

第七章、估计

§7.1 最大似然估计

- 总体 $X \sim F_\theta$. 目标: 给出 θ 的估计值 $\hat{\theta}$.
- **思想**: 大概率事件发生. **支撑**: 概率的主观置信度含义.
- **似然函数**:

$$L(\theta) = p(x_1, \theta) \cdots p(x_n, \theta),$$

其中, $p(x, \theta)$ 为总体的分布列/密度.

- 定义1.1. θ 的最大似然估计 $\hat{\theta} = \hat{\theta}(x_1, \cdots, x_n)$ 指 $L(\theta)$ 的最大值点. 即, $\hat{\theta} \in \Theta$, 且

$$L(\hat{\theta}) = \max_{\theta \in \Theta} p(x_1, \theta) \cdots p(x_n, \theta).$$

- 在似然函数 $L(\theta) = p(x_1, \theta) \cdots p(x_n, \theta)$ 中, 视数据 \vec{x} 为已知, 视参数 θ 为未知.
- 用 $\hat{\theta}(x_1, \cdots, x_n)$ 则强调计算, 用 $\hat{\theta}(X_1, \cdots, X_n)$ 则强调理论.
- $\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$ 是统计量.
- 必须有 $\hat{\theta} \in \Theta$.
- $g(\hat{\theta})$ 称为 $g(\theta)$ 的最大似然估计.
- 最大似然估计(最大值点)可能不唯一.

例1.1(正态模型/测量模型). 测试飞机最大飞行速度, 得到 $n = 15$ 个数据: x_1, \dots, x_n . 试估计飞机的最大飞行速度的均值.

- 假设飞机最大飞行速度 X 服从正态分布 $N(\mu, \sigma^2)$.
- $\theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$. μ 为待估参数, σ^2 为讨厌参数.
- 似然函数:

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

- $\hat{\mu} = \bar{x} := \frac{1}{n}(x_1 + \dots + x_n)$. 因为对每个固定的 σ^2 , 都有

$$L(\bar{x}, \sigma^2) = \max_{\mu \in \mathbb{R}} L(\mu, \sigma^2).$$

- 称 \bar{x} 为**样本均值**.

样本古典概型

- 古典概型: $\Omega = \{1, \dots, n\}$, $\mathbb{P}(\{i\}) = \frac{1}{n}$.
- 样本随机变量 ξ :

$\xi(i) = x_i, \forall i. \quad \eta(i) = i, f(i) := x_i$, 则 $\xi = f(\eta)$.

- 样本均值: ξ 的期望,

$$\mathbb{E}\xi = \mathbb{E}f(\eta) = \sum_{i=1}^n \frac{1}{n} \cdot f(i) = \sum_{i=1}^n \frac{1}{n} \cdot x_i = \bar{x}.$$

- 样本方差: ξ 的方差,

$$\mathbb{E}(\xi - \mathbb{E}\xi)^2 = \mathbb{E}(f(\eta) - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$\mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2.$$

样本古典概型(续)

- $\varphi(\mu) := \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \mathbb{E}(\xi - \mu)^2$.
- $\varphi(\mathbb{E}\xi) \leq \varphi(\mu)$:

$$\varphi(\mu) = \mathbb{E}((\xi - \mathbb{E}\xi) + (\mathbb{E}\xi - \mu))^2 = D(\xi) + (\mathbb{E}\xi - \mu)^2,$$

其中, $D(\xi) = \varphi(\mathbb{E}\xi)$, 交叉项为

$$\mathbb{E}(\xi - \mathbb{E}\xi)(\mathbb{E}\xi - \mu) = (\mathbb{E}\xi - \mu)\mathbb{E}(\xi - \mathbb{E}\xi) = 0.$$

- 进一步, 还可求 $\tau = \sigma^2$ 的最大似然估计.
- 将 $\hat{\mu} = \bar{x}$ 代入, 记 $a = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.

$$L(\bar{x}, \tau) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}$$

$$= \frac{1}{(\sqrt{2\pi})^n} \tau^{-\frac{n}{2}} e^{-\frac{n}{2} \cdot \frac{a}{\tau}} = \frac{1}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{n}{2} \left(\ln \tau + \frac{a}{\tau} \right) \right\}.$$

- $\hat{\tau}$ 即为 $\ln \tau + \frac{a}{\tau}$ 的最小值点:

$$\frac{d}{d\tau} \left(\ln \tau + \frac{a}{\tau} \right) = \frac{1}{\tau} - \frac{a}{\tau^2} = \frac{1}{\tau^2} (\tau - a) \Rightarrow \hat{\tau} = a.$$

- 称 $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ 为**样本方差**.

例1.2(次品率的估计). 某工人生产20 件产品, 检查出恰有一件为次品. 估计该工人生产的次品率.

- 总体 $X \sim B(1, p)$, $p \in [0, 1]$. 样本量: $n = 20$.

- 似然函数:

$$L(p) = p^s(1-p)^{n-s}, \quad \text{其中 } s = x_1 + \cdots + x_n.$$

- \hat{p} 也为 $\ln L(p) = s \ln p + (n-s) \ln(1-p)$ 的最大值点:

$$\frac{d}{dp} \ln L(p) = \frac{s}{p} - \frac{n-s}{1-p} \Rightarrow \hat{p} = \frac{s}{n}.$$

- $n = 20$, $s = 1$, 因此, $\hat{p} = \frac{1}{20}$.

