参数估计

September 12, 2017

统计推断的基本问题分成两大类：估计问题和假设检验问题。这里讨论总体参数的点估计和区间估计。

1 点估计

设总体 X 的分布函数的形式已知，但它的一个或多个参数未知，借助总体 X 的一个样本来估计总体未知参数的值的问题称为参数的点估计问题。

点估计问题的一般提法：设总体 X 的分布函数 $F(x;\theta)$ 的形式为已知，θ 是待估参数。$X_1, X_2, \cdots, X_n$ 是 X 的一个样本，$x_1, x_2, \cdots, x_n$ 是相应的一个样本值。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 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 作为未知参数 θ 的近似值。称$\hat{\theta}(X_1, X_2, \cdots, X_n)$为 θ 的估计量，$\hat{\theta}(x_1, x_2, \cdots, x_n)$为 θ 的估计值。

1.1 矩估计法

设 X 为连续分布随机变量，其概率密度为 $f(x;\theta_1, \theta_2, \cdots, \theta_k)$，或 X 为离散型随机变量，其分布规律为 $P\{X = x\} = p(x;\theta_1, \theta_2, \cdots, \theta_k)$，其中 $\theta_1, \theta_2, \cdots, \theta_k$ 为待估参数，

$X_1, X_2, \cdots, X_n$ 是来自 X 的样本。假设总体 X 的前 k 阶矩

$$\mu_l = E(X^l) = \int_{-\infty}^{\infty} x^l f(x; \theta_1, \theta_2, \cdots, \theta_k) \mathrm{d}x \ , \ l = 1, 2, \cdots, k$$

或

$$\mu_l = E(X^l) = \sum_{x \in R_X} x^l p(x; \theta_1, \theta_2, \cdots, \theta_k) \ , \ l = 1, 2, \cdots, k$$

存在，其中 R_X 是 X 可能取值的范围。基于样本矩

$$A_l = \frac{1}{n} \sum_{i=1}^{n} X_i^l$$

依概率收敛于相应的总体矩 $\mu_l (l = 1, 2, \cdots, k)$，样本矩的连续函数依概率收敛于相应的总体矩的连续函数，用样本矩作为相应的总体矩的估计量，以样本矩的连续函数作为相应的总体矩的连续函数的估计量。这种方法称为矩估计法。设

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

这是一个包含 k 个未知参数 $\theta_1, \theta_2, \cdots, \theta_k$ 的联立方程组，可以从中解出 $\theta_1, \theta_2, \cdots, \theta_k$，得到

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

以 A_i 分别代替 $\mu_i, i = 1, 2, \cdots, k$，得到以

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

分别作为 $\theta_i, i = 1, 2, \cdots, k$ 的估计量，称为矩估计量。矩估计量的观察值称为矩估计值。

1.2 最大似然估计法

若总体 X 属离散型，其分布律 $P\{X=x\}=p(x;\theta),\theta\in\Theta$ 的形式为已知，θ 为待估参数，Θ 是 θ 的可能取值的范围。设 $X_1,X_2,\cdots,X_n$ 是来自 X 的样本，则 $X_1,X_2,\cdots,X_n$ 的联合分布律为

$$\prod_{i=1}^{n}p(x_i;\theta)\ .$$

设 $x_1,x_2,\cdots,x_n$ 是相应于样本 $X_1,X_2,\cdots,X_n$ 的一个样本值。样本 $X_1,X_2,\cdots,X_n$ 取到观察值 $x_1,x_2,\cdots,x_n$ 的概率，即事件 $\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$ 发生的概率为

$$L(\theta)=L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}p(x_i;\theta)\ ,\ \theta\in\Theta\ . \tag{1}$$

这一概率随 θ 的取值而变化，$L(\theta)$ 称为样本的似然函数($x_1,x_2,\cdots,x_n$ 是已知的样本值，是常数)。

现已经取到样本值 $x_1,x_2,\cdots,x_n$，表明取到这一样本值的概率 $L(\theta)$ 比较大。我们不会考虑那些不能使样本 $x_1,x_2,\cdots,x_n$ 出现的 $\theta\in\Theta$ 作为 θ 的估计。若已知当 $\theta=\theta_0\in\Theta$ 时使 $L(\theta)$ 取很大值，而 Θ 中的其他 θ 的值使 $L(\theta)$ 取很小值，认为取 θ_0 作为未知参数 θ 的估计值。最大似然估计法就是固定样本观察值 $x_1,x_2,\cdots,x_n$，在 θ 取值的可能范围 Θ 内挑选使似然函数 $L(x_1,x_2,\cdots,x_n;\theta)$ 达到最大的参数值 $\hat{\theta}$，作为参数 θ 的估计值。即取 $\hat{\theta}$ 使

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta}L(x_1,x_2,\cdots,x_n;\theta)\ . \tag{2}$$

得到的 $\hat{\theta}$ 与样本值 $x_1,x_2,\cdots,x_n$ 有关，记为$\hat{\theta}(x_1,x_2,\cdots,x_n)$，称为参数 θ 的最大似然估计，相应的统计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$称为参数 θ 的最大似然估计量。

