点估计

Guangyao Zhao

2022-08-18

Contents

定义

假设总体 X 的分布函数的形式已经知道，但一个或多个参数未知。借助于总体 X 的一个样本来估计总体未知参数的值的问题，称之为参数的点估计问题。

- 已知：总体 X 的分布函数的 $F(x;\theta)$ 的形式
- 未知：待估参数 θ

$X_1, X_2, ..., X_n$ 是 X 的一个样本；$x_1, x_2, ..., x_n$ 是响应的样本值。在点估计中最常用的是矩估计和最大似然估计。

矩估计

样本的 k 阶矩依概率收敛于总体的 k 阶矩。即:

$$A^k = E(X^k) = \frac{1}{n}\sum_{i=1}^{n} X^k$$

- 总体中有几个未知参数，就建立几个方程。对于均匀分布有未知数 a, b，所以需要 2 个方程；对于求解分布的 μ 和 σ^2 时需要建立 2 个方程。
- 常用公式：$E(X^2) = D(X) + E^2(X)$

例子 1

假设总体 X 在 $[a, b]$ 上服从均匀分布，试着求 a, b 的矩估计量。

因为有两个未知参数，所以需要两个方程式，即:

一阶矩:

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{b-a}{2}$$

二阶矩:

$$E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = D(X) + E^2(X) = \frac{b-a}{12} + \left(\frac{b-a}{2}\right)$$

根据以上两个方程式即可求解出 a, b:

$$\hat{a} = \bar{X} - \sqrt{\frac{3}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\hat{b} = \bar{X} + \sqrt{\frac{3}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

例子 2

尝试求均值 μ 和 方差 σ^2 的矩估计量。

$$A_1 = E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} = \mu A_2 = E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = D(X) + E^2(X) = \sigma^2 + \mu^2$$

求解得:

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

该结论适用于任意分布的数据集。

最大似然估计（Maximum Likelihood Estimation, MLE）

最大似然估计的思想很简单：在已知分布的前提下，既然抽样的时候抽到了样本值 $x_1, x_2, ..., x_n$，那么就可以感性地认为这一组样本值的概率 $L(x_1, x_2, ..., x_n; \theta_1, \theta_2, ..., \theta_m)$ 比较大（最大），因此可以求出一组 $\theta_2, ..., \theta_m$，使得概率 $L(x_1, x_2, ..., x_n; \theta_1, \theta_2, ..., \theta_m)$ 最大。求解步骤（离散）如下：

写出似然函数：

$$L(x_1, x_2, ..., x_n; \theta_1, \theta_2, ..., \theta_m) = \prod_{i=1}^{n} P\{X = x_1\}$$

因为多个数相乘时，取对数时可以轻松地化简但单调性不变，所以对数似然更加方便。对似然函数两边取对数，得到对数似然函数：

$$\ln L(x_1, x_2, ..., x_n; \theta_1, \theta_2, ..., \theta_m) = \sum_{i=1}^{n} P\{X = x_1\}$$

对每一个参数求偏导：

$$\begin{cases} \frac{\partial \ln L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_m)}{\partial \theta_1} = 0 \\ \frac{\partial \ln L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_m)}{\partial \theta_2} = 0 \\ \cdots \\ \frac{\partial \ln L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_m)}{\partial \theta_m} = 0 \end{cases}$$

求解对数似然方程：

$$\begin{cases} \theta_1 = \theta_1(x_1, x_2, \cdots, x_n) \\ \theta_2 = \theta_2(x_1, x_2, \cdots, x_n) \\ \vdots \\ \theta_m = \theta_m(x_1, x_2, \cdots, x_n) \end{cases}, \text{则} \begin{cases} \hat{\theta}_1 = \theta_1\left(X_1, X_2, \cdots, X_n\right) \\ \hat{\theta}_2 = \theta_2\left(X_1, X_2, \cdots, X_n\right) \\ \vdots \\ \hat{\theta}_m = \theta_m\left(X_1, X_2, \cdots, X_n\right) \end{cases}$$

矩估计 vs 最大似然估计

- 矩估计：不需要知道总体分布，只要求各阶矩存在即可。但是精度差
- 最大似然估计：需要提前知道分布，计算繁琐。但是精度比矩估计高，应用也比其广泛

估计量的评选标准

无偏性（估计量和真实值尽量接近）

假设 $\hat{\theta} = \theta(X_1, X_2, ..., X_n)$ 是 θ 的估计量，若 $E(\hat{\theta}) = \theta$，则称 $\hat{\theta} = \theta(X_1, X_2, ..., X_n)$ 是 θ 的无偏估计。

有效性（估计量和真实值的离散程度要小）

假设 $\hat{\theta}_1, \hat{\theta}_2$ 均为 θ 的无偏估计，若 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效。（注意前提）

相合性

相合性说的是虽然 $\hat{\theta} \neq \theta$ 但是可依概率收敛于 θ 时，就称作相合估计量。