点估计

续本达

清华大学 工程物理系

2023-11-15

复习

统计量

样本 $(X_1, X_2, \cdots, X_n)$ 的不含有未知参数的连续函数 $g(X_1, X_2, \cdots, X_n)$ 称为 **统计量** 。

- 样本均值、样本方差、样本标准差

抽样分布

正态总体 $X \sim N(\mu, \sigma^2)$ 和 $X' \sim N(\mu', \sigma'^2)$

χ^2 分布 $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

F 分布 $\dfrac{S^2/S'^2}{\sigma^2/\sigma'^2} \sim F(n-1, n'-1)$

t 分布 $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

统计推断

引子

- 统计推断是数理统计理论的主要部分。
- 现行的统计推断理论，是建立在概率论的基础上的。
- 从总体中抽出的样本，去推断总体的性质（期望、方差、分布等）。

例

假定在教室里的学生中，身高服从正态分布 $N(\mu, \sigma^2)$，其中 μ, σ^2 是未知参数，即为推断的对象。

概念导图

- 统计推断的基本问题
  - 参数估计问题
    - 点估计
    - 区间估计
  - 假设检验问题

参数估计

- 参数 是刻画总体某方面概率特性的数量。
- 当此数量未知时，从总体抽出一个样本，用一定的方法对它进行参数估计。
  1. 方法问题：如何利用样本估计未知参数
  2. 评判标准：如何评价估计的好坏

参数估计类型

点估计 估计未知参数的值；

区间估计 估计未知参数的取值范围，并使此范围包含未知参数真值的概率为给定的值。

例

$X \sim N(\mu, \sigma^2)$，若 μ, σ^2 未知，通过构造统计量，给出它们的 估计值 （点估计） 或 取值范围 （区间估计） 就是参数估计的内容。

参数估计

- 参数 是刻画总体某方面概率特性的数量。
- 当此数量未知时，从总体抽出一个样本，用一定的方法对它进行参数估计。
  1. 方法问题：如何利用样本估计未知参数
  2. 评判标准：如何评价估计的好坏

参数估计类型

点估计 估计未知参数的值；

区间估计 估计未知参数的取值范围，并使此范围包含未知参数真值的概率为给定的值。

例

$X \sim N(\mu, \sigma^2)$，若 μ, σ^2 未知，通过构造统计量，给出它们的 估计值 （点估计） 或 取值范围 （区间估计） 就是参数估计的内容。

点估计

引子

例

设在炸药制造厂，一天中发生着火的次数 X 服从以 λ 为参数的泊松分布，参数未知，现有以下的样本值，试估计参数 λ 。

着火次数 k	0	1	2	3	4	5	6	7	
发生 k 次着火的天数 n_k	75	90	54	22	6	2	1	0	$\Sigma = 250$

由于 $X \sim \pi(\lambda)$ ，故有 $\lambda = \mathrm{E}(X)$，用样本均值估计总体均值 $\mathrm{E}(X)$，由数据计算得到

$$\overline{X} = 1.22$$

用程序计算样本方差

同样有 $\mathrm{Var}(X) = \lambda$

```
library(Hmisc)
x <- data.frame(fire=c(0, 1, 2, 3, 4, 5, 6, 7),
                days=c(75, 90, 54, 22, 6, 2, 1, 0))
wtd.var(x$fire, weights=x$days)
```

```
1.27042570281125
```

$$s^2 = 1.27$$

该用哪个?

定义

定义 (点估计)

用一个数值作为未知参数的估计值称为**点估计**。

- 设总体 X 的分布函数的形式已知，θ 是待估参数，$X_1, X_2, \cdots, X_n$ 为总体的一个样本。
- 点估计构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$，用它的观察值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 作为待估参数的近似值。

约定

尖帽符号 $\hat{}$ 表示估计量：它是统计量。

定义

定义 (点估计)

用一个数值作为未知参数的估计值称为点估计。

- 设总体 X 的分布函数的形式已知，θ 是待估参数，$X_1, X_2, \cdots, X_n$ 为总体的一个样本。
- 点估计构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$，用它的观察值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 作为待估参数的近似值。

约定

尖帽符号 ^ 表示估计量：它是统计量。

定义

定义 (点估计)