若总体 X 属连续型，其概率密度 $f(x;\theta),\theta\in\Theta$ 的形式已知，θ 为待估参数，Θ 是 θ 的可能取值的范围。设 $X_1,X_2,\cdots,X_n$ 是来自 X 的样本，则 $X_1,X_2,\cdots,X_n$ 的联合密度

为

$$\prod_{i=1}^{n} f(x_i;\theta)\ .$$

设 $x_1, x_2, \cdots, x_n$ 是相应于样本 $X_1, X_2, \cdots, X_n$ 的一个样本值,则随机点 $(X_1, X_2, \cdots, X_n)$ 落在点 $(x_1, x_2, \cdots, x_n)$ 的领域 (边长分别为 $\mathrm{d}x_1, \mathrm{d}x_2, \cdots, \mathrm{d}x_n$ 的 n 维立方体) 内的概率近似为

$$\prod_{i=1}^{n} f(x_i;\theta)\mathrm{d}x_i\ . \tag{3}$$

取 θ 的估计值 $\hat{\theta}$ 使以上概率取到最大值，但因子 $\prod_{i=1}^{n}\mathrm{d}x_i$ 不随 θ 而变，考虑函数

$$L(\theta) = L(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta) \tag{4}$$

的最大值。$L(\theta)$ 称为样本的似然函数。若

$$L(x_1, x_2, \cdots, x_n; \hat{\theta}) = \max_{\theta\in\Theta} L(x_1, x_2, \cdots, x_n; \theta)\ , \tag{5}$$

则称$\hat{\theta}(x_1, x_2, \cdots, x_n)$为 θ 的最大似然估计值，称$\hat{\theta}(X_1, X_2, \cdots, X_n)$为 θ 的最大似然估计量。

$\hat{\theta}$ 可以从方程

$$\frac{\mathrm{d}}{\mathrm{d}\theta}L(\theta) = 0 \tag{6}$$

或者对数似然方程

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta) = 0 \tag{7}$$

解得。

[实验的数学处理] 寻找未知参数的适当估计值叫做点估计。若 $x_1, x_2, \cdots, x_N$ 是观测值分布 $p(x;\theta)$ 的随机样本，把 $(x_1, x_2, \cdots, x_N)$ 的联合概率 (密度) 函数叫做样本的似然函数，记为 $L(\boldsymbol{x}|\theta) = L(x_1, x_2, \cdots, x_N|\theta)$。似然函数值为 N 次测量得到观测值

$(x_1, x_2, \cdots, x_N)$ 的概率 (密度)。似然函数值为各个观测值的概率 (密度) 之积，即

$$
\begin{aligned}
L(\boldsymbol{x}|\theta) &\equiv p(x_1, x_2, \cdots, x_N; \theta) \\
&= \prod_{i=1}^{N} p(x_i; \theta) = \prod_{i=1}^{N} p(x = x_i; \theta)
\end{aligned} \tag{8}
$$

对于一组确定的观测结果 $\boldsymbol{x} = (x_1, x_2, \cdots, x_N)$，似然函数 $L(\boldsymbol{x}|\theta)$ 的数值是参数 θ 的函数。最大似然法就是选择使观测结果具有最大概率 (密度) 的参数值作为未知参数的估计值。

若估计值 $\hat{\theta}$ 使似然函数值最大，

$$
L(\boldsymbol{x}|\theta)|_{\theta=\hat{\theta}} = \max \ , \tag{9}
$$

称 $\hat{\theta}$ 为参数 θ 的最大似然估计值。最大似然估计值是样本的函数

$$
\hat{\theta} = \hat{\theta}(\boldsymbol{x}) = \hat{\theta}(x_1, x_2, \cdots, x_N) \ , \tag{10}
$$

统计量 $\hat{\theta}(\boldsymbol{x})$ 就是参数 θ 的最大似然估计量。

2 基于截尾样本的最大似然估计

3 估计量的评选标准

4 区间估计

5 正态总体均值与方差的区间估计

6 $(0-1)$ 分布参数的区间估计

7 单侧置信区间

对于给定值 $\alpha(0<\alpha<1)$,若由样本 $X_1, X_2, \cdots, X_n$ 确定的统计量 $\underline{\theta}=\underline{\theta}(X_1, X_2, \cdots, X_n)$,对于任意 $\theta \in \Theta$ 满足

$$P\{\theta>\underline{\theta}\} \geqslant 1-\alpha\ , \tag{11}$$

称随机区间 $(\underline{\theta}, \infty)$ 是 θ 的置信水平为 $1-\alpha$的单侧置信区间,$\underline{\theta}$称为θ 的置信水平为 $1-\alpha$的单侧置信下限。

若统计量 $\overline{\theta}=\overline{\theta}(X_1, X_2, \cdots, X_n)$,对于任意 $\theta \in \Theta$ 满足

$$P\{\theta<\overline{\theta}\} \geqslant 1-\alpha\ , \tag{12}$$