估计问题: $X \sim F(x;\theta)$, θ未知, 用样本$X_1, X_2, \cdots, X_n$的函数估计θ (或X的数字特征).

$$\text{估计问题}\begin{cases}\text{点估计}\, g(X_1,\cdots,X_n)\begin{cases}\text{矩估计}\\ \text{最大似然估计}\end{cases}\\ \text{区间估计}\left[g_1(X_1,\cdots,X_n), g_2(X_1,\cdots,X_n)\right]\end{cases}$$

一、矩估计法 (略)

1. 理论依据:

记: 样本的k阶原点矩—— $A_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k,\ k = 1,2,\cdots$

总体的k阶原点矩—— $a_k = E(X^k),\ k = 1,2,\cdots$, 则

$$A_k \xrightarrow{P} a_k \quad (\text{即} \lim_{n\to\infty} P\left\{\left|A_k - a_k\right| < \varepsilon\right\} = 1)$$

$$g(A_1, A_2, \cdots, A_k) \xrightarrow{P} g(a_1, a_2, \cdots, a_k).$$

【例2.3(P_{31})】设总体X服从 $[\theta_1,\theta_2]$上的均匀分布, X_1, X_2, …, X_n是来自X的样本, 求θ_1,θ_2的矩估计量.

$$\hat{\theta}_{1矩估计}=A_1-\sqrt{3(A_2-A_1^2)},\quad \hat{\theta}_{2矩估计}=A_1+\sqrt{3(A_2-A_1^2)}$$

2. 具体做法:

若总体分布中有m个未知参数$\theta_1,\cdots,\theta_m$, 则

$$由\begin{cases} a_1=a_1(\theta_1,\cdots,\theta_m)\\ \quad\vdots\\ a_m=a_m(\theta_1,\cdots,\theta_m)\end{cases}\Rightarrow\begin{cases}\theta_1=\theta_1(a_1,\cdots,a_m)\\ \quad\vdots\\ \theta_m=\theta_m(a_1,\cdots,a_m)\end{cases},$$

用A_i代替$a_i(i=1,\ldots,m)$, 得:

$$\begin{cases}\hat{\theta}_1=\hat{\theta}_1(A_1,A_2,\cdots,A_m)\\ \quad\vdots\\ \hat{\theta}_m=\hat{\theta}_m(A_1,A_2,\cdots,A_m)\end{cases}$$

—$\theta_1,\cdots,\theta_m$的矩估计量.

注意1: 矩估计量的观察值称为矩估计值, 与样本值有关.

【例2.4(P_{31})】在n重贝努利试验中, 事件A发生了n_A次, 试求事件A发生概率的矩估计值.

$$\hat{P}_{矩估计} = \frac{n_A}{n} = f_n(A).$$

【例2.8(P_{35})】设X_1, $X_2,\ldots,X_n$是来自$N(\mu,\sigma^2)$的样本, 求μ,σ^2的矩估计量.

$$\hat{\mu}_{矩估计} = \bar{X}, \quad \hat{\sigma}^2_{矩估计} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

【例2.5(P_{32})】设X的概率密度为
$$f(x,\theta_1,\theta_2)=\begin{cases}\dfrac{\theta_2}{\Gamma\left(\dfrac{1+\theta_1}{\theta_2}\right)}x^{\theta_1}e^{-x^{\theta_2}}, & x>0\\ 0, & x\le 0\end{cases},$$

$\theta_1>-1,\theta_2>0$, $X_1, X_2,\ldots,X_n$是来自X的样本, 试求θ_1,θ_2的矩估计量.

结论1: 此分布无法得到θ_1,θ_2矩估计量的解析式, 只能求其数值解.

【例2.6(P_{33})】设总体X服从柯西 (Cauchy) 分布, 其概率密度函数为

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

$X_1, X_2,\ldots,X_n$是来自X的样本, 试求θ的矩估计量.

结论2: 此分布各阶矩均不存在, 故θ无矩估计量(值).

3. 矩估计法的优点与不足:

优点: 矩估计法简便易行, 且当n充分大时, 估计的精确度也很高;

不足: 矩估计法只用到总体的数字特征, 而未用到总体的具体分布形式, 损失了一部分很有用的信息, 显得粗糙.

二、极大似然估计法 (MLE——Maximum Likelihood Estimation, 由费歇 (R. A. Fisher) 提出)

【补例1】设总体 $X \sim b(1,p)$ (p 未知), 今抽得一样本 $(x_1=1, x_2=0, x_3=1, x_4=1, x_5=0)$, 试问该样本出现的概率有多大?

