4.3: 点估计的优良准则

金百锁

第四章：参数估计

4.3 点估计的优良准则

我们看到对同一个参数，有多个不同的估计量，因此，评选不同估计量的优劣性是需要考虑的。

4.3.1 无偏性

设 $\hat{g}(X_1, \cdots, X_n)$ 为待估参数函数 $g(\theta)$ 的一个估计量，若

$$E\hat{g}(X_1, \cdots, X_n) = g(\theta)$$

则称 $\hat{g}(X_1, \cdots, X_n)$ 为 $g(\theta)$ 的**无偏估计量** (Unbiased Estimator) 。无偏性的实际意义就是无系统误差. 因此在有多个估计量可供选择时，我们优先考虑无偏估计量。

很多时候我们得到的估计量是有偏，例如正态总体的方差 σ^2 的最大似然估计量 $\hat{\sigma}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$ 是有偏的，$E\hat{\sigma}^2 = \frac{n-1}{n}\sigma^2$. 若以 $\frac{n}{n-1}$ 乘以 $\hat{\sigma}^2$, 所得到的估计量就是无偏的. 这种方法称为修正.

若某一参数存在多个无偏估计时，如何来选择使用哪个估计量？人们又在无偏性的基础上增加了对方差的要求.

4.3.2 有效性

设 $\hat{g}_1(X_1, \cdots, X_n)$ 和 $\hat{g}_2(X_1, \cdots, X_n)$ 为待估参数函数 $g(\theta)$ 的两个不同的无偏估计量，若对任意的 $\theta \in \Theta$, 有

$$Var(\hat{g}_1(X_1, \cdots, X_n)) \leq Var(\hat{g}_2(X_1, \cdots, X_n))$$

而且至少对某个 $\theta_0 \in \Theta$ 使得严格不等式成立。则称 $\hat{g}_1$ 较 $\hat{g}_2$ 有效。

设 $X_1, \cdots, X_n$ 为从如下分布中抽取的简单样本，试证明样本方差为总体方差的无偏估计.

↑Example

↓Example

证: 显然

$$
\begin{aligned}
ES^2 &= E\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(E(X_i - EX_i + EX_i - \bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}[(E(X_i - EX_i)^2 - E(EX_i - \bar{X})^2] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}[\sigma^2 - \sigma^2/n)^2 = \sigma^2.
\end{aligned}
$$

设总体 X 服从 $(0,\theta)$ 上的均匀分布, $0<\theta$, 求参数 θ 的最大似然估计是否为无偏估计.

↑Example
↓Example

解: 易得似然函数为

$$L(\theta)=\frac{1}{\theta^n}\prod_{j=1}^{n} I(0\le x_j\le\theta)=\frac{1}{\theta^n} I(0\le x_{(1)}\le x_{(n)}\le\theta).$$

于是似然函数 $L(\theta)$ 在 $\theta=x_{(n)}$ 时取到最大值. 而 $X_{(n)}$ 的密度函数为

$$f(t)=\frac{nt^{n-1}}{\theta^n} I(0<t<\theta).$$

因此

$$EX_n=\int_0^{\theta} tf(t)dt=\frac{n}{n+1}\theta.$$

即 θ 的最大似然估计量 $X_{(n)}$ 不是 θ 的无偏估计, 但 $\frac{n+1}{n}X_{(n)}$ 为 θ 的无偏估计量.

设 $X_1,\ldots,X_n$ 来自均值为 μ, 方差为 σ^2 的总体分布的简单样本, $\omega_1,\ldots,\omega_n$ 为已知的非负权值, 且满足 $\sum\omega_i=1$, 试比较 μ 的两个估计估计 $\bar{X}$ 和 $\sum_{i=1}^n\omega_iX_i$.

↑Example
↓Example

解: 因为

$$Var(\bar{X})=\frac{\sigma^2}{n},\qquad Var(\sum\omega_iX_i)=\sum_{i=1}^n\omega_i^2\sigma^2,$$

所以

$$Var(\sum\omega_iX_i)\geq Var(\bar{X})$$

且等号成立当且仅当 $\omega_i=\frac{1}{n}$.

设总体分布服从 $U(0,\theta)$, 试比较 θ 的矩估计 $\hat{\theta}_M=2\bar{X}$ 和修正的最大似然估计 $\hat{\theta}_L=((n+1)/n)X_{(n)}$ 的有效性。

↑Example
↓Example

解：由前面例题, 对 $(0,\theta)$ 上的均匀分布, 有

$$E\hat{\theta}_M = 2E\bar{X} = \theta, \quad \operatorname{Var}\hat{\theta} = 4\frac{\theta^2}{12n} = \frac{\theta^2}{3n}$$
$$E\hat{\theta}_L = ((n+1)/n)EX_{(n)} = \theta$$
$$EX_{(n)}^2 = \int_0^\theta x^2 \frac{nx^{n-1}}{\theta^n}dx = \frac{n}{n+2}\theta^2$$

所以 $\hat{\theta}_L$ 的方差为

$$\operatorname{Var}\hat{\theta}_L = \left(\frac{n+1}{n}\right)^2 EX_{(n)}^2 - \left(E\hat{\theta}_L\right)^2 \tag{4.1}$$
$$= \frac{(n+1)^2}{n^2}\frac{n}{n+2}\theta^2 - \theta^2 = \frac{1}{n(n+2)}\theta^2 \tag{4.2}$$