用一个数值作为未知参数的估计值称为点估计。

- 设总体 X 的分布函数的形式已知，θ 是待估参数，$X_1, X_2, \cdots, X_n$ 为总体的一个样本。
- 点估计构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$，用它的观察值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 作为待估参数的近似值。

约定

尖帽符号 $\hat{}$ 表示估计量：它是统计量。

矩估计法

定义

定义 (矩估计)

用样本 k 阶矩作为总体 k 阶矩的估计量，建立含待估参数的方程，从而解出待估参数。

- 设随机变量 $X \sim f(x;\theta_1,\theta_2,\cdots,\theta_k)$，其中 $\theta_1,\theta_2,\cdots,\theta_k$ 为待估参数。假设总体的前 k 阶矩存在：

$$\mathrm{E}(X^r) = \mu_r(\theta_1,\theta_2,\cdots,\theta_k), 1 \le r \le k$$

- 假设 $X_1, X_2, \cdots, X_n$ 为来自 X 的一个样本，r 阶样本矩 $A_r = \dfrac{1}{n}\sum_{i=1}^{n} X_i^r$。
- A_r 及其函数依概率收敛于相应的总体矩。因此可以
  1. 用样本矩作为相应的总体矩的估计量；
  2. 用样本矩的函数作为相应的总体矩函数的估计量。

定义

定义 (矩估计)

用样本 k 阶矩作为总体 k 阶矩的估计量，建立含待估参数的方程，从而解出待估参数。

- 设随机变量 $X \sim f(x;\theta_1,\theta_2,\cdots,\theta_k)$，其中 $\theta_1,\theta_2,\cdots,\theta_k$ 为待估参数。假设总体的前 k 阶矩存在：

$$\mathrm{E}(X^r) = \mu_r(\theta_1,\theta_2,\cdots,\theta_k), 1 \le r \le k$$

- 假设 $X_1,X_2,\cdots,X_n$ 为来自 X 的一个样本，r 阶样本矩 $A_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$。
- A_r 及其函数依概率收敛于相应的总体矩。因此可以
  1. 用样本矩作为相应的总体矩的估计量；
  2. 用样本矩的函数作为相应的总体矩函数的估计量。

定义

定义 (矩估计)

用样本 k 阶矩作为总体 k 阶矩的估计量，建立含待估参数的方程，从而解出待估参数。

- 设随机变量 $X \sim f(x;\theta_1,\theta_2,\cdots,\theta_k)$，其中 $\theta_1,\theta_2,\cdots,\theta_k$ 为待估参数。假设总体的前 k 阶矩存在：

$$\mathrm{E}(X^r) = \mu_r(\theta_1,\theta_2,\cdots,\theta_k), 1 \le r \le k$$

- 假设 $X_1, X_2, \cdots, X_n$ 为来自 X 的一个样本，r 阶样本矩 $A_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$。
- A_r 及其函数依概率收敛于相应的总体矩。因此可以
  1. 用样本矩作为相应的总体矩的估计量；
  2. 用样本矩的函数作为相应的总体矩函数的估计量。

定义

定义 (矩估计)

用样本 k 阶矩作为总体 k 阶矩的估计量，建立含待估参数的方程，从而解出待估参数。

- 设随机变量 $X \sim f(x;\theta_1,\theta_2,\cdots,\theta_k)$，其中 $\theta_1,\theta_2,\cdots,\theta_k$ 为待估参数。假设总体的前 k 阶矩存在：

$$\mathrm{E}(X^r) = \mu_r(\theta_1,\theta_2,\cdots,\theta_k), 1 \le r \le k$$

- 假设 $X_1, X_2, \cdots, X_n$ 为来自 X 的一个样本，r 阶样本矩 $A_r = \dfrac{1}{n}\sum_{i=1}^{n} X_i^r$。
- A_r 及其函数依概率收敛于相应的总体矩。因此可以
  1. 用样本矩作为相应的总体矩的估计量；
  2. 用样本矩的函数作为相应的总体矩函数的估计量。