例1.4. 总体: $X \sim U(0, \theta)$, 数据: X_1, \dots, X_n , (样本量: n).
求: θ 的最大似然估计.

- 似然函数:

$$L(\theta) = \prod_{i=1}^n \frac{1}{\theta} 1_{\{0 \leq x_i \leq \theta\}} = \frac{1}{\theta^n} 1_{\{0 \leq x_1, \dots, x_n \leq \theta\}}.$$

- 仅当 $\theta \geq \max_{1 \leq i \leq n} x_i$ 时, $L(\theta) > 0$.
- 当 $\theta \geq \max_{1 \leq i \leq n} x_i$ 时, $L(\theta) = \frac{1}{\theta^n}$, 关于 θ 单调下降.
- 从而, $\hat{\theta} = \max_{1 \leq i \leq n} x_i$, 即 $\max_{1 \leq i \leq n} X_i$.

例1.5. 总体: $X \sim \text{Exp}(\lambda)$, 样本量: n . 求: EX 的最大似然估计.

- 似然函数:

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} = \exp \{ n(\ln \lambda - \lambda \bar{x}) \}.$$

- $\hat{\lambda}$ 是 $\ln \lambda - \bar{x}$ 的最大值点:

$$\frac{d}{d\lambda} (\ln \lambda - \bar{x}) = \frac{1}{\lambda} - \bar{x} \Rightarrow \hat{\lambda} = 1/\bar{x}.$$

- $EX = 1/\lambda$, 因此, $\widehat{EX} = 1/\hat{\lambda} = \bar{x}$, 或 \bar{X} .

例1.6. 总体: $X \sim \mathcal{P}(\lambda)$, 样本量: n . 求: λ 的最大似然估计.

- 似然函数:

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{1}{\prod_{i=1}^n x_i!} \lambda^{n\bar{x}} e^{-\lambda n} = \frac{1}{\prod_{i=1}^n x_i!} e^{n(\bar{x} \ln \lambda - \lambda)}.$$

- $\hat{\lambda}$ 是 $\bar{x} \ln \lambda - \lambda$ 的最大值点:

$$\frac{d}{d\lambda} (\bar{x} \ln \lambda - \lambda) = \frac{\bar{x}}{\lambda} - 1 \Rightarrow \hat{\lambda} = \bar{x}, \text{ 或 } \bar{X}.$$

§7.2 矩估计

- **思想**: 样本矩 \approx 真实矩. **理论**: LLN.
- 总体矩: $\alpha_k = \alpha_k(\theta) = E_\theta X^k$, 是参数的函数.
- 样本矩: $a_k = \overline{X^k} = \frac{1}{n} \sum_{i=1}^n X_i^k$ 或 $\overline{x^k}$, 是统计量.
- 定义2.1:
 - (1) 称 a_k 为 α_k 的矩估计.
 - (2) 若待估量 $g(\theta) = \phi(\alpha_1, \dots, \alpha_k)$, 其中 ϕ 为连续函数, 则称 $\phi(a_1, \dots, a_k)$ 为 $g(\theta)$ 的矩估计.

例2.1 (续例1.1) 飞机最大飞行速度 $X \sim N(\mu, \sigma^2)$, 样本量: n .
求 μ, σ^2 的矩估计.

- Step 1. 将待估量写为矩的函数:

$$\mu = \alpha_1, \quad \sigma^2 = \alpha_2 - \alpha_1^2.$$

- Step 2. 求涉及到的样本矩: $a_1 = \bar{x}, a_2 = \overline{x^2}$.
- Step 3. 代入函数:

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \overline{x^2} - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

例2.4. 总体: $X \sim U(0, \theta)$. 样本量: n . 求 θ 的矩估计.

- $\alpha_1 = \frac{1}{2}\theta$, 即 $\theta = 2\alpha_1$. 故 $\hat{\theta}_1 = 2\bar{X}$ 是 θ 的矩估计.
- $\alpha_2 = \frac{1}{3}\theta^2$, 即 $\theta = \sqrt{3\alpha_2}$. 从而 $\hat{\theta}_2 = \sqrt{3\bar{X}^2}$ 也是 θ 的矩估计.
- 比较 $\hat{\theta}_1$ 与最大似然估计 $\hat{\theta} = \max_{1 \leq i \leq n} X_i$.

(1) 当 $2\bar{X} < \max_{1 \leq i \leq n} X_i$ 时, $\hat{\theta}_1$ 不合理. 但, $\hat{\theta}$ 总是合理.

(2) 期望:

$$E_{\theta}\hat{\theta}_1 = 2E\bar{X} = \theta, \quad E_{\theta}\hat{\theta} = \frac{n}{n+1}\theta \quad (\text{例3.6.7}).$$

(3) 方差:

$$\begin{aligned} \text{var}_{\theta}(\hat{\theta}_1) &= \frac{4}{n} \text{var}_{\theta}(X) = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{1}{3n} \theta^2, \\ \text{var}_{\theta} \left(\frac{n+1}{n} \hat{\theta} \right) &= \frac{(n+1)^2}{n^2} \cdot \frac{n}{(n+1)^2(n+2)} \theta^2 = \frac{1}{n(n+2)} \theta^2. \end{aligned}$$

§7.3 估计的无偏性

- 定义3.1. 若统计量 $T = T(X_1, \dots, X_n)$ 满足

$$E_{\theta}T = g(\theta), \quad \forall \theta \in \Theta.$$

则称 T 为 $g(\theta)$ 的无偏估计.

