

$$\hat{p}_j^* = \frac{n_{\cdot j}}{n}, j = 1, 2, \dots, s$$

此时

$$K_n^* = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n \hat{p}_{\cdot j}^*)^2}{n_{i \cdot} \hat{p}_{\cdot j}^*} = n \left(\sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_{i \cdot} n_{\cdot j}} - 1 \right)$$

这里不能直接套用定理 5.3.2, 因为 $n_{i \cdot}$ 是事先给定的, 并不是随机变量。但是经过复杂的论证, 可以证明若原假设成立, 那么

$$K_n^* \xrightarrow{\mathcal{L}} \chi_{(r-1)(s-1)}^2$$

5.5 其他的非参数检验方法 *

这一小节老师讲的很粗略。

5.5.1 Kormogorov 检验 (1933)

总体 $\sim F(x)$, $F(x)$ 未知, 样本 $(X_1, \dots, X_n)$, 已知一个分布 $F_0(x)$, 原假设:

$$H_0: F(x) = F_0(x)$$

首先构造经验分布函数

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}} = \frac{1}{n} \sum_{i=1}^n I_{\{X_{(i)} \leq x\}}$$

记 $D_n(x) = |F_n(x) - F(x)|$, $D_n = \sup_x |F_n(x) - F(x)|$, 如果 $H_0: F(x) = F_0(x)$ 成立, 则因为 $D_n(x) \rightarrow 0$ a.s.0, 所以

$$D_n^* \stackrel{\text{def}}{=} \sup_x |F_n(x) - F_0(x)| \rightarrow 0$$

D_n^* 较大时否定原假设 H_0 .

注 提出的 $F_0(x)$ 应当连续。

5.5.2 Smirnov 检验

F_1, F_2 是连续的分布函数, 分别抽取的样本为 $(X_1, \dots, X_{n_1})$ 和 $(Y_1, \dots, Y_{n_2})$, 原假设:

$$H_0: F_1(x) = F_2(x)$$

F_{1n_1}, F_{2n_2} 是对应的经验分布函数, 设

$$D_{n_1 n_2}^+ = \sup_x (F_{1n_1}(x) - F_{2n_2}(x)), D_{n_1 n_2} = \sup_x |F_{1n_1}(x) - F_{2n_2}(x)|$$

原假设成立时, $D_{n_1 n_2}^+ = D_{n_1 n_2}$. 二者相差较大时否定原假设。

第 6 章 贝叶斯方法简介

6.1 引言和若干基本概念

$B_1, B_2, \dots, B_n$ 是事件 A 的一个划分, 称为完备事件群, 那么

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}$$

因果互换位置, 就要用到贝叶斯公式。

$$\begin{aligned} f(x, y) &= f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x) \\ \Rightarrow f_{X|Y}(x|y) &= \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\int f(x, y)dx} = \frac{f_{Y|X}(y|x)f_X(x)}{\int f_{Y|X}(y|x)f_X(x)dx} \end{aligned}$$

点估计中认为, 参数 θ 是一个未知但确定的量, 所以用样本得到估计 (频率派):

$$m(\{|\hat{p}_n - p| \geq \varepsilon\}) \rightarrow 0$$

但贝叶斯观点下, 参数 θ 也具有随机性, 进而有一定的概率分布结构:

$$X \sim f(x, \theta) = f_{X|\theta}(x|\theta)\pi(\theta)$$

其中 $\pi(\theta)$ 称为 θ 的先验分布。

一、三种信息

1. 总体信息: 总体 X 所服从的分布信息。
2. 样本信息: 样本 $(X_1, \dots, X_n)$ 的观测值 $(x_1, \dots, x_n)$ 。
3. 先验信息: 抽样之前, 根据历史事件、客观规律等, 对事件的主观推测。

频率派只使用前两种信息进行统计推断, 而贝叶斯学派将使用三种信息。

例 6.1.1.

每天抽出 n 个产品进行检查, 关注其次品率。抽了 m 天, 得到的样本数据为 $\theta_1, \dots, \theta_m$ 。抽样之前, 主观上推断的

$$P(\theta_i = \frac{k}{n}) = p_k, \quad k = 0, 1, \dots, n$$

就为先验信息。

二、先验分布和后验分布

定义 6.1.1

参数空间 Θ 上的任一概率分布称为先验分布, 记为 $\theta \sim F^\pi(\theta)$ 或者 $\theta \sim \pi(\theta)$. 是在取得样本之前, 对 θ 的认识。

定义 6.1.2

获得样本 $(x_1, \dots, x_n)$ 之后 θ 的条件分布, 即将 (X, θ) 视为随机变量, $\theta|X=x$ 的条件分布 $\pi(\theta|x)$, 称为后验分布。

记边缘分布 $X \sim m(x)$, 则

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}$$

后验分布 $\pi(\theta|x)$ 是贝叶斯学派推断的出发点。

6.2 先验分布的确定

一、主观概率: 是指人们从经验中取得的事件发生概率的个人信念。如何得到主观概率?