具体做法

总体的前 k 阶矩构成联立方程组，含 k 个未知参数。一般情况下可以解出这 k 个参数 $\theta_1, \theta_2, \cdots, \theta_k$：

$$\begin{cases} \mu_1 = \mu_1(\theta_1, \theta_2, \cdots, \theta_k), \\ \mu_2 = \mu_2(\theta_1, \theta_2, \cdots, \theta_k), \\ \vdots \\ \mu_k = \mu_k(\theta_1, \theta_2, \cdots, \theta_k), \end{cases} \Rightarrow \begin{cases} \theta_1 = \theta_1(\mu_1, \mu_2, \cdots, \mu_k), \\ \theta_2 = \theta_2(\mu_1, \mu_2, \cdots, \mu_k), \\ \vdots \\ \theta_k = \theta_k(\mu_1, \mu_2, \cdots, \mu_k), \end{cases}$$

以样本矩 A_r 代替总体矩 $\mu_r(r = 1, 2, \cdots, k)$ 就得到待估参数的估计量，称为 矩估计量：

$$\hat{\theta}_i = \theta_i(A_1, A_2, \cdots, A_k), i = 1, 2, \cdots, k$$

矩估计量的观察值称为 矩估计值 。

具体做法

总体的前 k 阶矩构成联立方程组，含 k 个未知参数。一般情况下可以解出这 k 个参数 $\theta_1, \theta_2, \cdots, \theta_k$：

$$\boxed{\begin{cases} \boldsymbol{\mu}_1 = \boldsymbol{\mu}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_k), \\ \boldsymbol{\mu}_2 = \boldsymbol{\mu}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_k), \\ \qquad \vdots \\ \boldsymbol{\mu}_k = \boldsymbol{\mu}_k(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_k), \end{cases} \Rightarrow \begin{cases} \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_k), \\ \boldsymbol{\theta}_2 = \boldsymbol{\theta}_2(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_k), \\ \qquad \vdots \\ \boldsymbol{\theta}_k = \boldsymbol{\theta}_k(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_k), \end{cases}}$$

以样本矩 A_r 代替总体矩 $\mu_r(r = 1, 2, \cdots, k)$ 就得到待估参数的估计量，称为 矩估计量 ：

$$\hat{\theta}_i = \theta_i(A_1, A_2, \cdots, A_k), i = 1, 2, \cdots, k$$

矩估计量的观察值称为 矩估计值 。

矢量表述

- 随机变量 $X \sim f(x;\vec{\theta}), \vec{\theta} \equiv (\theta_1, \cdots, \theta_k)$ 为待估参数。总体的前 k 阶矩构成矢量 $\vec{\mu}$ $(\vec{\theta})$：每个分量都是 $\vec{\theta}$ 的函数，$\mu_i = \mu_i(\vec{\theta})$。
- 一般情况下可反解出 $\vec{\theta}(\vec{\mu})$。
- 设 $\vec{X} = (X_1, \cdots, X_n)$ 是总体的一个样本，样本的前 k 阶样本矩构成矢量 $\vec{A} = (A_1, \cdots, A_k)$，则待估函数的矩估计量为

$$\vec{\theta} = \vec{\theta}(\vec{\mu} = \vec{A})$$

矢量表述

- 随机变量 $X \sim f(x;\vec{\theta}), \vec{\theta} \equiv (\theta_1, \cdots, \theta_k)$ 为待估参数。总体的前 k 阶矩构成矢量 $\vec{\mu}$ $(\vec{\theta})$：每个分量都是 $\vec{\theta}$ 的函数，$\mu_i = \mu_i(\vec{\theta})$。
- 一般情况下可反解出 $\vec{\theta}(\vec{\mu})$。
- 设 $\vec{X} = (X_1, \cdots, X_n)$ 是总体的一个样本，样本的前 k 阶样本矩构成矢量 $\vec{A} = (A_1, \cdots, A_k)$，则待估函数的矩估计量为

$$\hat{\vec{\theta}} = \vec{\theta}(\vec{\mu} = \vec{A})$$

例：指数分布的矩估计

例

设总体 $X \sim \mathrm{Exp}(\lambda)$，即 $f(x) = \lambda e^{-\lambda x}$。 $X_1, \cdots, X_n$ 为总体的一个样本。求参数 λ 的矩法估计量。