- 例3.1. 样本均值 \bar{X} 是期望 μ 的无偏估计.

$$\begin{aligned} E_{\theta}\bar{X} &= E_{\theta}\frac{1}{n}(X_1 + \dots + X_n) \\ &= \frac{1}{n}(E_{\theta}X_1 + \dots + E_{\theta}X_n) = \mu. \end{aligned}$$

例3.1(续). 方差 σ^2 的无偏估计.

- 样本方差 $\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 不是 σ^2 的无偏估计.
- $x_i - \mu = (x_i - \bar{x}) + (\bar{x} - \mu)$:

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

- $\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - (\bar{X} - \mu)^2.$
- $E_{\theta} \widehat{\sigma}^2 = \text{var}(X) - \text{var}(\bar{X}) = \sigma^2 - \frac{1}{n} \sigma^2 = \frac{n-1}{n} \sigma^2.$
- 定理3.1. 若总体方差 σ^2 存在, 则 S^2 是 σ^2 的无偏估计, 其中,

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

例3.2. 总体: $X \sim \text{Exp}(\lambda)$. 样本量: n . 寻找 λ 的无偏估计.

- 最大似然估计& 矩估计: $\hat{\lambda} = 1/\bar{X} = \frac{n}{S_n}$, 其中,

$$S_n = X_1 + \cdots + X_n \sim \Gamma(n, \lambda), \quad p_{S_n}(y) = \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y}, \quad y > 0.$$

- 于是,

$$\begin{aligned} E\hat{\lambda} &= E \frac{n}{S_n} = n \int_0^{\infty} \frac{1}{y} \cdot \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y} dy \\ &= n \frac{\lambda \Gamma(n-1)}{\Gamma(n)} \int_0^{\infty} \frac{\lambda^{n-1}}{\Gamma(n-1)} y^{n-2} e^{-\lambda y} dy = \frac{n}{n-1} \lambda. \end{aligned}$$

- $n \geq 2$ 时, $\frac{n-1}{n} \hat{\lambda}$ 为 λ 的无偏估计.
- $n = 1$ 时, $E\hat{\lambda} = \int_0^{\infty} \frac{1}{x} \times \lambda e^{-\lambda x} dx = \infty$.

§7.4 无偏估计的优良性

- 假设 $T = T(X_1, \dots, X_n)$ 是 $g(\theta)$ 的估计量, 称

$$R(\theta, T) := E_{\theta}(T - g(\theta))^2$$

为 T 的均方误差/风险函数.

- 定义4.1. 假设 $T = T(X_1, \dots, X_n)$ 是 $g(\theta)$ 的估计量. 如果
 - (1) T 是 $g(\theta)$ 的无偏估计,
 - (2) 对于 $g(\theta)$ 的任意无偏估计 $\tilde{T} = \tilde{T}(X_1, \dots, X_n)$, 都有

$$\text{var}_{\theta}(T) \leq \text{var}_{\theta}(\tilde{T}), \quad \forall \theta \in \Theta.$$

则称 T 为 $g(\theta)$ 的(一致)最小方差无偏估计(Uniformly Minimum Variance Unbiased, UMVU).

- 定义4.2. 假设统计量 $T = T(X_1, \dots, X_n)$ 满足:

对任意统计量 $\tilde{T} = \tilde{T}(X_1, \dots, X_n)$ 都有,

在 $T = t$ 的条件下, \tilde{T} 的分布与参数 θ 无关,

那么, 称 T 为充分统计量.

- 即, 在 $T = t$ 的条件下, (X_1, \dots, X_n) 的分布与参数 θ 无关.
- 定理4.1. 若 $T = T(X_1, \dots, X_n)$ 满足:

$$\prod_{i=1}^n p(x_i, \theta) = q_{\theta}(T(x_1, \dots, x_n))h(x_1, \dots, x_n),$$

则 T 是充分统计量.

例4.3. 总体: $B(1, p)$, 样本量: n . 考虑 $T = X_1 + \cdots + X_n$.

- 对 $t = 0, 1, \cdots, n$, 令

$$S_t := \{(x_1, \cdots, x_n) : x_i \in \{0, 1\}, \forall i; \text{ 且 } x_1 + \cdots + x_n = t\}.$$

- 那么, $\forall (x_1, \cdots, x_n) \in S_t$,

$$\begin{aligned} & P_p(X_1 = x_1, \cdots, X_n = x_n | T = t) \\ &= \frac{P_p(X_1 = \mathbf{x}_1, \cdots, X_n = \mathbf{x}_n)}{P_p(T = t)} = \frac{p^t (1-p)^{n-t}}{C_n^t p^t (1-p)^{n-t}} = \frac{1}{C_n^t}. \end{aligned}$$

- $\forall t$, 在 $T = t$ 的条件下, (X_1, \cdots, X_n) 服从 S_t 上的均匀分布. 该分布与 p 无关. 因此, T 是充分统计量.

例4.6. 总体: $X \sim \text{Exp}(\lambda)$, 样本量: n .