1. 相对似然性;
2. 专家意见;
3. 历史资料。

二、利用先验信息确定先验分布:

1. 直方图 $\rightarrow$ 密度;
2. 相对似然性;
3. 选定先验分布的类型 $\pi(\theta, \alpha)$, α 称为超参数 (参数分布的参数), 对 α 作估计 $\hat{\alpha}$, 把 $\pi(\theta, \hat{\alpha})$ 作为先验分布。

三、无信息先验: “同等无知”, 即对 Θ 中任一点 θ 无任何偏向, 类似于均匀分布。

1. 均匀分布和广义均匀分布:

- (a). $\Theta = \{1, 2, \dots, n\}$, 离散均匀分布: $P(\theta = i) = \frac{1}{n}, i = 1, 2, \dots, n$;
- (b). $\Theta = \{\theta \in (a, b)\}$, 均匀分布 $U(a, b)$.
- (c). $\Theta = \mathbb{R}$, 广义 (improper) 先验 (均匀分布): $\pi(\theta) = 1$, 或者其他正的常数, 并没有必要满足全空间积分为 1. 因为我们只需确保后验分布

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$$

是个密度函数即可, 而上下 $\pi(\theta)$ 齐次, 所以并不影响。

2. 位置参数的无信息先验: 总体 $X \sim f(x - \theta)$, 这里的 θ 就称为位置参数 (location parameter), 取广义先验 $\pi(\theta) = 1$, 估计 $\hat{\theta} = \mathbb{E}(\theta|x)$.
3. 刻度参数的无信息先验: 总体 $X \sim \frac{1}{\sigma}\varphi(\frac{x}{\sigma})$, 这里的 σ 就称为刻度参数 (scale parameter), 一般取广义先验 $\pi(\sigma) = \frac{1}{\sigma}I_{(0, +\infty)}$.
4. 一般情况下的无信息先验: 通常取 $\pi(\theta) = \det(I(\theta))^{\frac{1}{2}}$, 其中 $I(\theta)$ 为信息函数:

$$I(\theta) = \mathbb{E}\left(\frac{\partial \ln(x, \theta)}{\partial \theta}\right)^2$$

例 6.2.1.

样本观测值 $(x_1, \dots, x_n)$, 总体 $X \sim N(\theta, 1)$, 取 $\theta \sim \pi(\theta) = 1$, 求 $\pi(\theta, x)$.

解答:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)dx} = \frac{\exp\left\{-\frac{1}{2}\sum(x_i - \theta)^2\right\}}{\int \exp\left\{-\frac{1}{2}\sum(x_i - \theta)^2\right\}dx} = \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\theta - \bar{x})^2\right\} \sim N(\bar{x}, \frac{1}{n})$$

可以取 $\mathbb{E}(\theta|x) = \bar{x}$ 作为 θ 的估计, 与经典结论相同。

四、共轭先验分布: 要求先验分布和后验分布为同一个分布族, 常见的只有四个: Beta 分布族、正态分布族、Gamma 分布族和逆 Gamma 分布族。

例 6.2.2.

总体 $X \sim B(n, \theta)$,

- (1). 给定先验分布 $\theta \sim U(0, 1)$, 则 $\pi(\theta|x)$ 为 Beta 分布。
- (2). 给定先验分布 $\theta \sim \beta(a, b)$, 则 $\pi(\theta|x)$ 也为 Beta 分布。

$$\pi(\theta, a, b) = \frac{1}{B(a, b)} \theta^{a-1} (1-\theta)^{b-1} I_{(0,1)}$$

这里计算过程省略, 直接给出结论: $\beta(x+1, n-x+1)$ 和 $\beta(x+a, n-x+b)$ 。

关于后验分布的计算, 借助广义均匀分布的思想, 我们可以只关注包含 θ 的因子部分 (称为“核”), 而系数部分可直接忽略。求后验分布的计算过程可以简化为:

1. 分别写出似然函数 $L(\theta)$ 和先验分布 $\pi(\theta)$ 的核 $\tilde{L}(\theta)$ 和 $\tilde{\pi}(\theta)$;
2. 后验分布 $\pi(\theta|x) \propto \tilde{L}(\theta) \cdot \tilde{\pi}(\theta)$
3. 添加正则化常数因子, 即 $1/\int \tilde{L}(\theta) \cdot \tilde{\pi}(\theta)d\theta$, 即可得到后验分布 $\pi(\theta|x)$ 。

例 6.2.3.