当 $n>1$ 时, $n(n+2)>3n$ 对一切 θ 都成立, 所以 $\hat{\theta}_L$ 比 $\hat{\theta}_M$ 更有效。

4.3.3 克拉美-劳方差下界

设 $\mathbf{X} = (X_1, \cdots, X_n)$ 是总体分布 $f(x, \theta)$ 的一个样本, $\hat{g}(\mathbf{X})$ 是 $g(\theta)$ 的任一无偏估计量, 取

$$S(\mathbf{X}, \theta) = \sum_{i=1}^{n} \frac{\partial \ln f\left(X_i, \theta\right)}{\partial \theta}$$

有如下结论:

$$\mathrm{ES}(\mathbf{X}, \theta) = 0,$$

$$\operatorname{Var} S(\mathbf{X}, \theta) = n \int_{-\infty}^{\infty} \left[\frac{\partial f(x, \theta)}{\partial \theta} / f(x, \theta) \right]^2 f(x, \theta) dx \equiv n I(\theta)$$

$$\operatorname{Cov}(\hat{g}(\mathbf{X}), S(\mathbf{X}, \theta)) = g'(\theta),$$

因为

$$[\operatorname{Cov}(\hat{g}(\mathbf{X}), S(\mathbf{X}, \theta))]^2 \leq \operatorname{Var} S(\mathbf{X}, \theta) \operatorname{Var} \hat{g}(\mathbf{X}).$$

$$\operatorname{Var}(\hat{g}(\mathbf{X})) \geq \left(g'(\theta)\right)^2 [n I(\theta)]^{-1}$$

这就是克拉美-劳方差下界.

设 $\hat{\theta}$ 是 $g(\theta)$ 的一个无偏估计，若对 $g(\theta)$ 的任一无偏估计 $\hat{\theta}_1$，都有

$$\operatorname{Var}(\hat{\theta}) \leq \operatorname{Var}\left(\hat{\theta}_1\right), \quad \forall \theta \in \Theta,$$

则称 $\hat{\theta}$ 是 $g(\theta)$ 的一个最小方差无偏估计 **(Minimum Variance Unbiased Estimate，简称 MVUE)**

Definition

设 $(X_1, \cdots, X_n)$ 是从总体 $N\left(\mu, \sigma^2\right)$ 中抽取的一个样本，其中 σ^2 未知。求 μ 的 MUVE.

↑Example

↓Example

解: 对 μ 作点估计。这儿密度函数为

$$f(x,\mu) = (\sqrt{2\pi}\sigma)^{-1}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

所以

$$I(\mu) = (\sqrt{2\pi}\sigma)^{-1}\int_{-\infty}^{\infty}\frac{1}{\sigma^4}(x-\mu)^2\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}dx$$
$$= \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2}$$

由克拉美-劳方差下界不等式, 任一 μ 的无偏估计量的方差不能小于 σ^2/n, 而 $\bar{X}$ 是 μ 的无偏估计, 方差恰为 σ^2/n, 故 $\bar{X}$ 是 μ 的 MUVE.

4.3.4 相合性

设总体分布依赖于参数 $\theta_1, \cdots, \theta_k, g(\theta_1, \cdots, \theta_k)$ 是待估参数函数。设 $X_1, \cdots, X_n$ 为自该总体中抽取的样本，$T(X_1, \cdots, X_n)$ 为 $g(\theta_1, \cdots, \theta_k)$ 的一个估计量，如果对任意的 $\epsilon > 0$ 和 $\theta_1, \cdots, \theta_k$ 的一切可能值都有

$$\lim_{n \to \infty} P_{\theta_1, \cdots, \theta_k}(|T(X_1, \cdots, X_n) - g(\theta_1, \cdots, \theta_k)| \geq \epsilon) = 0$$

我们则称 $T(X_1, \cdots, X_n)$ 为 $g(\theta_1, \cdots, \theta_k)$ 的一个 (弱)**相合估计量** (Consistent Estimator)。

相合性是对一个估计量的最基本的要求，如果一个估计量没有相合性，那么无论样本大小多大，我们也不能把未知参数估计到任意预定的精度。这种估计量显然是不可取的。

矩估计量是满足相合性的，最大似然估计量在很一般的条件下也是满足相合性的。

4.3.5 渐近正态性

估计量是样本 $X_1, \cdots, X_n$ 的函数，其确切的分布一般不是容易得到。但是，许多形式很复杂的统计量 (未必是和)，当 n 很大时，其分布都渐近于正态分布，这个性质称为统计量的 “渐近正态性”。

无偏性和有效性都是对固定的样本大小 n 而言的，这种性质称为估计量的 “小样本性质”，而相合性和渐近正态性都是考虑在样本大小趋于无穷时的性质，这种性质称为 “大样本性质”。

↑Example

设从总体

X	0	1	2	3
P	$\theta/2$	θ	$3\theta/2$	$1-3\theta$

抽取的一个简单样本 $X_1,\cdots,X_{10}$ 的观察值为 $(0,3,1,1,0,2,0,0,3,0)$,
(1) 求 θ 的矩估计量 $\hat{\theta}_M$ 和最大似然估计量 $\hat{\theta}_L$，并求出估计值。
(2) 上述估计量是否为无偏的？若不是，请作修正.
(3) 比较修正后的两个估计量，指出那个更有效.

↓Example

由有效性的定义，我们自然会问在一切可能的无偏估计里，能否找到具有最小方差的无偏估计量？如果存在这样的估计量，我们称其为最小方差无偏估计量, 详细地可以参考课本。