- 联合密度为

$$p_{\lambda}(x_1, \dots, x_n) = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} \cdot \mathbf{1}_{\{x_1, \dots, x_n > 0\}}.$$

- 令 $T = T(X_1, \dots, X_n) = X_1 + \dots + X_n$. 则 T 是充分统计量.
- 则 $\forall t > 0$, 在 $T = t$ 的条件下, (X_1, \dots, X_n) 服从

$$S_t = \{(x_1, \dots, x_n) : x_i > 0, \forall i; \text{ 且 } x_1 + \dots + x_n = t\}$$

上的均匀分布.

例4.4. 总体: $X \sim N(\mu, \sigma^2)$, 样本量: n .

- 联合密度为

$$\begin{aligned}\prod_{i=1}^n p(x_i, \theta) &= \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left\{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left\{-\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right\}.\end{aligned}$$

因此, $(\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)$ 是充分统计量.

- 若总体改为 $X \sim N(\mu, 1)$, 则联合密度改为

$$\begin{aligned}\prod_{i=1}^n p(x_i, \theta) &= \frac{1}{\sqrt{2\pi}^n} \exp\left\{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2}\right\} \\ &= \frac{1}{\sqrt{2\pi}^n} \exp\left\{-\frac{1}{2} \sum_{i=1}^n x_i^2\right\} \exp\left\{\mu \sum_{i=1}^n x_i\right\} \exp\left\{-\frac{n}{2} \mu^2\right\}.\end{aligned}$$

因此, \bar{X} 是充分统计量.

例4.4.(续) & 4.7. 总体: $X \sim N(\mu, 1)$, 样本量: n .

- 已有, $T_1 = \bar{X}$ 是充分统计量.
- $(\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)$ 也是充分统计量.
- $T_2 = (\bar{X}, X_1 - X_2)$ 也是充分统计量.
- $\psi_1(T_1) := T_1$ 是 μ 的无偏估计.
- 令 $\phi(T_2) = X_1 - X_2$, 则 $E_\mu \phi(T_2) = 0, \forall \mu$.
因此, $\psi_2(T_2) := T_1 + X_1 - X_2$ 也是 μ 的无偏估计.
- 用 T_2 造的无偏估计不好:

$$\begin{aligned} & \text{var}_\mu(T_1 + X_1 - X_2) \\ &= \text{var}_\mu \left(\frac{n+1}{n} X_1 - \frac{n-1}{n} X_2 + \frac{1}{n} X_3 + \cdots + \frac{1}{n} X_n \right) \\ &= \frac{(n+1)^2 + (n-1)^2 + n - 2}{n^2} = 2 + \frac{1}{n} > \frac{1}{n} = \text{var}_\mu(T_1). \end{aligned}$$

- 定义4.3. 设 $T = T(X_1, \dots, X_n)$ 为充分统计量. 若对任意 ϕ ,

$$E_{\theta}\phi(T) = 0, \forall \theta \in \Theta \text{ 可推出 } P_{\theta}(\phi(T) = 0) = 1, \forall \theta \in \Theta.$$

则称 T 为完全充分统计量.

- 定理4.2. 设 $T = T(X_1, \dots, X_n)$ 为完全充分统计量. 若

$$E_{\theta}\phi(T) = g(\theta), \forall \theta,$$

则 $\phi(T)$ 是 $g(\theta)$ 的UMVU 估计.

- 定义4.4. 若密度或分布列 $p(x, \theta)$ 能进行如下分解:

$$p(x, \theta) = S(\theta)h(x) \exp \left\{ \sum_{k=1}^m C_k(\theta)T_k(x) \right\},$$

则称 $p(x, \theta), \theta \in \Theta$ 为指数族分布.

- 注: x 可为高维向量, 于是 $p(x, \theta)$ 为联合密度/联合分布列.
- 引理4.1. 若总体 X 是指数族, 则样本 (X_1, \dots, X_n) 也是.
- 定理4.3. 总体分布如上; $\Theta \in \mathbb{R}^m$ 且含内点; (C_1, \dots, C_m) 在 Θ 上一对一、连续; 诸 C_i 之间(T_i 之间)无线性关系. 则

$$\left(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i) \right)$$

是完全充分统计量.

例4.9 & 4.14. 总体: $X \sim N(\mu, \sigma^2)$, 样本量: n .

- $\theta = (\mu, \sigma^2)$. $m = 2$:

$$p(x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}} e^{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2}.$$

- $T_1 = \sum_{i=1}^n X_i$, $T_2 = \sum_{i=1}^n X_i^2$, 则 (T_1, T_2) 是完全充分统计量.
- \bar{X} , S^2 是 μ , σ^2 的UMVU 估计:

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

是 (T_1, T_2) 的函数, 且是 μ , σ^2 的无偏估计.

例4.9 & 4.14(续).

- 改为已知 μ (例如, 已知 $\mu = 1$). 则 $\theta = \sigma^2$, $m = 1$:

$$p(x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-1)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-1)^2}.$$

- $T_1 = \sum_{i=1}^n (X_i - \mu)^2$ 是完全充分统计量.
- $\hat{\sigma}^2$ 是 σ^2 的UMVU 估计, 其中

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

例4.15. 总体: $X \sim N(\mu, 1)$, 样本量: n , 求 μ^2 的UMVU 估计.

