第 8 讲*

点估计量和点估计方法

麻省理工学院 14.30 2006 年春季

Herman Bennett

假定某一参数值未知的情况下，点估计的目标就是用一个样本计算出一个数值，该数值在某种意义上代表对该参数真实值的优良估计。

19 定义

19．1 参数

概率质量/密度函数可以记作 $f_X(x)$ 或 $f_X(x/\theta)$ ，其中θ代表能全定义分布的常数。例如，如果x是正态分布的随机变量，则常数μ和σ将能完全定义其分布状态。这些常数被称为参数，并且通常被记作希腊字母θ。[1]

例 19. 1：

—正态分布：$f_X(x/\theta) = f(x/\mu,\sigma)$，两个参数：$\theta_1=\mu,\theta_2=\sigma$；

—二项分布：$f_X(x/\theta) = f(x/n,p)$，两个参数：$\theta_1=n,\theta_2=p$ ；

—泊松分布：$f_X(x/\theta) = f(x/\lambda)$，一个参数：$\theta=\lambda$；

—伽玛分布：$f_X(x/\theta) = f(x/\alpha,\beta)$，两个参数：$\theta_1=\alpha,\theta_2=\beta$ 。

19.2（点）估计量

<u>θ的点估计量</u>，记作$\hat{\theta}$，它是一个统计量（是随机样本的一个函数）：

$$\hat{\theta}=r(X_1,X_2,\ldots,X_n) \qquad (63)$$

注意：这些讲义不一定是自封的。它们只是对讲座的一种补充而不是替代。

[1]对此也可以这样解释：参数是一个在同一个分布族中都适用的常数。

- 注意$\hat{\theta}$的值并不是直接取决于 θ，而仅仅是间接取决于每一个X_i的随机过程。

- θ<u>的点估计</u>是一种估计量$\hat{\theta}$的实现值（即随机样本实现值的一个函数）：

$$\hat{\theta} = r(x_1, x_2, \ldots, x_n)\text{。} \tag{64}$$

例 19.2 设样本$X_1, \ldots, X_{10}$服从正态分布$\mathrm{N}(\mu, \sigma^2)$，我们想估计参数μ（未知）。我们可以构造出无数个μ的估计量。实际上，任何一个随机样本的函数都可以看作是μ的一个估计量，例如：

$$\hat{\theta} = r(X_1, X_2, \ldots, X_{10}) = \begin{cases} X_{10} \\ 2 \\ \dfrac{X_{10} + X_1}{2} \\ \overline{X} = \dfrac{1}{10}\sum_{i=1}^{10} X_i \\ 1.5X_2 \\ etc. \end{cases}$$

例19.3 设随机样本$X_1, \ldots, X_n$服从均匀分布$U(0, \theta)$，其中θ未知。计算出它的3个不同的估计量。

20 （点）估计量的评价

由于有许多可能的估计量，我们需要定义一些性质对它们进行评价和排序。

20.1 无偏性

如果对于θ的每一个可能值，都有：

$$E(\hat{\theta}) = \theta \tag{65}$$

则称估计量$\hat{\theta}$是<u>参数θ的无偏估计量</u>。

如果$\hat{\theta}$不是无偏的，则称它是一个<u>有偏估计量</u>，其中差值$E(\hat{\theta}) - \theta$称作$\hat{\theta}$的偏差。

- 如果随机样本$X_1, \ldots, X_n$独立同分布，且$E(X_i) = \theta$，则样本均值的估计量

$$\hat{\theta} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{66}$$

是总体均值的无偏估计量：$E(\overline{X}_n) = \theta$（见第7讲中的例18.1）

- 如果随机样本$X_1, \ldots, X_n$独立同分布, 且$E(X_i) = \mu$, $Var(X_i) = \theta$，则样本方差的估计量

$$\hat{\theta} = S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \tag{67}$$

