

随机变量

总体

随机抽样

样本

统计量

代表性
独立性

频率直方
图……

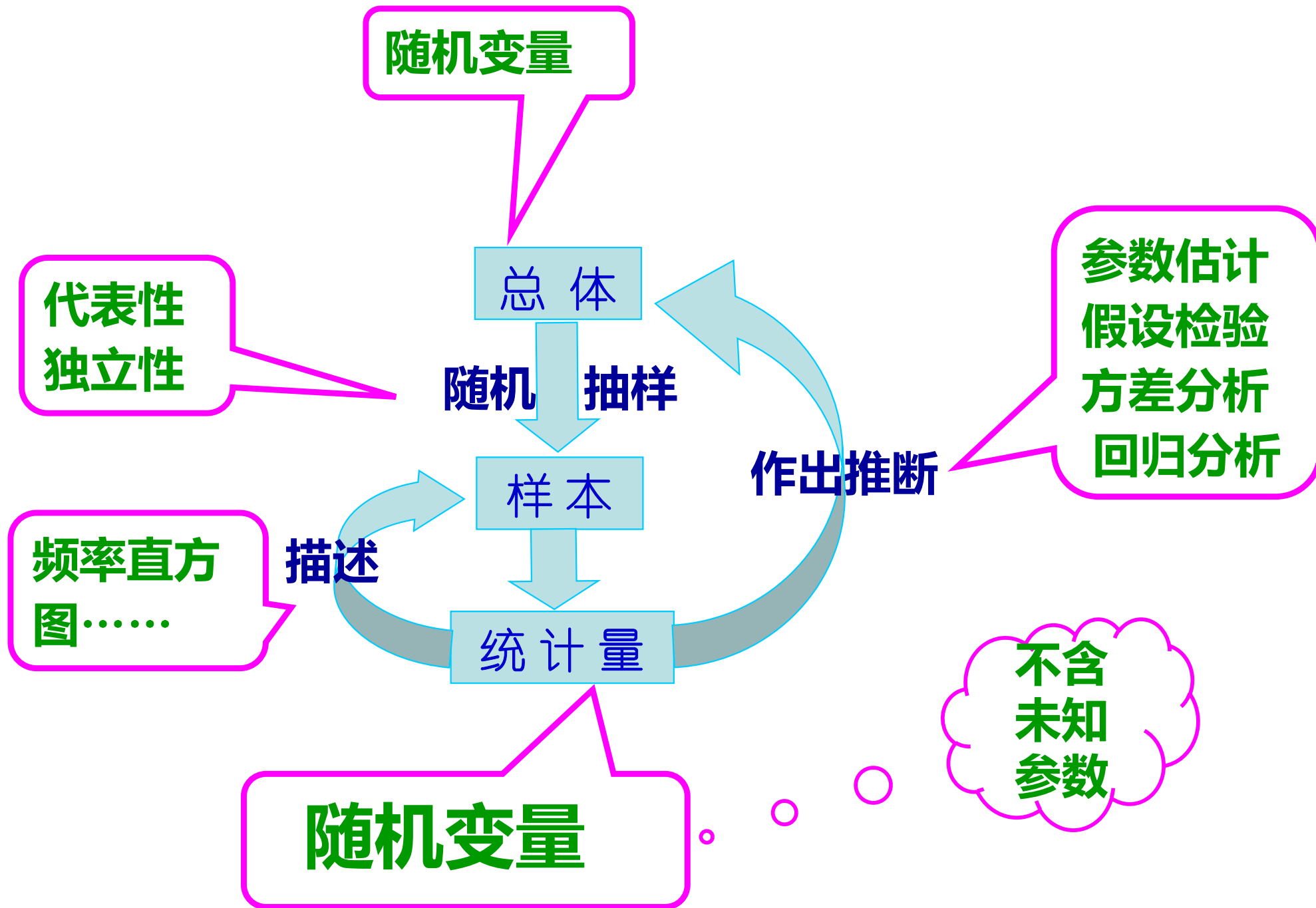
描述

作出推断

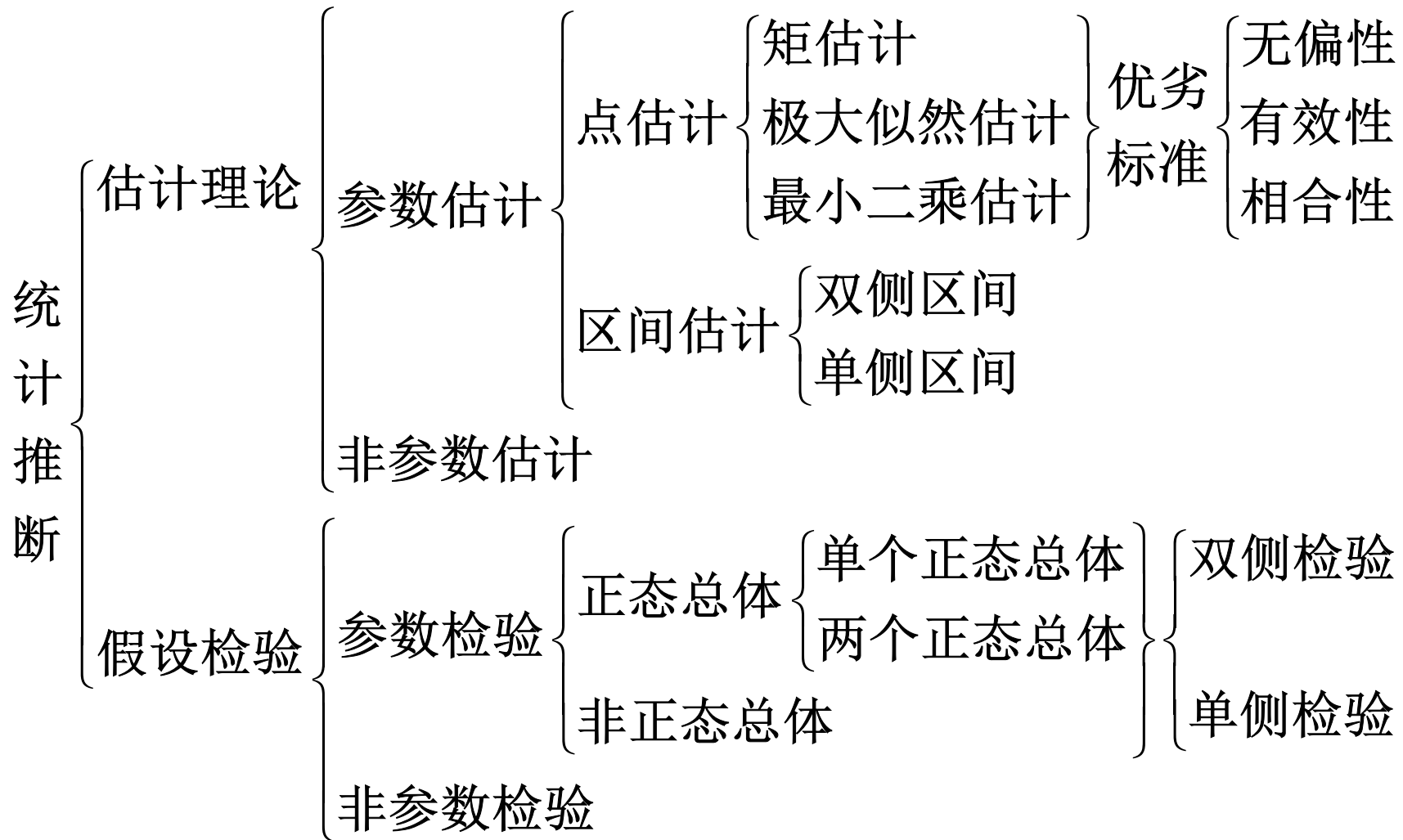
参数估计
假设检验
方差分析
回归分析

随机变量

不含
未知
参数



§ 6.1 点估计方法



§ 6.1 点估计方法

设总体 X 的分布函数 $F(x; \theta)$ 的形式已知, $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ 是待估参数. X_1, X_2, \dots, X_n 是 X 的一个样本, x_1, x_2, \dots, x_n 为相应的一个样本值.

