

Lec-21. 点估计: 矩估计和最大似然估计

主讲教师: 吴利苏 (wulisu@sdust.edu.cn)

主 页: wulisu.cn

本次课内容

点估计

- 矩估计
- 最大似然估计

总体的分布函数形式已知，但参数未知.

- 点估计:

样本统计量 $\xrightarrow{\text{估计}}$ 总体未知参数

- 参数是指反映总体某方面特征的量，比如：合格率, 均值, 方差, 中位数...
- 天气预报明天的最高温度: 15°C . —点估计.
- 明天的温度: 9°C - 15°C . —区间估计 (下周介绍)

点估计

设总体 X 有未知参数 θ , X_1, \dots, X_n 是总体 X 的简单随机样本.

- **点估计量**: 构造合适的统计量

$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$$

用来估计未知参数 θ , 则 $\hat{\theta}$ 称为参数 θ 的点估计量.

- **点估计值**: $\hat{\theta}$ 的观察值 $\hat{\theta}(x_1, \dots, x_n)$ 称为参数 θ 的点估计值.
- 常用的点估计: **矩估计**、**最大似然估计**.

例

某校大二学生有 4 千人参加第二学期的《概率论》考试. 现随机选出 100 名学生, 计算得他们的平均成绩为 72.3 分, 标准差为 15.8 分. 试估计全部学生的平均成绩.

解: 设总体 (4000 个学生成绩) 的均值为 μ , 则 μ 的估计值为 72.3 分. □

例

某校大二学生有 4 千人参加第二学期的《概率论》考试. 现随机选出 100 名学生, 计算得他们的平均成绩为 72.3 分, 标准差为 15.8 分. 试估计全部学生的平均成绩.

解: 设总体 (4000 个学生成绩) 的均值为 μ , 则 μ 的估计值为 72.3 分. □

- μ : 总体均值 (总体的一阶矩)
- 72.3 分: 样本均值 (样本的一阶矩) 的观测值

例

某校大二学生有 4 千人参加第二学期的《概率论》考试. 现随机选出 100 名学生, 计算得他们的平均成绩为 72.3 分, 标准差为 15.8 分. 试估计全部学生的平均成绩.

解: 设总体 (4000 个学生成绩) 的均值为 μ , 则 μ 的估计值为 72.3 分. □

- μ : 总体均值 (总体的一阶矩)
- 72.3 分: 样本均值 (样本的一阶矩) 的观测值

♠ 用样本矩 作为 总体矩 的估计即为矩估计.

记号: 样本矩 $\xrightarrow{\text{估计}}$ 总体矩

样本矩

总体矩

原点矩

$$A_j = \frac{1}{n} \sum_{i=1}^n X_i^j$$

$$\mu_j = E(X^j)$$

中心矩

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

$$\nu_k = E([X - E(X)]^k)$$

- $A_1 = \bar{X}$, $B_2 = A_2 - A_1^2$;
- $\mu_1 = E(X)$, $\nu_2 = \mu_2 - \mu_1^2$,

$$\text{即 } D(X) = E(X^2) - (E(X))^2.$$

矩估计法

- 统计思想:

样本矩 $\xrightarrow{\text{估计}}$ 总体矩

样本矩的函数 $\xrightarrow{\text{估计}}$ 总体矩的函数

- 理论根据: 辛钦大数定律和依概率收敛的性质.

矩估计法

假设总体矩 $\mu_j = E(X^j)$ 存在, $j = 1, \dots, k$.
则由辛钦大数定律

$$\hat{\mu}_j = A_j = \frac{1}{n} \sum_{i=1}^n X_i^j \xrightarrow{P} \mu_j, j = 1, \dots, k$$

进一步设 h 为连续函数, 由依概率收敛的性质

$$h(\hat{\mu}_1, \dots, \hat{\mu}_k) = h(A_1, \dots, A_k) \xrightarrow{P} h(\mu_1, \dots, \mu_k).$$

矩估计步骤

设总体 X 有 k 个未知参数 $\theta_1, \dots, \theta_k$, 且前 k 阶矩存在. X_1, \dots, X_n 是来自总体 X 的样本.

矩估计步骤:

矩估计步骤

设总体 X 有 k 个未知参数 $\theta_1, \dots, \theta_k$, 且前 k 阶矩存在. X_1, \dots, X_n 是来自总体 X 的样本.