是总体方差的无偏估计量：$E(S^2) = \theta$（见第7讲中的例18.1）

例20.1 设X_i服从$U[0, \theta]$。随机样本容量为n，定义θ的一个估计量如下：

$$\hat{\theta} = \frac{2}{n}\sum_{i=1}^{n} X_i \text{。}$$ $\hat{\theta}$**是有偏的吗？**

20.2 有效性

设$\hat{\theta}_1$与$\hat{\theta}_2$是θ的无偏估计量。如果对于给定容量为n的样本，

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2) \quad (68)$$

则称$\hat{\theta}_1$比$\hat{\theta}_2$更有效，其中$Var(\hat{\theta}_i)$是估计量的方差。

设$\hat{\theta}_1$是θ的一个无偏估计量。如果对于θ的任何无偏估计量，$\hat{\theta}_k$，

$$Var(\hat{\theta}_1) \le Var(\hat{\theta}_k), \quad (69)$$

则称$\hat{\theta}_1$是有效的，或者是方差最小的无偏估计量。

- 不要将估计量$\hat{\theta}$的方差-$Var(\hat{\theta})$，与样本方差估计量S^2相混淆，S^2是总体方差σ^2 (!)的无偏估计量。

例20.2 如何比较例19.2中估计量的有效性?这些估计量中哪些是无偏的?

20.3 均方误差

为什么强调无偏估计量呢？均方误差（MSE）用于对每一个估计量$\hat{\theta}$在偏差和有效性之间进行权衡。

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + (bias(\hat{\theta}))^2 \quad \textbf{(70)}$$

在θ的所有可能估计量中，对于给定容量为n的样本，如果$\hat{\theta}$有最小的均方误差，则$\hat{\theta}$称为θ的最小均方误差估计量。

例20.3 画图描绘两个估计量的概率密度函数,使第一个估计量的偏差小于其无效性(另一个估计量反之)。

20.4 渐近准则

20.4.1 一致性

设$\hat{\theta}$是θ的一个估计量。如果$\hat{\theta} \xrightarrow{p} \theta$，则$\hat{\theta}$具有一致性（大数定律：讲义第 7 讲）。

例 20.4 设来自总体 $f(x)$ 容量为 n 的随机样本，其中 $E(X)=\mu$（未知）。下列估计量中哪个具有一致性？

$$\hat{\mu}_1=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad \hat{\mu}_2=\frac{1}{n-5}\sum_{i=1}^{n}X_i \qquad \hat{\mu}_3=\frac{1}{n-5}\sum_{i=1}^{n-5}X_i$$

- 当$n\to\infty$时，MSE$\to 0$ $\Rightarrow$一致性

20.4.2 渐近有效性

设$\hat{\theta}_1$是θ的一个估计量。当$n\to\infty$时，如果$\hat{\theta}_1$满足有效估计量定义，则称它渐近有效。

21 点估计方法

下面介绍两种用于构造(点)估计量的标准方法。

21.1 矩法估计（MM）

设 $X_1, X_2, \ldots, X_n$ 是来自总体概率质量/密度函数为 $f(x/\theta_1,\ldots,\theta_k)$ 的随机样本，其中 $\theta_1,\ldots,\theta_k$ 是未知参数。一种估计这些参数的方法就是使前k个总体矩与对应的k个样本矩相等。合成k估计量被称作参数 $\theta_1,\ldots,\theta_k$ 的矩法估计值。[2]

计算过程概括如下：

方程系统：
k个方程与k个未知数
$\Rightarrow$ 输出 $\hat{\theta}_1,\ldots,\hat{\theta}_k$
注意
(1）$E(X_i^j)=g_j(\theta_1,\ldots,\theta_k)$
(2) 样本矩的实现值是一个标量

$$\begin{cases} \text{样本矩} & & \begin{array}{c}\text{总体矩}\\ \text{(理论上)}\end{array} & \\ \frac{1}{n}\sum_{i=1}^{n} X_i & = & E(X_i^1) & \text{一阶矩} \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 & = & E(X_i^2) & \text{二阶矩} \\ \frac{1}{n}\sum_{i=1}^{n} X_i^3 & = & E(X_i^3) & \text{三阶矩} \\ \vdots & \vdots & \vdots & \\ \frac{1}{n}\sum_{i=1}^{n} X_i^k & = & E(X_i^k) & k\text{阶矩} \end{cases}$$