$$\mu_1 = \mathrm{E}(X) = \frac{1}{\lambda}$$

$$\Longrightarrow \lambda = \frac{1}{\mu_1}$$

$$\Longrightarrow \hat{\lambda} = \frac{1}{A_1} = \frac{1}{\overline{X}}$$

例：指数分布的矩估计

例

设总体 $X \sim \mathrm{Exp}(\lambda)$，即 $f(x) = \lambda e^{-\lambda x}$。 $X_1, \cdots, X_n$ 为总体的一个样本。求参数 λ 的矩法估计量。

$$
\begin{aligned}
\mu_1 &= \mathrm{E}(X) = \frac{1}{\lambda} \\
\Longrightarrow \lambda &= \frac{1}{\mu_1} \\
\Longrightarrow \hat{\lambda} &= \frac{1}{A_1} = \frac{1}{\overline{X}}
\end{aligned}
$$

例：期望与方差的矩估计

例

如果总体的期望 μ 与方差 σ^2 都存在（大部分常用分布满足该条件），则它们的矩估计量分别为

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

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

套用矩估计法

$$\mu_1 = \mathrm{E}(X) = \mu$$

$$\mu_2 = \mathrm{E}(X^2) = \mathrm{Var}(X) + [\mathrm{E}(X)]^2 = \sigma^2 + \mu^2$$

代入前两阶样本矩即可反解出。

例：期望与方差的矩估计

例

如果总体的期望 μ 与方差 σ^2 都存在（大部分常用分布满足该条件），则它们的矩估计量分别为

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

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

套用矩估计法

$$\mu_1 = \mathrm{E}(X) = \mu$$

$$\mu_2 = \mathrm{E}(X^2) = \mathrm{Var}(X) + [\mathrm{E}(X)]^2 = \sigma^2 + \mu^2$$

代入前两阶样本矩即可反解出。

例：均匀分布的矩估计

例

设总体 $X \sim U(a,b)$，a,b 未知，求参数 a,b 的 矩估计量.

一二阶矩为

$$\mu_1 = \mathrm{E}(X) = \frac{a+b}{2}$$

$$\mu_2 = \mathrm{E}(X^2) = \mathrm{Var}(X) + [\mathrm{E}(X)]^2 = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4}$$

解得 $a = \mu_1 - \sqrt{3(\mu_2 - \mu_1^2)}, b = \mu_1 + \sqrt{3(\mu_2 - \mu_1^2)}$，代之以样本矩 A_1 和 A_2 即得到 a 和 b 的矩估计量

$$\hat{a} = A_1 - \sqrt{3(A_2 - A_1^2)} = A_1 - \sqrt{3B_2}$$

$$\hat{b} = A_1 + \sqrt{3(A_2 - A_1^2)} = A_1 + \sqrt{3B_2}$$

例：均匀分布的矩估计

例

设总体 $X \sim U(a,b)$, a,b 未知, 求参数 a,b 的 矩估计量.

一二阶矩为

$$\mu_1 = \mathrm{E}(X) = \frac{a+b}{2}$$

$$\mu_2 = \mathrm{E}(X^2) = \mathrm{Var}(X) + [\mathrm{E}(X)]^2 = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4}$$

解得 $a = \mu_1 - \sqrt{3(\mu_2 - \mu_1^2)}, b = \mu_1 + \sqrt{3(\mu_2 - \mu_1^2)}$，代之以样本矩 A_1 和 A_2 即得到 a 和 b 的矩估计量

$$\hat{a} = A_1 - \sqrt{3(A_2 - A_1^2)} = A_1 - \sqrt{3B_2}$$

$$\hat{b} = A_1 + \sqrt{3(A_2 - A_1^2)} = A_1 + \sqrt{3B_2}$$

例：均匀分布的矩估计

例

设总体 $X \sim U(a,b)$，a,b 未知，求参数 a,b 的 矩估计量.