矩估计步骤:

- (1)** 建立 $(\theta_1, \dots, \theta_k)$ 与 (μ_1, \dots, μ_k) 的联系:
求总体前 k 阶矩关于 k 个参数的函数,

$$\mu_i = E(X^i) = h_i(\theta_1, \dots, \theta_k), \quad i = 1, \dots, k.$$

矩估计步骤

设总体 X 有 k 个未知参数 $\theta_1, \dots, \theta_k$, 且前 k 阶矩存在. X_1, \dots, X_n 是来自总体 X 的样本.

矩估计步骤:

- (1)** 建立 $(\theta_1, \dots, \theta_k)$ 与 (μ_1, \dots, μ_k) 的联系:
求总体前 k 阶矩关于 k 个参数的函数,

$$\mu_i = E(X^i) = h_i(\theta_1, \dots, \theta_k), \quad i = 1, \dots, k.$$

- (2)** 求各参数关于各阶总体矩的反函数,

$$\theta_i = g_i(\mu_1, \dots, \mu_k), \quad i = 1, \dots, k.$$

矩估计步骤

- (1) 建立 $(\theta_1, \dots, \theta_k)$ 与 (μ_1, \dots, μ_k) 的联系：
求总体前 k 阶矩关于 k 个参数的函数，

$$\mu_i = E(X^i) = h_i(\theta_1, \dots, \theta_k), \quad i = 1, \dots, k.$$

- (2) 求各参数关于各阶总体矩的反函数，

$$\theta_i = g_i(\mu_1, \dots, \mu_k), \quad i = 1, \dots, k.$$

- (3) 以样本各阶矩 A_1, \dots, A_k 代替总体 X 各阶矩 μ_1, \dots, μ_k ，得各参数的矩估计

$$\hat{\theta}_i = g_i(A_1, \dots, A_k), \quad i = 1, \dots, k.$$

在实际应用时，为求解方便，也可以样本中心矩 B_i 估计总体中心矩 ν_i . ($E(X) = \mu$)

$$\hat{\nu}_j = B_j = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^j \xrightarrow{P} \nu_j, j = 1, \dots, k$$

采用的矩不同，得出的矩估计也可能不同。

例

4 千人中随机选出 100 名学生, 平均成绩为 72.3 分, 标准差为 15.8 分. 试求总体标准差 σ 的矩估计值.

例

4千人中随机选出100名学生，平均成绩为72.3分，标准差为15.8分。试求总体标准差 σ 的矩估计值。

解： $\mu_1 = E(X) = \mu$, $\mu_2 = E(X^2) = \mu^2 + \sigma^2$,

则 $\mu = \mu_1$, $\sigma = \sqrt{\mu_2 - \mu_1^2}$. 因此,

$$\begin{cases} \hat{\mu} = A_1 = \bar{X} = 72.3, \\ \hat{\sigma} = \sqrt{A_2 - \bar{X}^2} = \sqrt{B_2} = \sqrt{\frac{99}{100} \times 15.8^2} = 15.7. \end{cases}$$

其中 $B_2 = \frac{n-1}{n} S^2$. □

例

4千人中随机选出100名学生，平均成绩为72.3分，标准差为15.8分。试求总体标准差 σ 的矩估计值。

解： $\mu_1 = E(X) = \mu$, $\mu_2 = E(X^2) = \mu^2 + \sigma^2$,

则 $\mu = \mu_1$, $\sigma = \sqrt{\mu_2 - \mu_1^2}$. 因此,

$$\begin{cases} \hat{\mu} = A_1 = \bar{X} = 72.3, \\ \hat{\sigma} = \sqrt{A_2 - \bar{X}^2} = \sqrt{B_2} = \sqrt{\frac{99}{100} \times 15.8^2} = 15.7. \end{cases}$$

其中 $B_2 = \frac{n-1}{n} S^2$. □

注：矩估计不涉及总体分布。

例

设总体 $X \sim b(1, p)$, p 未知, X_1, \dots, X_n 是总体 X 的样本. 求 p 的矩估计量.

$$\text{解: } \mu_1 = EX = p,$$

$$\therefore p = \mu_1,$$

$$\therefore \hat{p} = \bar{X}.$$



即用样本比例来估计总体比例.

例 (标记重捕法)

一个很大的罐子里装满了糖，如何估计糖的数目 n ?