点估计问题就是要构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$, 用它的观察值 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 来估计未知参数 θ .

由于估计量是样本的函数, 是随机变量, 故对不同的样本值, 得到的参数值往往不同, 如何求估计量是关键问题.

一、矩估计法

是英国统计学家K.皮尔逊最早提出的，其基本思想是用样本矩估计总体矩. 这种方法的理论依据是大数定律. 记

$$\mu_i = E(X^i) = g_i(\theta_1, \dots, \theta_k), \quad A_i = \frac{1}{n} \sum_{j=1}^n X_j^k$$

分别表示总体和样本的 i 阶矩. 用样本 i 阶矩代替总体的 i 阶矩, 建立如下方程(组):

$$\left\{ \begin{array}{l} g_1(\theta_1, \dots, \theta_k) = A_1 \\ g_2(\theta_1, \dots, \theta_k) = A_2 \\ \dots\dots\dots \\ g_k(\theta_1, \dots, \theta_k) = A_k \end{array} \right.$$

用样本矩估计 总体矩

该方程组的解 $\hat{\theta}_j = h_j(X_1, \dots, X_k)$, $j = 1, 2, \dots, k$ 称为 $\theta_1, \dots, \theta_k$ 的矩估计量, $\hat{\theta}_j = h_j(x_1, \dots, x_k)$ 称为估计值.

不论总体服从什么分布, 若总体期望 μ 与方差 σ^2 存在, 则它们的矩估计量分别为

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

事实上, 按矩法原理, 令

$$\begin{cases} \mu = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \\ \mu^2 + \sigma^2 = E(X^2) = \frac{1}{n} \sum_{i=1}^n X_i^2 = \bar{X}^2 + S_n^2 \end{cases}$$

其解为 $\hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = S_n^2$. 得证.

例1 设总体 $X \sim N(\mu, \sigma^2)$, X_1, X_2, \dots, X_n 为总体的样本, 求 μ, σ^2 的矩法估计量.

解 $\hat{\mu}_{\text{矩}} = \bar{X}, \quad \hat{\sigma}_{\text{矩}}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = S_n^2.$

例2 设总体 $X \sim E(\lambda)$, X_1, X_2, \dots, X_n 为总体的样本. 求 λ 的矩法估计量.

解 由于 $E(X) = 1/\lambda$, 因此 $1/\lambda = \bar{X}$,
所以 $\hat{\lambda}_{\text{矩}} = 1/\bar{X}.$

例3 设从某灯泡厂某天生产的灯泡中随机抽取10只灯泡, 测得其寿命为(单位: 小时).

1050, 1100, 1080, 1120, 1200, 1250, 1040, 1130, 1300, 1200. 试用矩法估计该厂这天生产的灯泡的平均寿命及寿命分布的方差.

解 $\hat{\mu} = \bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 1147(h);$

$$\hat{\sigma}^2 = \frac{1}{10} \sum_{i=1}^{10} x_i^2 - \bar{x}^2 = 6821(h^2).$$

例4 设总体 $X \sim U(a, b)$, a, b 未知, 求参数 a, b 的矩法估计量.

解 由于 $E(X) = \frac{a+b}{2}, \quad D(X) = \frac{(b-a)^2}{12}.$

$$E(X^2) = D(X) + E^2(X) = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2$$

$$\text{令} \begin{cases} \frac{a+b}{2} = \bar{X} \\ \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 = \bar{X}^2 + S_n^2 \end{cases}$$

解得

$$\hat{a}_{\text{矩}} = \bar{X} - \sqrt{3(A_2 - \bar{X}^2)} = \bar{X} - \sqrt{3}S_n,$$

$$\hat{b}_{\text{矩}} = \bar{X} + \sqrt{3(A_2 - \bar{X}^2)} = \bar{X} + \sqrt{3}S_n.$$

矩估计法的优点是简单易行,并不需要事先知道总体是什么分布.

缺点是,当总体类型已知时,没有充分利用分布提供的信息.

一般场合下,矩估计量不具有唯一性. 其主要原因在于建立矩法方程时,选取那些总体矩用相应样本矩代替带有一定的随意性(例如:对于泊松分布,参数 λ 既是期望也是方差,因此矩估计 \bar{X}, S_n^2 都可看作是参数 λ 的矩估计).