一二阶矩为

$$\mu_1 = \mathrm{E}(X) = \frac{a+b}{2}$$

$$\mu_2 = \mathrm{E}(X^2) = \mathrm{Var}(X) + [\mathrm{E}(X)]^2 = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4}$$

解得 $a = \mu_1 - \sqrt{3(\mu_2-\mu_1^2)}, b = \mu_1 + \sqrt{3(\mu_2-\mu_1^2)}$，代之以样本矩 A_1 和 A_2 即得到 a 和 b 的矩估计量

$$\hat{a} = A_1 - \sqrt{3(A_2-A_1^2)} = A_1 - \sqrt{3B_2}$$

$$\hat{b} = A_1 + \sqrt{3(A_2-A_1^2)} = A_1 + \sqrt{3B_2}$$

最大似然估计

引子

直觉

一次试验就出现的事件有较大的概率

例

有两外形相同的箱子, 各装 100 个球

一箱 99 个白球，1 个红球

一箱 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。所取的球来自哪一箱?

引子

直觉

一次试验就出现的事件有较大的概率

例

有两外形相同的箱子, 各装 100 个球

一箱 99 个白球，1 个红球

一箱 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。所取的球来自哪一箱?

引例

例

设总体 X 服从 0-1 分布, 且 $P(X=1)=p$, 用最大似然法求 p 的估计值.

总体 X 的概率分布为 $P(X=x)=p^x(1-p)^{1-x}, x=0,1$。设 $x_1, x_2, \cdots, x_n$ 为总体的样本 $X_1, X_2, \cdots, X_n$ 的观测值，则得到该样本值的概率为

$$P(X_1=x_1, \cdots, X_n=x_n)=\prod_{i=1}^{n} P(X_i=x_i)=p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \equiv L(p)$$

对于不同的 p，$L(p)$ 不同，取 p 使这个事件发生的概率最大

$$\hat{p}=\arg\max L(p)$$

引例

例

设总体 X 服从 0-1 分布, 且 $P(X=1)=p$, 用最大似然法求 p 的估计值.

总体 X 的概率分布为 $P(X=x)=p^x(1-p)^{1-x}, x=0,1$。设 $x_1,x_2,\cdots,x_n$ 为总体的样本 $X_1,X_2,\cdots,X_n$ 的观测值，则得到该样本值的概率为

$$P(X_1=x_1,\cdots,X_n=x_n)=\prod_{i=1}^{n}P(X_i=x_i)=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}\equiv L(p)$$

对于不同的 p，$L(p)$ 不同，取 p 使这个事件发生的概率最大

$$\hat{p}=\arg\max L(p)$$

求解

log 是单调增函数

$$\hat{p} = \arg\max L(p) = \arg\max \log L(p)$$

$$0 = \frac{\mathrm{d}\log L}{\mathrm{d}p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$\implies \hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

所以，$\hat{p} = \overline{x}$ 是 p 的最大似然估计值。它恰好也是矩估计值。

定义：似然函数

1. 设 X 为离散型随机变量，其分布律为 $P(X=x)=p(x,\theta), x=u_1,u_2,\cdots,\theta\in\Theta$ 样本 $X_1,X_2,\cdots,X_n$ 的联合概率分布为

$$P(X_1=x_1,\cdots,X_n=x_n)=p(x_1,\theta)p(x_2,\theta)\cdots p(x_n,\theta)=L(\vec{x},\theta)$$

2. 设 X 是连续型变量，取 $f(X,\theta)$ 为 X 的密度函数，似然函数定义为

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

最大似然法：maximum likelihood estimation (MLE)

$$\hat{\theta}(\vec{x})=\arg\max L(\vec{x};\theta)$$

称为**最大似然估计值**。

- 称统计量 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为参数 θ 的**最大似然估计量**。

多参数推广

未知参数可以不止一个，如 $\vec{\theta}$。设 X 的概率密度为 $f(x,\vec{\theta})$，则定义似然函数

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

若对于某组给定的样本值 $\vec{x}$，参数 $\hat{\vec{\theta}}$ 使似然函数取得最大值，即 $\hat{\vec{\theta}} = \arg\max L(\vec{x},\vec{\theta})$， 则称 $\hat{\theta}_r(\vec{x})$ 为 $\theta_r(r=1,2,\cdots,k)$ 的最大似然估计值，称统计量 $\hat{\theta}_r(\vec{X})$ 为 θ_r 的最大似然估计量。

似然方程组

若 $L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k)$ 关于 $\vec{\theta}$ 可微，则称

$$\frac{\partial}{\partial\theta_r}L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k) = 0, r=1,2,\cdots,k$$