- $m = 1, \theta = \mu$:

$$p(x, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu^2} e^{-\frac{1}{2}x^2} e^{\mu x}.$$

- $T_1 = \sum_{i=1}^n X_i$ 是完全充分统计量, 因此 \bar{X} 也是.
- 由 $\text{var}(Y) = EY^2 - (EY)^2$ 知,

$$\mu^2 = (E_{\mu} \bar{X})^2 = E_{\mu} \bar{X}^2 - \text{var}_{\mu}(\bar{X}) = E_{\mu} \bar{X}^2 - \frac{1}{n} = E_{\mu} \left(\bar{X}^2 - \frac{1}{n} \right).$$

- 因此, $\bar{X}^2 - \frac{1}{n}$ 是 μ^2 的UMVU 估计.

例4.10 & 4.13. 某工人生产20 件产品, 其中1 件为次品.
求: 次品率的UMVU 估计.

- 总体: $Y \sim B(1, p)$, $p = \theta \in [0, 1]$, 样本量: $n = 20$.
- 分布列: (记 $k = y_1 + \cdots + y_n$)

$$\begin{aligned} P_p(Y_1 = y_1, \cdots, Y_n = y_n) &= \prod_{i=1}^n p^{y_i} (1-p)^{1-y_i} \\ &= p^k (1-p)^{n-k} = e^{k \log p + (n-k) \log(1-p)} = e^{n \log(1-p)} e^{(\log p - \log(1-p))k}. \end{aligned}$$

- 或者, 总体: $X \sim B(20, p)$, $p = \theta \in [0, 1]$, 样本量: 1. 分布列:

$$P_p(X = k) = C_n^k p^k (1-p)^{n-k} = e^{n \log(1-p)} C_n^k e^{(\log p - \log(1-p))k}.$$

- $T_1 = X = Y_1 + \cdots + Y_{20}$ 是完全充分统计量.
- 因此, $\hat{p} = X/20$ 是UMVU 估计.

§7.5 估计的相合性

- 定义5.1. 设 $T_n = T_n(X_1, \dots, X_n)$ 满足: $\forall \varepsilon > 0$,

$$P_\theta(|T_n - g(\theta)| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0, \quad \forall \theta \in \Theta.$$

则称 T_n 为 $g(\theta)$ 的相合估计, 或 T_n 具有相合性.

- 定理5.1, 推论5.1.

样本矩 $a_\ell = \overline{X^\ell}$ 为总体矩 $\alpha_\ell = EX^\ell$ 的相合估计.

- 定理5.2. $g(\theta)$ 的矩估计具有相合性.

(注: $g(\theta) = \phi(\alpha_1, \dots, \alpha_k)$, 其中 ϕ 为连续函数).

例5.2. 总体: $X \sim U(0, \theta)$, 样本量: n .

- 矩估计 $2\bar{X}$ 具有相合性.
- 最大似然估计 $T_n = \max_{1 \leq i \leq n} X_i$ 也具有相合性: $\forall 0 < \varepsilon < \theta$,

$$P_{\theta}(|T_n - \theta| \geq \varepsilon) = P_{\theta}(T_n \leq \theta - \varepsilon)$$

$$= P_{\theta}(X \leq \theta - \varepsilon)^n = \left(\frac{\theta - \varepsilon}{\theta}\right)^n \rightarrow 0.$$

§7.6 估计的渐近分布

- 定义6.2. 设 $T_n = T_n(X_1, \dots, X_n)$ 满足:

$$\sqrt{n}(T_n - g(\theta)) \xrightarrow{d} Z \sim N(0, \sigma^2), \quad \forall \theta \in \Theta,$$

则称 T_n 是渐近正态的, 其中 $\sigma^2 = \sigma_\theta^2$ 称为渐进方差.

- 工具: CLT & Δ 方法.
- 定理6.3 (Δ 方法). 设 $\sqrt{n}(T_n - \theta) \xrightarrow{d} Z \sim N(0, \tau^2)$, $h'(\theta)$ 存在且不为0, 则

$$\sqrt{n}(h(T_n) - h(\theta)) \xrightarrow{d} W \sim N(0, h'(\theta)^2 \tau^2).$$

例6.1. 总体: $X \sim N(\mu, \sigma^2)$, 样本量: n .

- UMVU 估计: $\hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.
- $\hat{\mu}$ 渐近正态: 事实上,

$$\sqrt{n}(\bar{X} - \mu) \sim N(0, \sigma^2).$$

- 定理7.1. $(n-1)S^2 \stackrel{d}{=} \sigma^2 K_{n-1}$, 其中, $K_{n-1} \sim \chi^2(n-1)$.
- S^2 渐近正态: CLT,

$$\frac{\sum_{i=1}^{n-1} Z_i^2 - (n-1)}{\sqrt{n-1}} \xrightarrow{d} W \sim N(0, \text{var}(Z^2))$$

$$\Rightarrow \sqrt{n}(S^2 - \sigma^2) \approx \sqrt{n-1}(S^2 - \sigma^2) \xrightarrow{d} \sigma^2 W \sim N(0, 2\sigma^4).$$

例6.3. 总体: $X \sim N(\mu, 1)$, 待估量: $g(\mu) = P_\mu(X \leq x_0)$.