[2]此估计方法是由卡尔. 皮尔逊于 1894 年提出的。

例 21.1 设容量为 n 的随机样本服从总体 $N(\mu,\sigma^2)$，其中 μ,σ^2 为未知参数。计算两个参数的矩法估计值。

例21.2 设容量为 n 的随机样本服从伽玛分布：$f(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}e^{-x/\beta}$，对于 $0<x<\infty$。假定随机样本的实现值为 $\frac{1}{n}\sum_{i=1}^{n}x_i=7.29$，$\frac{1}{n}\sum_{i=1}^{n}x_i^2=85.59$，计算参数 α 和 β 的矩法估计值。记住 $E(X_i)=\alpha\beta$ 和 $Var(X_i)=\alpha\beta^2$。

21.2 极大似然估计（MLE）

设 $X_1, X_2, \ldots, X_n$ 是来自总体概率质量/密度函数为 $f(x/\theta_1, \ldots, \theta_k)$ 的随机样本，其中 $\theta_1, \ldots, \theta_k$ 是未知参数。估计这些参数的另一种方法就是找到由 $f(x/\hat{\theta}_1, \ldots, \hat{\theta}_k)$ 产生的样本观测值出现的概率为最大的 $\hat{\theta}_1, \ldots \hat{\theta}_k$ 的值。

随机样本的联合概率密度函数， $f(x_1, x_2, \ldots, x_n / \theta_1, \ldots, \theta_k)$，称作似然函数，记作 $L(\theta / x)$。

$$L(\theta/x) = L(\theta_1, \ldots, \theta_k / x_1, \ldots, x_n) = \begin{cases} f(x_1, \ldots, x_n / \theta_1, \ldots \theta_k) & \text{一般情况} \\ \prod_{i=1}^{n} f(x_i / \theta_1, \ldots, \theta_k) & \text{随机样本(独立同分布)} \end{cases}$$

其中 θ 和 x 都是向量，$\theta = (\theta_1, \ldots \theta_k)$，$x = (x_1, \ldots, x_k)$.

对于一个给定的样本向量X，当 $L(\theta / x)$ 达到最大值时，θ 的参数值记作 $\hat{\theta}_{MLE}(x)$。那么 $\hat{\theta}_{MLE}(x)$ 称作未知参数 $\theta_1, \ldots, \theta_k$ 的最大似然估计值**(MLE)**。[3]

- 直观认识:离散型分布的情形

[3] 该估计方法是由R. A. 费舍于 1912 年提出的。

- 存在总体最大值吗?能找到吗?唯一吗?. 参考微积分101:

$$foc:\frac{\partial L(\theta/x)}{\partial\theta_i}=0,\qquad i=1,\ldots k$$ （对于一个良态函数来说）

你需要检验它是否真的是最大值而并非最小值. （求二阶导数验证）

- 多数情况下，较容易找到 $LnL(\theta/x)$ 的最大值（既然是一个单调变换最大值唯一……参考微积分101）。

- **MLE**的不变性: $\hat{\tau}_{MLE}(\theta)=\tau(\hat{\theta}_{MLE})$ 。

- 对于大样本来说, **MLE**可以得到θ的一个非常优良的估计量(满足一致性和渐进有效)。毫无疑问，此种方法被广泛运用。

- 但是…

（1） 数值灵敏度(稳健性)

（2） 不一定是无偏的

（3） 可能难以计算

例21.3 设一个随机样本来自服从正态分布 $N(\mu,\sigma^2)$ **的总体**, 其中参数 μ,σ^2 未知。计算参数的极大似然估计值。

例 **21.4** 设随机样本来自服从 $U(0,\theta)$ 分布的总体, θ 未知。计算 $\hat{\theta}_{MLE}$ 。