解：从罐子里取 k 颗糖，做上记号，再放回罐子中，然后有放回取 m 颗. 设取到做记号的糖数为 k_1 . 则带记号的糖的总体比例为 $\frac{k}{n}$, 样本比例为 $\frac{k_1}{m}$.

$$\therefore \frac{k}{\hat{n}} = \frac{k_1}{m} \Rightarrow \hat{n} = k \frac{m}{k_1}.$$

□

类似方法可以估计池塘里鱼的数目，森林里某动物的数目等.

例

设总体 X 的密度为：

$$f(x; \theta) = \begin{cases} \sqrt{\theta} x^{\sqrt{\theta}-1} & 0 \leq x \leq 1; \\ 0 & \text{其他.} \end{cases} \cdot \theta > 0 \text{ 未知,}$$

X_1, \dots, X_n 为样本，求 θ 的矩估计量。

若已获得 $n = 10$ 的样本值如下，

0.43 0.01 0.30 0.04 0.54

0.14 0.99 0.18 0.98 0.02

求 θ 的矩估计值。

解:(1)

$$\mu_1 = E(X) = \int_{-\infty}^{+\infty} xf(x) dx = \int_0^1 \sqrt{\theta} x^{\sqrt{\theta}} dx = \frac{\sqrt{\theta}}{\sqrt{\theta} + 1}$$

$$(2) \theta = \left(\frac{\mu_1}{1-\mu_1} \right)^2$$

(3) 矩估计量

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

$$(4) \bar{x} = 0.363, \text{矩估计值} \hat{\theta} = \left(\frac{0.363}{1-0.363} \right)^2 = 0.325. \quad \square$$

例

设总体 X 服从均匀分布 $U(a, b)$, a, b 未知.
 X_1, \dots, X_n 为样本, 求 a, b 的矩估计量.

解: (1) 求矩关于参数的函数

$$\mu_1 = E(X) = \frac{a+b}{2}, \quad \nu_2 = \frac{(b-a)^2}{12} = \frac{1}{3} \left(\frac{b-a}{2} \right)^2,$$

(2) 求参数关于矩的反函数

$$a = \mu_1 - \sqrt{3\nu_2}, \quad b = \mu_1 + \sqrt{3\nu_2}$$

(3) 以样本矩 $A_1 = \bar{X}$ 代替总体矩 μ_1 ,

$$B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ 代替 } \nu_2,$$

得参数 a 和 b 的矩估计量:

$$\hat{a} = \bar{X} - \sqrt{3B_2}, \quad \hat{b} = \bar{X} + \sqrt{3B_2}.$$

最（极）大似然估计的原理介绍

例

假设在一个罐中放着许多白球和黑球，并假定已经知道两种球的数目之比是 1:3，但不知道哪种颜色的球多。如果用放回抽样方法从罐中取 5 个球，观察结果为：黑、白、黑、黑、黑，估计取到黑球的概率 p .

解: 设 $X = \begin{cases} 1, & \text{取到黑球,} \\ 0, & \text{取到白球.} \end{cases}$ 则 $X \sim b(1, p)$.

p 为取到黑球的概率, 未知, $p = \frac{1}{4}$ 或 $\frac{3}{4}$.
抽取容量为 5 的样本 X_1, \dots, X_5 , 观测值为

1, 0, 1, 1, 1.

当 $p = \frac{1}{4}$ 时, 出现本次观察结果的概率为 $\left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right) = \frac{3}{1024}$.

当 $p = \frac{3}{4}$ 时, 出现本次观察结果的概率为 $\left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right) = \frac{81}{1024}$.

由于 $\frac{3}{1024} < \frac{81}{1024}$, 因此认为 $p = \frac{3}{4}$ 比 $p = \frac{1}{4}$ 更有可能.

于是 \hat{p} 取为 $p = \frac{3}{4}$ 更合理. □

解: 设 $X = \begin{cases} 1, & \text{取到黑球,} \\ 0, & \text{取到白球.} \end{cases}$ 则 $X \sim b(1, p)$.

p 为取到黑球的概率, 未知, $p = \frac{1}{4}$ 或 $\frac{3}{4}$.
抽取容量为 5 的样本 X_1, \dots, X_5 , 观测值为

1, 0, 1, 1, 1.

当 $p = \frac{1}{4}$ 时, 出现本次观察结果的概率为 $\left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right) = \frac{3}{1024}$.

