参数估计

- 定义：参数是刻画总体概率特性的量，当此数量未知时，通过从总体中抽取的样本对这个参数值进行推断就是参数估计。

- 方式：
  1. 点估计：直接推断未知参数的值。
  2. 区间估计：推断参数可能的取值范围，使得这个区间包含未知参数真值的概率为给定的值。

点估计

- 思想：设所求的参数为 $\theta_1, \theta_2, ..., \theta_k$ ，基于总体的一个样本 $X_1, X_2, ..., X_k$ ，构造 k 个统计量：

$$\begin{aligned} &\hat{\theta}_1(X_1, X_2, \cdots, X_n) \\ &\hat{\theta}_2(X_1, X_2, \cdots, X_n) \\ &\vdots \\ &\hat{\theta}_k(X_1, X_2, \cdots, X_n) \end{aligned}$$

  将这些统计量称为对应参数的（点）估计量。当测得样本值时，带入这些统计量可得 $\theta_1, \theta_2, ..., \theta_k$ 的估计值。

矩估计法

- 定义：对总体 k 阶矩，以样本相应 k 阶矩（原点矩或中心矩）作为其估计值；若是总体矩的连续函数，以样本的相应连续函数作为其估计量。

- 合理性：辛钦定律的推广：

$$\frac{1}{n}\sum_{i=1}^{n} X_i^r \xrightarrow{P} E(X^r)$$

  即

$$A_r \xrightarrow{P} \mu_r$$

  其中 μ_r 为总体的真实数字特征，A_r 为样本的对应特征，在这里就是矩。对于依概率收敛的序列，有：

$$g(A_1, A_2, \cdots, A_k) \xrightarrow{P} g(\mu_1, \mu_2, \cdots, \mu_k)$$

- 求解方式：设待估计的参数值为 $\theta_1, \theta_2, ..., \theta_k$ 。设样本的 r 阶矩存在，并可用待求参数的函数来表示，即：

$$E(X^r) = \mu_r(\theta_1, \theta_2, \cdots, \theta_k)$$

  。我们用样本的相应 r 阶矩估计总体的 r 阶矩的值，得到

$$\mu_r(\theta_1, \theta_2, \cdots, \theta_k) = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

  令 r = 1,2,...,k ，就得到了 k 个方程，解这个方程组就得到了 k 个未知参数的估计值。

极大似然估计

- 似然函数：设总体的密度函数（或概率分布）为 $f(x, \theta)$ ，其中 θ 是参数。对于来自总体的简单随机样本 $(X_1, X_2, \cdots, X_n)$ 其联合密度函数为：

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

当样本的观测值给定后，联合分布函数退化为 $L(\theta)$ ，称其为样本的似然函数。

- 极大似然估计：选择适当的 θ ，使得 $L(\theta)$ 取最大值。称这样得到的 $\overset{\frown}{\theta}$ 为参数 θ 的极大似然估计值。对于有多个参数的情形与之类似，通过取最大值找出对应参数的估计值。
  NOTE: 不存在最大值时取极大值

- 合理性：一次试验（取一次样本）中，所出现的事件具有较大的概率。

- 求解方式：求导（若是多个参数则对每个参数求偏导）为 0 。称求导的方程为似然方程。也可以求 $ln(L(\theta))$ 的最大值，称为对数似然方程。

估计量的评价

对于同一个未知参数，不同方式得到的估计量可能不相同，于是选取估计量时有几个标准：

1. 无偏性
2. 有效性
3. 一致性

无偏性

- 定义：设 $(X_1,X_2,\cdots,X_n)$ 是总体 X 的样本，

  $$\hat{\theta}(X_1,X_2,\cdots,X_n)$$

  是总体参数 θ 的估计值，如果 $E(\hat{\theta})=\theta$ ，则称 $\hat{\theta}$ 是 θ 的无偏估计值。

- 结论：
  1. 无论 X 服从什么分布，$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 是总体的 k 阶矩 μ_k 的无偏估计量。
     - 证明：

       $$E(\frac{1}{n}\sum_{i=1}^{n}X_i^k)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^k)=\mu_k$$

  2. $S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 不是 $D(X)=\sigma^2$ 的无偏估计量。

  3. $S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 不是 $D(X)=\sigma^2$ 的无偏估计量。

  4. 设 $S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ ，$\sqrt{S}$ 不是 $\sqrt{D(X)}=\sigma$ 的偏小估计量。
     - 证明：事实上，设

       $$X\sim N(\mu,\sigma^2)$$

       则

       $$Y=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$$

       对$\sqrt{Y}$ 直接求期望，即

       $$\begin{aligned}E(\sqrt{Y})&=\int_0^{+\infty}\sqrt{y}\cdot\frac{1}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})}y^{\frac{n-1}{2}-1}e^{-\frac{y}{2}}dy\\&=\frac{\sqrt{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\end{aligned}$$

       于是

$$E(S)=\frac{\sqrt{2}\Gamma(\frac{n}{2})}{\sqrt{n-1}\Gamma(\frac{n-1}{2})}\sigma<\sigma$$

有效性

- 定义：设 $\hat\theta_1$ 和 $\hat\theta_2$ 都是 θ 的无偏估计量，若

$$D(\hat\theta_1)<D(\hat\theta_2)$$

  则称 $\hat\theta_1$ 比 $\hat\theta_2$ 更有效。

一致性

- 定义：设 $\hat\theta=\hat\theta(X_1,X_2,\cdots,X_n)$ 是总体参数 θ 的估计量，若对任意待估 $\theta\in\Theta$ ，当 $n\to+\infty$ 时，$\hat\theta$ 依概率收敛于 θ ，即 $\forall\epsilon>0$ ，

$$\lim_{n\to\infty}P(|\hat\theta-\theta|<\epsilon)=1$$

  或

$$\lim_{n\to\infty}P(|\hat\theta-\theta|\geq\epsilon)=0$$

  则称 $\hat\theta$ 是总体参数 θ 的一致（或相合）估计量。

- 意义：
  1. 一致估计量表明当 n 很大时，$\hat\theta$ 与 θ 非常接近的可能性很大。
  2. 一致估计量只有在样本容量足够大时，才能显示出优越性。

- 例子：
  1. S^2 是 σ^2 的一致估计量。
     - 证明：$E(S^2)=\sigma^2$，$D(S^2)=\frac{2\sigma^4}{n-1}$ 因此

$$P\{|S^2-\sigma^2|\geq\epsilon\}\leq\frac{2\sigma^4}{(n-1)\epsilon^2}\to0$$

  2. B_2 估计 σ^2 ，不是无偏却是一致估计量。

  3. S 估计 σ ，不是无偏却是一致估计量。
     - 证明：对于依概率收敛，有以下结论：若

$$\xi_n\xrightarrow{P}\xi$$

       则对于连续函数 $f(x)$ ，有

$$f(\xi_n)\xrightarrow{P}f(\xi)$$

       在这里只需要令 $f(x)=\sqrt{x}$ ，然后利用 S^2 是 σ^2 的一致估计量即可。
- 结论：
  1. 样本 k 阶矩是总体 k 阶矩的一致性估计量。
     - 证明：由大数定律可得。

  2. 设 $\hat\theta$ 是 θ 的无偏估计量，且：

$$\lim_{n\to\infty}D(\hat\theta)=0$$

     则 $\hat\theta$ 是 θ 的一致估计量。

  3. 矩估计法得到的一般是一致估计量。在一定条件下，极大似然估计也具有一致性。