二、极大似然估计法

是在总体类型已知条件下使用的一种参数估计方法. 它首先是由德国数学家 **Gauss** 在 **1821** 年提出的, **Fisher** 在 **1922** 年重新发现了这一方法, 并首先研究了这种方法的一些性质.

基本思想: 一次试验就出现的事件有较大的概率.

例如: 有两外形相同的箱子, 各装 **100** 个球.

甲箱 **99** 个白球 **1** 个红球

乙箱 **1** 个白球 **99** 个红球

现从两箱中任取一箱, 并从箱中任取一球, 结果所取得的球是白球. 问这个球是从哪个箱子中取出的?

若取到的是甲箱，从中取出白球的概率为**0.99**，
若取到的是乙箱，从中取出白球的概率为**0.01**。现在
一次试验白球出现了，人们的直觉就是：

“此白球**最像**从甲箱取出的”。“极大似然”之意

分析

对离散型总体 $P(X = x_k) = P(x_k; \theta)$, $k = 1, 2, \dots$

抽样结果：事件 $\{X_1 = x_1, \dots, X_n = x_n\}$ 发生的概率为：

$$f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n P(x_i; \theta), \quad \theta \in \Theta$$

对连续型总体 密度函数为 $f(x; \theta)$

抽样结果：事件 $\{X_1 = x_1, \dots, X_n = x_n\}$ 可理解成事件

$$\{X_1 \in (x_1, x_1 + dx_1), \dots, X_n \in (x_n, x_n + dx_n)\},$$

其发生的概率为:

$$f(x_1, x_2, \dots, x_n; \theta) \approx \prod_{i=1}^n f(x_i; \theta) dx_i, \quad \theta \in \Theta.$$

极大似然估计的思想是选取 $\theta \in \Theta$, 使得上述概率达到最大.

极大似然估计原理:

设 X_1, X_2, \dots, X_n 是取自总体 X 的一个样本, 样本的联合密度(连续型)为.

$$f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta), \quad \theta \in \Theta.$$

或联合概率函数(离散型)

$$f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n p(x_i; \theta), \quad \theta \in \Theta,$$

(其中 Θ 是参数空间—— θ 可能的取值范围).

当给定样本值 x_1, x_2, \dots, x_n 时, 定义似然函数为

$$L(\theta) = f(x_1, x_2, \dots, x_n; \theta).$$

$L(\theta)$ 看作参数 θ 的函数, 它可作为将以多大可能产生样本值 x_1, x_2, \dots, x_n 的一种度量.

极大似然估计法就是用使 $L(\theta)$ 达到最大值的 $\hat{\theta}$ 去估计 θ . 即 $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$. 其解

$$\hat{\theta} = g(x_1, x_2, \dots, x_n)$$

为参数 θ 的极大似然估计值, 简记为 $\hat{\theta}_{mle}$. 称统计量

$$\hat{\theta} = g(X_1, X_2, \dots, X_n)$$

为参数 θ 的极大似然估计量. 简记为 $\hat{\theta}_{MLE}$.

求极大似然估计的步骤

(一) 写出似然函数

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n p(x_i; \theta)$$

或

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta);$$

(二) 取对数 ($L(\theta)$ 与 $\ln L(\theta)$ 具有相同的极值点)

$$\ln L(\theta) = \sum_{i=1}^n \ln p(x_i; \theta) \quad \text{或} \quad \ln L(\theta) = \sum_{i=1}^n \ln f(x_i; \theta);$$

对数似然函数

(三) 求 $\ln L(\theta)$ 的驻点, 即令 $\frac{d \ln L(\theta)}{d \theta} = 0$,

对数似然方程

(四) 验证是否是极大值点. 即得极大似然估计值 $\hat{\theta}$.

当总体分布中含有多个未知参数时, 此时需令

$$\frac{\partial}{\partial \theta_i} \ln L = 0, \quad i = 1, 2, \dots, k$$

对数似然方程组

求解上述方程组, 即可得极大似然估计值 $\hat{\theta}_1, \dots, \hat{\theta}_k$.