- 方法一、 $g(\mu) = P_\mu(X - \mu \leq x_0 - \mu) = \Phi(x_0 - \mu)$.
- 由CLT, μ 的最大似然估计 $\hat{\mu} = \bar{X}$ 渐近正态, 渐近方差= 1.
- 再由 Δ 方法, $g(\mu)$ 的最大似然估计 $g(\hat{\mu}) = \Phi(x_0 - \bar{X})$ 渐近正态, 渐近方差为

$$\sigma_1^2 = g'(\mu)^2 \cdot 1 = \varphi(x_0 - \mu)^2.$$

- 方法二、 $\frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x_0\}} \xrightarrow{P_\mu} P_\mu(X_i \leq x_0) = g(\mu)$, 渐近正态, 渐近方差为

$$\sigma_2^2 = \text{var}(1_{\{X \leq x_0\}}) = g(\mu)(1-g(\mu)) = \Phi(x_0 - \mu)(1 - \Phi(x_0 - \mu)).$$

- 习题七、30: $\sigma_1^2 \leq \sigma_2^2$.

例6.3 (续). 总体: $X \sim N(\mu, 1)$, 待估量: $g(\mu) = \Phi(x_0 - \mu)$.

- $\hat{\mu} = \bar{X}$ 是完全充分统计量, 但 $g(\hat{\mu})$ 不是 $g(\mu)$ 的无偏估计.
- 令 $h(\mu) = \Phi\left(\sqrt{\frac{n}{n-1}}(x_0 - \mu)\right)$.
- 记 $p_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, 则

$$\begin{aligned} E_{\mu} h(\hat{\mu}) &= E_{\mu} \Phi\left(\sqrt{\frac{n}{n-1}}(x_0 - \bar{X})\right) \\ &= \int_{-\infty}^{\infty} p_{\mu, \frac{1}{n}}(y) \Phi\left(\sqrt{\frac{n}{n-1}}(x_0 - y)\right) dy \\ &= \int_{-\infty}^{\infty} p_{\mu, \frac{1}{n}}(y) \int_{-\infty}^{\sqrt{\frac{n}{n-1}}(x_0 - y)} p_{0,1}(z) dz dy \\ &= P\left(Z \leq \sqrt{\frac{n}{n-1}}(x_0 - Y)\right), \end{aligned}$$

其中, Y, Z 相互独立, $Y \sim N(\mu, \frac{1}{n})$, $Z \sim N(0, 1)$.

例6.3 (续). 总体: $X \sim N(\mu, 1)$, 待估量: $g(\mu) = \Phi(x_0 - \mu)$.

- 已有: 取 Y, Z 相互独立, $Y \sim N(\mu, \frac{1}{n})$, $Z \sim N(0, 1)$, 则

$$E_{\mu}h(\hat{\mu}) = P\left(Z \leq \sqrt{\frac{n}{n-1}}(x_0 - Y)\right),$$

- $\sqrt{\frac{n-1}{n}}Z + Y \sim N(\mu, 1)$. 因此,

$$E_{\mu}h(\hat{\mu}) = P\left(\sqrt{\frac{n-1}{n}}Z + Y - \mu \leq x_0 - \mu\right) = \Phi(x_0 - \mu).$$

- $\hat{\mu}$ 是完全充分统计量, $h(\hat{\mu})$ 是 $g(\mu)$ 的无偏估计, 因此是UMVU 估计.

§7.7 置信区间和置信限

定义7.1. 假设 $\underline{T} = \underline{T}(X_1, \dots, X_n)$ 与 $\bar{T} = \bar{T}(X_1, \dots, X_n)$ 为统计量, $\alpha \in (0, 1)$.

(1) 若 $\underline{T} < \bar{T}$ 且

$$P_{\theta}(\underline{T} \leq g(\theta) \leq \bar{T}) \geq 1 - \alpha, \quad \forall \theta \in \Theta,$$

则称 $[\underline{T}, \bar{T}]$ 为 $g(\theta)$ 的置信度为 $1 - \alpha$ 的置信区间.

(2) 若

$$P_{\theta}(\underline{T} \leq g(\theta)) \geq 1 - \alpha, \quad \forall \theta \in \Theta,$$

则称 \underline{T} 为 $g(\theta)$ 的置信度为 $1 - \alpha$ 的置信下限.

(3) 若

$$P_{\theta}(g(\theta) \leq \bar{T}) \geq 1 - \alpha, \quad \forall \theta \in \Theta,$$

则称 \bar{T} 为 $g(\theta)$ 的置信度为 $1 - \alpha$ 的置信上限.

1. 枢轴量法

- 定义7.2. 假设 $g(\theta)$ 是待估量. 若

$$h = h(X_1, \dots, X_n; g(\theta))$$

的分布与 θ 无关, 则称 h 为枢轴量.

- Step 1. 找枢轴量 $h = h(\vec{X}, g(\theta))$ 及其分布 F .
- Step 2. 利用 F 选择 a, b , 使得:

$$P(a \leq h \leq b) \geq 1 - \alpha.$$

- Step 3. 将 $a \leq h \leq b$ 化为 $\underline{T} \leq g(\theta) \leq \bar{T}$, 于是得到

$$P(\underline{T} \leq g(\theta) \leq \bar{T}) \geq 1 - \alpha.$$

1. 枢轴量法

例7.1. 总体: $X \sim \text{Exp}(\lambda)$. 样本量: n . 求 λ 的置信区间.