当 $p = \frac{3}{4}$ 时, 出现本次观察结果的概率为 $\left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right) = \frac{81}{1024}$.

由于 $\frac{3}{1024} < \frac{81}{1024}$, 因此认为 $p = \frac{3}{4}$ 比 $p = \frac{1}{4}$ 更有可能.

于是 \hat{p} 取为 $p = \frac{3}{4}$ 更合理. □

分布的参数 θ 未知, 但参数的可能取值范围 Θ 已知. 估计未知参数在 Θ 中最可能的取值称为最大似然估计

最大似然估计

设离散型 总体 $X \sim P\{X = x\} \triangleq p(x; \theta), \theta \in \Theta$,
 θ 未知. X_1, \dots, X_n 为样本, 其观察值为 x_1, \dots, x_n ,
则事件 $\{X_1 = x_1, \dots, X_n = x_n\}$ 发生的概率为

- 似然函数: $L(\theta) = \prod_{i=1}^n p(x_i; \theta)$.
- 最大似然原理:

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

其中 $\hat{\theta}(x_1, \dots, x_n)$ 称为 θ 的**最大似然估计值**,
统计量 $\hat{\theta}(X_1, \dots, X_n)$ 称为 θ 的**最大似然估计量**.

设连续型 总体 X 概率密度为 $f(x; \theta)$, $\theta \in \Theta$, θ 未知. X_1, \dots, X_n 为样本, 则样本在观察值 (x_1, x_2, \dots, x_n) 邻域发生的概率

$$\prod_{i=1}^n P(x_i < X_i < x_i + \Delta x_i) \approx \prod_{i=1}^n f(x_i; \theta) \Delta x_i,$$

Δx_i 与参数 θ 无关. 因此,

- 似然函数: $L(\theta) = \prod_{i=1}^n f(x_i; \theta)$
- 最大似然原理:

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

注

1. 未知参数可能不是一个, 设为

$$\theta = (\theta_1, \theta_2, \dots, \theta_k);$$

注

1. 未知参数可能不是一个, 设为

$$\theta = (\theta_1, \theta_2, \dots, \theta_k);$$

2. 求 $L(\theta)$ 的最大值时, 可转换为求 $\ln L(\theta)$ 的最大值, $\ln L(\theta)$ 称为**对数似然函数**. 利用

$$\frac{\partial \ln L(\theta)}{\partial \theta_i} \Big|_{\theta=\hat{\theta}} = 0,$$

解得 $\hat{\theta}_i, i = 1, 2, \dots, k.$

注

1. 未知参数可能不是一个, 设为

$$\theta = (\theta_1, \theta_2, \dots, \theta_k);$$

2. 求 $L(\theta)$ 的最大值时, 可转换为求 $\ln L(\theta)$ 的最大值, $\ln L(\theta)$ 称为**对数似然函数**. 利用

$$\frac{\partial \ln L(\theta)}{\partial \theta_i} \Big|_{\theta=\hat{\theta}} = 0,$$

解得 $\hat{\theta}_i, i = 1, 2, \dots, k$.

3. 若 $L(\theta)$ 关于某个 θ_i 是单调增(减)函数, 则 θ_i 的最大似然估计为 θ_i 的最大(小)值(与样本有关);

注

1. 未知参数可能不是一个, 设为

$$\theta = (\theta_1, \theta_2, \dots, \theta_k);$$

2. 求 $L(\theta)$ 的最大值时, 可转换为求 $\ln L(\theta)$ 的最大值, $\ln L(\theta)$ 称为**对数似然函数**. 利用

$$\frac{\partial \ln L(\theta)}{\partial \theta_i} \Big|_{\theta=\hat{\theta}} = 0,$$

解得 $\hat{\theta}_i, i = 1, 2, \dots, k$.

3. 若 $L(\theta)$ 关于某个 θ_i 是单调增(减)函数, 则 θ_i 的最大似然估计为 θ_i 的最大(小)值(与样本有关);
4. 若 $\hat{\theta}$ 是 θ 的最大似然估计, 则 $g(\theta)$ 的最大似然估计为 $g(\hat{\theta})$.

例

设 X 的概率密度为

$$f(x; \theta) = \begin{cases} \sqrt{\theta} x^{\sqrt{\theta}-1}, & 0 \leq x \leq 1, \\ 0, & \text{其他} \end{cases}$$

X_1, \dots, X_n 是样本, 求 θ 的最大似然估计量.