6.3 贝叶斯统计推断

6.3.1 贝叶斯估计

通过分析后验分布得到对参数的估计, 称为贝叶斯估计。

6.3.1.1 点估计

下设样本 $X = (X_1, \dots, X_n)$, 总体 $X \sim f(x, \theta)$, $\theta \sim \pi(\theta) \Rightarrow \pi(\theta|x) \propto f(x|\theta)\pi(\theta)$ 。

定义 6.3.1

使得 $\pi(\theta|x)$ 达到最大的 $\hat{\theta}$ 称为后验众数, 也称为广义似然 (GMLE), 记作 $\hat{\theta}_{MD}$; 用后验分布分布的中位数作为 θ 的估计, 称为后验中位数估计, 记作 $\hat{\theta}_{ME}$; 用后验均值作为 θ 的估计, 称为后验期望估计, 记作 $\hat{\theta}_E$ 。三者都称为 θ 的贝叶斯估计, 不引起歧义的情况下记作 $\hat{\theta}_B$ 。

$\pi(\theta|x)$ 为对称分布时 (例如正态分布), 三者相等。

例 6.3.1.

样本 $(X_1, \dots, X_n)$, 总体 $X \sim N(\theta, \sigma^2)$, σ^2 已知, 给出先验分布 $\pi(\theta) \sim N(\mu, \tau^2)$, 求 θ 的贝叶斯估计。

解答：计算可得 $\pi(\theta|x) \sim N(\mu_n(x), \eta_n^2)$ ，其中

$$\mu_n(x) = \frac{\mu\sigma^2/n + \tau^2\bar{x}}{\sigma^2/n + \tau^2}, \quad \eta_n^2 = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}$$

对称分布，所以贝叶斯估计为 $\hat{\theta}_B = \mu_n(x)$ 。

例 6.3.2.

产品检验：样本 $(X_1, \dots, X_n)$ ， $X_i = I_{A_i}$ 。取共轭先验 $\theta \sim \beta(a, b)$ ， $P(X_i = 1) = \theta$ ，求 $\hat{\theta}_{MD}$ 和 $\hat{\theta}_E$ 。

解答：已经计算过， $\pi(\theta|x) \sim \beta(x+a, n-x+b)$ ，计算可得

$$\hat{\theta}_{MD} = \frac{x+a-1}{n+a+b-2}, \quad \hat{\theta}_E = \frac{x+a}{n+a+b}$$

如果取 $a = b = 1$ ，即 $\pi(\theta) \sim U(0, 1)$ ，可以得到

$$\hat{\theta}_{MD} = \frac{x}{n}, \quad \hat{\theta}_E = \frac{x+1}{n+2}$$

前者和经典结论一致，为频数。实际上我们更信任后者，例如检测 10 次全部都不是次品，得到 $\hat{\theta}_E = 0.0833$ 而非 0，这符合次品率不会严格等于 0 的事实，而且如果检测了更多次都没有发现次品，相应的 $\hat{\theta}_E$ 也会越来越小，即次品率越来越小，也很符合事实和人们的一般印象。¹

例 6.3.3.

总体 $X \sim f(x|\theta) = e^{-(x-\theta)} I_{(\theta, +\infty)}(x)$ ， $\theta \in \mathbb{R}$ ，给出先验 $\theta \sim \pi(\theta) = \frac{1}{\pi(1+\theta^2)}$ ，求 $\hat{\theta}_{MD}$ 。

解答：这是一个极大值在边界上达到的例子，最后结果为 $\hat{\theta}_{MD} = x_{(1)}$ 。此外，另外两个贝叶斯估计无显式表达。

关于贝叶斯估计精度的相关讨论此处省略。

6.3.1.2 区间估计

6.3.2 贝叶斯检验

定义 6.3.2

对于检验问题

$$H_0: \theta \in \Theta_0 \longleftrightarrow H_1: \theta \in \Theta_1$$

在得到参数 θ 的后验分布 $\pi(\theta|\mathbf{x})$ 之后，分别计算

$$\alpha_0 = P(H_0|\mathbf{x}) = P(\theta \in \Theta_0|\mathbf{x})$$

$$\alpha_1 = P(H_1|\mathbf{x}) = P(\theta \in \Theta_1|\mathbf{x})$$

设定检验法则：当 $\alpha_0/\alpha_1 < 1$ 时拒绝 H_0 ，否则接受 H_0 。这样的检验称为贝叶斯检验。

¹归一化平均分会不会是类似的原理？有空研究一下。