- $\lambda X \sim \text{Exp}(1)$, 因此,

$$h_1 = \lambda(X_1 + \cdots + X_n) \sim \Gamma(n, 1).$$

- $2\lambda X \sim \text{Exp}\left(\frac{1}{2}\right)$, 因此

$$h_2 = 2\lambda(X_1 + \cdots + X_n) \sim \Gamma\left(n, \frac{1}{2}\right) = \chi^2(2n).$$

- 查 $\chi^2(2n)$ 的表获得 $\lambda_1 = \chi_{\alpha/2}^2(2n)$, $\lambda_2 = \chi_{1-\alpha/2}^2(2n)$. 于是, $P_\lambda(\lambda_1 \leq h_2 \leq \lambda_2) = 1 - \alpha$. 从而, 所求为 $[\underline{T}, \bar{T}]$, 其中,

$$\underline{T} = \frac{\lambda_1}{2(X_1 + \cdots + X_n)}, \quad \bar{T} = \frac{\lambda_2}{2(X_1 + \cdots + X_n)}.$$

2. 正态总体 $X \sim N(\mu, \sigma^2)$ 中参数的置信区间

- 定义3.6.8 & 7.3. n 维正态分布 $N(\vec{\mu}, \Sigma)$ 的联合密度为:

$$p(\vec{x}) = \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma|}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu}) \right\}.$$

其中, $\vec{x} = (x_1, \dots, x_n)^T$, $\mu = (\mu_1, \dots, \mu_n)^T$, $\Sigma = (\sigma_{ij})_{n \times n}$ 正定.

- 定义3.6.6, 3.6.7 & 7.4.

$\vec{X} = (X_1, \dots, X_n)^T$ 的期望与协方差阵分别指

$$E\vec{X} = (EX_1, \dots, EX_n)^T, \quad \text{cov}(\vec{X}, \vec{X}) = (\text{cov}(X_i, X_j))_{n \times n}.$$

\vec{X} 与 $\vec{Y} = (Y_1, \dots, Y_m)^T$ 的协方差阵指

$$\text{cov}(\vec{X}, \vec{Y}) = \begin{pmatrix} \text{cov}(X_1, Y_1) & \cdots & \text{cov}(X_1, Y_m) \\ \text{cov}(X_n, Y_1) & \cdots & \text{cov}(X_n, Y_m) \end{pmatrix}_{n \times m}.$$

● 引理7.1. 矩阵运算, 例: $\text{cov}(\mathbf{A}\vec{X}, \vec{Y}) = \mathbf{A}\text{cov}(\vec{X}, \vec{Y})$.

● 引理7.2. 假设 $\vec{X} \sim N(\vec{\mu}, \Sigma)$.

(1) $E\vec{X} = \vec{\mu}$, $\text{cov}(\vec{X}, \vec{X}) = \Sigma$;

(2) 设 $\mathbf{B}_{n \times n}$ 非退化, $\vec{a} \in \mathbb{R}^n$, 则 $\vec{a} + \mathbf{B}\vec{X}$ 服从 n 维正态分布;

(3) 设 $1 \leq m < n$, 则 $(X_1, \dots, X_m)^T$ 服从 m 维正态分布;

(4) 设 $1 \leq m < n$, $\mathbf{A}_{m \times n}$ 行满秩, 则 $\mathbf{A}\vec{X}$ 服从 m 维正态分布.

• 定理7.1. 假设总体: $X \sim N(\mu, \sigma^2)$, 样本量: n . 则

(1) $\bar{X} \sim N\left(\mu, \frac{1}{n}\sigma^2\right)$;

(2) $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$;

(3) \bar{X} 与 $\sum_{i=1}^n (X_i - \bar{X})^2$ 相互独立.

• 证: $X_i = \mu + \sigma Z_i$, 其中 $Z_i = X_i^* \sim N(0, 1)$, i.i.d..

• $\bar{X} = \mu + \sigma \bar{Z}$, $\star = \sum_{i=1}^n (Z_i - \bar{Z})^2 = \sum_{i=1}^n Z_i^2 - n\bar{Z}^2$.

• 取正交矩阵 $\mathbf{A}_{n \times n}$, 其第一行是 $\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$. 令 $\vec{Y} = \mathbf{A}\vec{Z}$.

• 由 \mathbf{A} 正交, $\vec{Y} \sim N(\vec{0}, \mathbf{I}_{n \times n})$ 且 $\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n Y_i^2$.

由 \mathbf{A} 的第一行, $Y_1^2 = n\bar{Z}^2$. 于是, $\star = \sum_{i=2}^n Y_i^2 \sim \chi^2(n-1)$.

• $\bar{Z} = \frac{1}{\sqrt{n}}Y_1 \sim N\left(0, \frac{1}{n}\right)$, 且与 $\sum_{i=2}^n Y_i^2$ 独立. 故, (1), (3) 成立.

例7.2 & 7.3. σ^2 已知, (例如, $X \sim N(\mu, 1)$).

求: μ 的置信度为 $1 - \alpha$ 的(1) 置信区间, (2) 置信上限.

- 取 $h = h(X_1, \dots, X_n, \mu) := \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$.
- (1) 查表获得 $z_{1-\alpha/2}$, 于是 $P_\mu(|h| \leq z_{1-\alpha/2}) = 1 - \alpha$. 因此,

$$P_\mu \left(|\bar{X} - \mu| \leq \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right) = 1 - \alpha.$$

- 概率论角度: $\bar{X} \in [\mu - \frac{\sigma}{\sqrt{n}}x, \mu + \frac{\sigma}{\sqrt{n}}x]$,
未知的随机点 \bar{X} 落在已知的确定区间中.
- 统计学角度: $\mu \in [\bar{X} - \frac{\sigma}{\sqrt{n}}x, \bar{X} + \frac{\sigma}{\sqrt{n}}x]$ (此即所求置信区间),
已知的随机区间(可由数据得到)覆盖未知参数 μ (确定的点).
- (2) 置信上限为 $\bar{\mu} = \bar{X} + \frac{\sigma}{\sqrt{n}}z_{1-\alpha}$:

$$P_\mu(h \geq z_\alpha) = 1 - \alpha \Rightarrow P_\mu \left(\bar{X} \leq \mu - \frac{\sigma}{\sqrt{n}}z_\alpha \right) = 1 - \alpha.$$

例7.4. σ^2 未知. 求: μ 的置信度为 $1 - \alpha$ 的置信区间.