为 似然方程组 。$\frac{\partial}{\partial\theta_r}\log L = 0$ 为 对数似然方程组 。

多参数推广

未知参数可以不止一个，如 $\vec{\theta}$。设 X 的概率密度为 $f(x,\vec{\theta})$，则定义似然函数

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

若对于某组给定的样本值 $\vec{x}$，参数 $\hat{\vec{\theta}}$ 使似然函数取得最大值，即 $\hat{\vec{\theta}} = \arg\max L(\vec{x},\vec{\theta})$，则称 $\hat{\theta}_r(\vec{x})$ 为 $\theta_r(r=1,2,\cdots,k)$ 的**最大似然估计值**，称统计量 $\hat{\theta}_r(\vec{X})$ 为 θ_r 的**最大似然估计量**。

似然方程组

若 $L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k)$ 关于 $\vec{\theta}$ 可微，则称

$$\frac{\partial}{\partial\theta_r}L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k) = 0, r = 1,2,\cdots,k$$

为 似然方程组 。$\frac{\partial}{\partial\theta_r}\log L = 0$ 为 对数似然方程组 。

多参数推广

未知参数可以不止一个，如 $\vec{\theta}$。设 X 的概率密度为 $f(x,\vec{\theta})$，则定义似然函数

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

若对于某组给定的样本值 $\vec{x}$，参数 $\hat{\vec{\theta}}$ 使似然函数取得最大值，即 $\hat{\vec{\theta}} = \arg\max L(\vec{x},\vec{\theta})$， 则称 $\hat{\theta}_r(\vec{x})$ 为 $\theta_r(r=1,2,\cdots,k)$ 的最大似然估计值，称统计量 $\hat{\theta}_r(\vec{X})$ 为 θ_r 的最大似然估计量。

似然方程组

若 $L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k)$ 关于 $\vec{\theta}$ 可微，则称

$$\frac{\partial}{\partial\theta_r}L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k) = 0, r = 1,2,\cdots,k$$

为 似然方程组 。$\dfrac{\partial}{\partial\theta_r}\log L = 0$ 为 对数似然方程组 。

例：正态总体的最大似然估计

例

设总体 $X \sim N(\mu, \sigma^2)$, $X_1, \cdots, X_n$ 是 X 的样本, 求 μ, σ^2 的最大似然估计量.

$$L(\vec{x}; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\Longrightarrow \log L = -\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2} - \frac{n}{2}\log\sigma^2 + C$$

C 是与 μ 和 σ^2 无关的常数。

$$\begin{cases} \frac{\partial}{\partial\mu}\log L = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \\ \frac{\partial}{\partial\sigma^2}\log L = \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{n}{2(\sigma^2)} = 0 \end{cases} \Longrightarrow \begin{cases} \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \\ \hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \end{cases}$$

与矩估计法一致。

例：正态总体的最大似然估计

例

设总体 $X \sim N(\mu, \sigma^2)$, $X_1, \cdots, X_n$ 是 X 的样本, 求 μ, σ^2 的最大似然估计量.

$$L(\vec{x}; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\Longrightarrow \log L = -\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2} - \frac{n}{2} \log \sigma^2 + C$$

C 是与 μ 和 σ^2 无关的常数。

$$\begin{cases} \frac{\partial}{\partial \mu} \log L = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \\ \frac{\partial}{\partial \sigma^2} \log L = \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2 - \frac{n}{2(\sigma^2)} = 0 \end{cases} \Longrightarrow \begin{cases} \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \\ \hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \end{cases}$$

与矩估计法一致。

例：正态总体的最大似然估计

例

设总体 $X \sim N(\mu, \sigma^2)$, $X_1, \cdots, X_n$ 是 X 的样本, 求 μ, σ^2 的最大似然估计量.

$$L(\vec{x}; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\Longrightarrow \log L = -\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2} - \frac{n}{2} \log \sigma^2 + C$$

C 是与 μ 和 σ^2 无关的常数。

$$\begin{cases} \frac{\partial}{\partial \mu} \log L = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i-\mu) = 0 \\ \frac{\partial}{\partial \sigma^2} \log L = \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i-\mu)^2 - \frac{n}{2(\sigma^2)} = 0 \end{cases} \Longrightarrow \begin{cases} \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x} \\ \hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \end{cases}$$