1. 样本的似然函数:

离散型总体: $$L(\theta)=P(X_1=x_1,\cdots,X_n=x_n)\overset{X_i\text{独立}}{\underset{\text{且与}X\text{同分布}}{=\!=\!=\!=\!=}}\prod_{i=1}^{n}p(x_i;\theta)$$

连续型总体: $$L(\theta)=f_{(X_1,\cdots,X_n)}(x_1,\cdots,x_n)\overset{X_i\text{独立}}{\underset{\text{且与}X\text{同分布}}{=\!=\!=\!=\!=}}\prod_{i=1}^{n}f(x_i;\theta)$$

2. 理论依据:

$$\boxed{L(\hat{\theta})=\max L(\theta)}$$

θ的最大似然估计量

3. 具体做法:

(1) 写出似然函数 $L(\theta)$;

(2) 求 $L(\theta)$ 或 $\ln L(\theta)$ 在 θ 可能取值范围 Θ 内的最大值点 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$.

$$\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n) \quad \text{——}\theta\text{的最大似然估计值}$$

$$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n) \quad \text{——}\theta\text{的最大似然估计量}$$

注意2: 极大似然估计值与样本值有关.

【例2.8(P_{35})】设$X_1, X_2,\ldots,X_n$是来自 $N(\mu,\sigma^2)$ 的样本, 求 μ,σ^2 的MLE.

$$\boxed{\hat{\mu}_{矩估计}=\hat{\mu}_{MLE}=\bar{X}, \quad \hat{\sigma}^2_{矩估计}=\hat{\sigma}^2_{MLE}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

【例2.4(P_{31})】在n重贝努利试验中，事件A发生了n_A次，试求事件A发生概率的MLE.

$$\boxed{\hat{P}(A)_{MLE}=\hat{P}(A)_{矩估计}=\frac{n_A}{n}=f_n(A)}$$

【例2.9(P_{36})】设有k个事件$A_1, A_2, \ldots, A_k$两两互斥，其概率$p_1, p_2, \ldots, p_k$之和为1，做n次重复独立试验，以n_i表示事件$A_i(i=1,2, \ldots,k)$发生的次数，求$p_1, p_2, \ldots, p_k$的MLE.

$$\boxed{\hat{P}(A_i)_{MLE}=\frac{n_i}{n}}$$

【例2.10(P_{37})】设总体X服从 $[\theta_1,\theta_2]$上的均匀分布, X_1, X_2,…,X_n是来自X的样本, 求θ_1,θ_2的MLE.

$$\boxed{\hat{\theta}_{1矩估计}=A_1-\sqrt{3(A_2-A_1^2)},\quad \hat{\theta}_{2矩估计}=A_1+\sqrt{3(A_2-A_1^2)}}$$

$$\boxed{\hat{\theta}_{1MLE}=X_{(1)},\quad \hat{\theta}_{2MLE}=X_{(n)}}$$

【例2.11(P_{37})】设总体X服从柯西 (Cauchy) 分布, 其概率密度函数为

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

X_1, X_2,…,X_n是来自X的样本, 试求θ的MLE.

结论3: 并非每个MLE问题都可得到解析解, 某些只能得到数值解.

结论4: 矩估计与极大似然估计是两种不同的点估计法, 可能出现:

(1) 对同一分布中的同一参数, 两种方法结果相同.

如 $X \sim b(1, p)\,(p = P(A))$, $\boxed{\hat{p}_{矩} = \hat{p}_{MLE} = f_n(A)}$.

如 $X \sim N(\mu, \sigma^2)$, $\boxed{\begin{cases} \hat{\mu}_{矩} = \hat{\mu}_{MLE} = \bar{X} \\ \hat{\sigma}^2_{矩} = \hat{\sigma}^2_{MLE} = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2 \end{cases}}$.

(2) 对同一分布中的同一参数, 两种方法结果不同.

如 $X \sim U(\theta_1, \theta_2)$, $\boxed{\begin{cases} \hat{\theta}_{1矩估计} = A_1 - \sqrt{3(A_2 - A_1^2)} \\ \hat{\theta}_{2矩估计} = A_1 + \sqrt{3(A_2 - A_1^2)} \end{cases}}$,

$\boxed{\begin{cases} \hat{\theta}_{1MLE} = X_{(1)} \\ \hat{\theta}_{2MLE} = X_{(n)} \end{cases}}$.