- $\tilde{h} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$ 仍然成立.

- 但是, \tilde{h} 不是枢轴量!

枢轴量只能含数据和待估量, 不能含讨厌参数 σ^2 .

- 例7.2中的 $\bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$ 含 σ^2 , 不是统计量.

- 用 σ^2 的UMVU 估计 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 代替 σ^2 .

将证明

$$h = h(X_1, \dots, X_n; \mu) := \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{S^2}}$$

是枢轴量.

- 定理7.1 的证明: $X_i = \mu + \sigma Z_i$. 正矩阵 $\mathbf{A}_{n \times n}$ 的第一行是 $(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$. 令 $\vec{Y} = \mathbf{A}\vec{Z}$, 则 $Y_1, \dots, Y_n \sim \text{i.i.d. } N(0, 1)$.
- 考察

$$h = \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{S^2}}, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

- 考察分子: $\sqrt{n}(\bar{X} - \mu) = \sigma\sqrt{n}\bar{Z} = \sigma Y_1$.
- 考察分母: $\sum_{i=1}^n (X_i - \bar{X})^2 = \sigma^2 \sum_{i=2}^n Y_i^2$. 于是,

$$h = \frac{Y_1}{\sqrt{\frac{1}{n-1} \sum_{i=2}^n Y_i^2}}.$$

- 自由度为 n 的 t 分布(记为 $t(n)$) 指的是

$$T_n := \frac{Z}{\sqrt{\frac{1}{n}(Z_1^2 + \dots + Z_n^2)}}$$

服从的分布, 其中 Z, Z_1, \dots, Z_n 独立同分布, $Z \sim N(0, 1)$.

- 因此,

$$h = h(X_1, \dots, X_n; \mu) = \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{S^2}} \sim t(n-1),$$

其中

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

为 σ^2 的UMVU 估计.

- 查表获得 $t_{1-\alpha/2}(n-1)$. 因此

$$P_{\mu} \left(|\bar{X} - \mu| \leq \frac{\hat{\sigma}}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right) = 1 - \alpha.$$

- 所求置信区间为

$$\left[\bar{X} - \frac{\hat{\sigma}}{\sqrt{n}} t_{1-\alpha/2}(n-1), \bar{X} + \frac{\hat{\sigma}}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right].$$

例7.5. $\theta = (\mu, \sigma^2)$. 求 σ^2 的置信度为 $1 - \alpha$ 的置信上限.

- 枢轴量: 由定理7.1,

$$h := \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1).$$

- 查表获得 $\chi_\alpha^2(n-1)$, 于是

$$P_\theta (h \geq \chi_\alpha^2(n-1)) = 1 - \alpha.$$

因此,

$$P_\theta \left(\sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_\alpha^2(n-1)} \right) = 1 - \alpha.$$

- 所求置信上限为 $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_\alpha^2(n-1)}$.

- 注: 若 μ 已知, 则枢轴量和置信上限应为

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2(n), \quad \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_\alpha^2(n)}.$$

3. 参数的近似置信区间

- 定理7.2. 假设 $T_n = T_n(X_1, \dots, X_n)$ 满足

$$\sqrt{n}(T_n - g(\theta)) \xrightarrow{d} N(0, \sigma^2),$$

则 $g(\theta)$ 的近似置信区间估计为:

(1) σ^2 已知: $[T_n - \frac{\sigma}{\sqrt{n}}z_{1-\alpha/2}, T_n + \frac{\sigma}{\sqrt{n}}z_{1-\alpha/2}]$,

(2) σ^2 未知: $[T_n - \frac{\hat{\sigma}_n}{\sqrt{n}}z_{1-\alpha/2}, T_n + \frac{\hat{\sigma}_n}{\sqrt{n}}z_{1-\alpha/2}]$, 其中 $\hat{\sigma}_n$ 为 σ 的相合估计.

例7.6. 90 人中15 人反应课业负担重, 求: 负担重的百分比 θ 的0.95 置信区间.

- 总体: $X \sim B(1, \theta)$, 样本量: $n = 90$.
- 由CLT,

$$\frac{\sqrt{n}(\bar{X} - \theta)}{\sqrt{\theta(1 - \theta)}} \xrightarrow{d} Z \sim N(0, 1).$$

于是,

$$P_{\theta} \left(\left| \frac{\sqrt{n}(\bar{X} - \theta)}{\sqrt{\theta(1 - \theta)}} \right| \leq z_{1-\alpha/2} \right) \approx 1 - \alpha.$$

- 查表获得 $z_{0.975} = 1.96$, 于是

$$\left| \frac{\sqrt{90}(\frac{15}{90} - \theta)}{\sqrt{\theta(1 - \theta)}} \right| \leq 1.96 \Rightarrow \theta \in [0.1037, 0.2569],$$

此即所求的近似置信区间.