与矩估计法一致。

例：正态总体的最大似然估计

例

设总体 $X \sim N(\mu, \sigma^2)$, $X_1, \cdots, X_n$ 是 X 的样本, 求 μ, σ^2 的最大似然估计量.

$$L(\vec{x}; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\Longrightarrow \log L = -\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2} - \frac{n}{2} \log \sigma^2 + C$$

C 是与 μ 和 σ^2 无关的常数。

$$\begin{cases} \frac{\partial}{\partial \mu} \log L = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \\ \frac{\partial}{\partial \sigma^2} \log L = \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2 - \frac{n}{2(\sigma^2)} = 0 \end{cases} \Longrightarrow \begin{cases} \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x} \\ \hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \end{cases}$$

与矩估计法一致。

例：正态总体的最大似然估计

例

设总体 $X \sim N(\mu, \sigma^2)$, $X_1, \cdots, X_n$ 是 X 的样本, 求 μ, σ^2 的最大似然估计量.

$$L(\vec{x}; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\Longrightarrow \log L = -\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2} - \frac{n}{2} \log \sigma^2 + C$$

C 是与 μ 和 σ^2 无关的常数。

$$\begin{cases} \frac{\partial}{\partial \mu} \log L = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \\ \frac{\partial}{\partial \sigma^2} \log L = \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2 - \frac{n}{2(\sigma^2)} = 0 \end{cases} \Longrightarrow \begin{cases} \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x} \\ \hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \end{cases}$$

与矩估计法一致。

例：均匀分布的最大似然估计

例

设 $X \sim U(a,b)$，$X_1, X_2, \cdots, X_n$ 是 X 的一个样本，求 a, b 的最大似然估计量。

似然函数为

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n}, & a < x_i < b \\ 0, & \text{其它} \end{cases} \implies \begin{cases} \frac{1}{(b-a)^n}, & a \le \min x_i \le \max x_i \le b \\ 0, & \text{其它} \end{cases}$$

当 $a = \min x_i, b = \max x_i$ 时，$L(a,b)$ 最大。所以

$$\hat{a} = \min X_i, \hat{b} = \max X_i$$

对比矩估计：两者不一致

$$\hat{a} = A_1 - \sqrt{3B_2}, \hat{b} = A_1 + \sqrt{3B_2}$$

例：均匀分布的最大似然估计

例

设 $X \sim U(a,b)$，$X_1, X_2, \cdots, X_n$ 是 X 的一个样本，求 a, b 的最大似然估计量。

似然函数为

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n}, & a < x_i < b \\ 0, & \text{其它} \end{cases} \Longrightarrow \begin{cases} \frac{1}{(b-a)^n}, & a \le \min x_i \le \max x_i \le b \\ 0, & \text{其它} \end{cases}$$

当 $a = \min x_i, b = \max x_i$ 时，$L(a,b)$ 最大。所以

$$\hat{a} = \min X_i, \hat{b} = \max X_i$$

对比矩估计：两者不一致

$$\hat{a} = A_1 - \sqrt{3B_2}, \hat{b} = A_1 + \sqrt{3B_2}$$

例：均匀分布的最大似然估计

例

设 $X \sim U(a,b)$，$X_1, X_2, \cdots, X_n$ 是 X 的一个样本，求 a,b 的最大似然估计量。

似然函数为

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n}, & a < x_i < b \\ 0, & \text{其它} \end{cases} \implies \begin{cases} \frac{1}{(b-a)^n}, & a \le \min x_i \le \max x_i \le b \\ 0, & \text{其它} \end{cases}$$

当 $a = \min x_i, b = \max x_i$ 时，$L(a,b)$ 最大。所以

$$\hat{a} = \min X_i, \hat{b} = \max X_i$$

对比矩估计：两者不一致

$$\hat{a} = A_1 - \sqrt{3B_2}, \hat{b} = A_1 + \sqrt{3B_2}$$

最小二乘估计

特点

- 历史悠久；
- 计算高效；
- 估计量是样本的线性组合；
  - 线性组合是一类简单的函数形式，估计量的复杂性可控。
- 在线性回归中近一步讨论。