若已获得 $n = 10$ 的样本值如下:

0.43	0.01	0.30	0.04	0.54
0.14	0.99	0.18	0.98	0.02

求 θ 的最大似然估计值.

$$\text{解： } L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \sqrt{\theta} x_i^{\sqrt{\theta}-1} = \theta^{\frac{n}{2}} \left(\prod_{i=1}^n x_i \right)^{\sqrt{\theta}-1}$$

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

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

$$\implies \frac{n}{\sqrt{\theta}} = - \sum_{i=1}^n \ln x_i \implies \sqrt{\theta} = -n / \sum_{i=1}^n \ln x_i$$

- 参数 θ 的最大似然估计量为：

$$\hat{\theta} = \frac{n^2}{\left(\sum_{i=1}^n \ln X_i\right)^2}.$$

- 将上面的样本值代入估计量，得 θ 的最大似然估计值为： $\hat{\theta} = 0.305$. □
- 比较矩估计量：

$$\hat{\theta} = \left(\frac{\bar{X}}{1 - \bar{X}}\right)^2.$$

例

设总体 $X \sim N(\mu, \sigma^2)$, X_1, \dots, X_n 是样本, μ, σ^2 均未知. 求 μ, σ^2 的最大似然估计.

例

设总体 $X \sim N(\mu, \sigma^2)$, X_1, \dots, X_n 是样本, μ, σ^2 均未知. 求 μ, σ^2 的最大似然估计.

解:

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

$$\ln L(\mu, \sigma^2) = n \ln \frac{1}{\sqrt{2\pi}} - \frac{n}{2} \ln \sigma^2 - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

$$\frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = \frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0$$

所以

$$\hat{\mu} = \bar{X}$$

$$\hat{\sigma}^2 = B_2$$

例

设 X 服从均匀分布 $U(a, b)$, a 和 b 未知, 样本 X_1, \dots, X_n

- (1) 求 a 和 b 的最大似然估计.
- (2) 求 $E(X)$ 的最大似然估计.
- (3) 若已获得 $n = 5$ 的样本值如下,
0.34 0.59 0.16 0.96 0.84
求 $a, b, E(X)$ 的最大似然估计值.

解：(1) 似然函数

$$L(a, b) = \begin{cases} \frac{1}{(b-a)^n}, & a \leq x_i \leq b, i = 1, \dots, n. \\ 0, & \text{其他.} \end{cases}$$

关于 a 单调增，关于 b 单调减. 并且在得到样本值 x_1, \dots, x_n 后，只有当 a 的取值 $\leq \min \{x_1, \dots, x_n\}$ ， b 的取值 $\geq \max \{x_1, \dots, x_n\}$ 时，才能使似然函数 $L(a, b)$ 不为零.

因此, a 达到最大值 $\min\{x_1, \dots, x_n\}$, b 达到最小值 $\max\{x_1, \dots, x_n\}$, 就能使 $L(\alpha, b)$ 达到最大.
所以

$$\hat{a} = \min\{X_1, \dots, X_n\} = X_{(1)},$$
$$\hat{b} = \max\{X_1, \dots, X_n\} = X_{(n)}.$$

□

比较矩估计量:

$$\hat{a} = \bar{X} - \sqrt{3B_2},$$

$$\hat{b} = \bar{X} + \sqrt{3B_2}.$$

(2) $E(X) = \frac{a+b}{2}$ 是参数 a, b 的函数, 因此 $E(X)$ 最大似然估计量为

$$E(\hat{X}) = \frac{\hat{a} + \hat{b}}{2} = \frac{X_{(1)} + X_{(n)}}{2}.$$

(3) 将样本值分别代入 $a, b, E(X)$ 最大似然估计量,

$$\hat{a} = 0.16, \hat{b} = 0.96, E(X) = 0.56.$$



例 (课上练习)

设总体 $X \sim b(1, p)$, p 未知, X_1, \dots, X_n 是总体 X 的样本. 求 p 的矩估计量和最大似然估计量.

对总体的未知参数可用不同方法求得不同的估计量，如何评价不同估计量的好坏？

对总体的未知参数可用不同方法求得不同的估计量，如何评价不同估计量的好坏？

⇒ 下次课：估计量的评价准则！