注：(1) 当似然函数是未知参数的单调函数时，此时不能利用导数为零的方法求解，只能从定义出发求解（见例题8）。

(2) 待估参数的MLE可能不存在，也可能不唯一。

例5 设 $X \sim B(1, p)$, X_1, X_2, \dots, X_n 是来自 X 的一个样本, 求 p 的极大似然估计量。

解 设 x_1, x_2, \dots, x_n 为相应于样本 X_1, X_2, \dots, X_n 的一个样本值, X 的分布律为

$$P\{X = x\} = p^x (1 - p)^{1-x}, \quad x = 0, 1.$$

似然函数
$$L(p) = \prod_{i=1}^n p^{x_i} (1 - p)^{1-x_i} = p^{\sum x_i} (1 - p)^{n - \sum x_i},$$

$$\ln L(p) = \left(\sum_{i=1}^n x_i\right) \ln p + \left(n - \sum_{i=1}^n x_i\right) \ln(1-p),$$

令

$$\frac{d}{dp} \ln L(p) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p} = 0,$$

解得 p 的极大似然估计值

$$\hat{p}_{mle} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

p 的极大似然估计量为

$$\hat{p}_{MLE} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

与矩估计相同

例6 设 X 服从参数为 $\lambda (\lambda > 0)$ 的泊松分布, X_1, X_2, \dots, X_n 是来自 X 的一个样本, 求 λ 的极大似然估计量.

解 X 的分布律为

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

似然函数为

$$L(\lambda) = \prod_{i=1}^n \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) = e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \left(\prod_{i=1}^n (x_i!) \right)^{-1},$$

$$\ln L(\lambda) = -n\lambda + \left(\sum_{i=1}^n x_i \right) \ln \lambda - \sum_{i=1}^n (x_i!),$$

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

解得 p 的极大似然估计值 $\hat{p}_{mle} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$.

p 的极大似然估计量为 $\hat{p}_{MLE} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$.

与矩估计相同

例7 设总体 $X \sim N(\mu, \sigma^2)$, X_1, X_2, \dots, X_n 为总体的样本. 求 μ, σ^2 的极大似然估计.

解 似然函数为

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \\ &= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}}, \end{aligned}$$

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

似然方程组为

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

解得

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

μ , σ^2 的极大似然估计量分别为

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = S_n^2.$$

与矩估计相同

例8 设 $X \sim U(a, b)$, x_1, x_2, \dots, x_n 是 X 的一个样本值. 求 a, b 的极大似然估计值与极大似然估计量.

解 X 的密度函数为

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{其它} \end{cases}$$

似然函数为

$$L(x_1, x_2, \dots, x_n; a, b) = \begin{cases} \frac{1}{(b-a)^n}, & a \leq x_i \leq b, \\ & i = 1, 2, \dots, n \\ 0, & \text{其它.} \end{cases}$$

似然函数 L 是 a, b 的单调函数, a 越大, b 越小, L 越大.

显然 a 的最大可能取值为 x_{\min} ， b 的最小可能取值为 x_{\max} 。因此 a, b 的极大似然估计值分别为 x_{\min}, x_{\max} 。
估计量为

$$\hat{a} = X_{\min}, \quad \hat{b} = X_{\max}.$$

与矩估计不同

练习题1 设某种电子器件的寿命 X 服从双参数的指数分布, X_1, \dots, X_n 是来自 X 的简单样本, X 的概率密度为

$$f(x; \lambda, \theta) = \begin{cases} \lambda e^{-\lambda(x-\theta)}, & x \geq \theta, \\ 0, & \text{其他.} \end{cases}$$

其中 λ, θ ($\lambda, \theta > 0$)为未知参数.

(1)求 λ 与 θ 的矩估计量,

(2)求 λ 与 θ 的极大似然估计量.

练习题1答案

(1) 矩估计 $\hat{\lambda}_1 = 1/S_n, \hat{\theta}_1 = \bar{X} - S_n;$

(2) 极大似然估计 $\hat{\lambda}_2 = \frac{1}{\bar{X} - X_{(1)}}, \hat{\theta}_2 = X_{(1)}.$

二者不同