点估计 (Point estimation)

1. 简介及定义 Introductions and definitions

2. 矩估计法 The Method of Moment

总体的 k 阶原点矩 = 样本的 k 阶原点矩

总体的 k 阶中心矩 = 样本的 k 阶中心矩

3. 极大（最大）似然估计 Maximum Likelihood Estimation

复习总体和随机样本的概念

考虑一个随机试验：令 X 为该随机试验对应的随机变量，通常称为总体。其元素通常称为个体 (individuals)
重复该随机试验 n 次，令 X_k 为第 k 次试验对应的随机变量.

这些随机变量的 集合$\{X_1, X_2, \dots, X_n\}$ 就称为总体X的一个容量为 n 的随机样本，简记为 $X_1, X_2, \dots, X_n$

随机样本 $X_1, X_2, \dots, X_n$ 是独立同分布的的(independent and identically distributed): 如果 $F(x)$ 是总体 X 的分布函数, 那么 $F(x)$ 也是 X_i 的分布函数

每一个 X_i 通常称为总体 X 的一次观察（observation）

随机样本的应用

考虑总体的均值 μ 及方差 σ^2:

点估计的目标就是应用随机样本计算出一个数，在一定程度上，表示参数的估计值

The objective of point estimation is to use a sample to compute a number that represents, in some sense, an educated guess for the true value of the parameter.

该结果就称为点估计.

1. 简介.

定义 1：点估计 point estimate

假设总体 X 的分布函数为 $F(x;\theta)$， θ 是未知参数.

$X_1, X_2, \dots, X_n$ 是总体 X 的随机样本, 观察值分别为 $x_1, x_2, \dots, x_n$,

$\hat{\theta}(X_1, X_2, \dots, X_n)$ 是一个统计量， 它的观察值 （**observed value** ）称为 θ 的一个估计, 称为 θ 的点估计.

$\hat{\theta}(X_1, X_2, \dots, X_n)$ 称为θ的估计量 (estimator), 这是一个随机变量

$\hat{\theta}(x_1, x_2, \dots, x_n)$ 称为θ的估计值 （estimate） , 这是一个值

定义 2 点估计 参数 θ 的点估计是一个数， 可以视为θ的一个合理的（sensible）值。

先建立一个合适的统计量， 再由一组样本值计算出统计量的值

估计量的评价标准 1　无偏估计（Unbiased Estimators）

无偏性：若估计量 $\hat{\theta}$（$X_1, X_2, \ldots, X_n$）的数学期望等于未知参数，即 $E(\hat{\theta}) = \theta$. 则称 $\hat{\theta}$ 为θ的无偏估计量（Unbiased Estimator）

如果 $\hat{\theta}(X_1, X_2, \ldots, X_n)$是有偏的，则差 $E(\hat{\theta}) - \theta$ 称为$\hat{\theta}$ 的偏差(bias)

例: $X_1, X_2, \ldots, X_n$ 是来自总体 X 的随机样本　$E(X) = \mu.$

样本均值 $\bar{X}$ 可以作为 μ 的一个无偏估计.

X_1 能作为 μ 的无偏估计吗？　YES

例: $X_1, X_2, \dots, X_n$ 是来自总体 X 的随机样本。 $D(X)=\sigma^2$.

$$\text{样本方差}\ S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)$$

是总体方差 σ^2的无偏估计。

$E(X_i)=\mu, D(X_i)=\sigma^2=E\left(X_i^2\right)-E^2(X_i)$。 得 $E\left(X_i^2\right)=\sigma^2+\mu^2$。

$E(\bar{X})=\mu, D(\bar{X})=\dfrac{\sigma^2}{n}=E(\bar{X}^2)-E^2(\bar{X})$。 得 $E(\bar{X}^2)=\dfrac{\sigma^2}{n}+\mu^2$。

$$E(S^2)=\frac{1}{n-1}\left(\sum_{i=1}^{n}E\left(X_i^2\right)-nE(\bar{X}^2)\right)$$

$$=\frac{1}{n-1}(n\sigma^2+n\mu^2-\sigma^2-n\mu^2)=\sigma^2$$

注: $X_1, X_2, \dots, X_n$ 是来自总体 X 的随机样本，总体的方差为 σ^2.

则估计量

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \quad \text{是有偏估计.}$$

$$E(\hat{S}^2) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\sigma^2.$$

偏差 $E(\hat{S}^2) = -\frac{1}{n}\sigma^2 < 0$.

估计量 $\hat{S}^2$ 通常欠估计σ^2， tends to underestimate σ^2.

这就是为什么很多统计学家更加青睐 S^2 的原因(though when n is large, the bias is small and there is little difference between the two).

如何选择一个统计量?

Principle of unbiased Estimation:

1. 首先选择一个无偏估计量

2. 如果 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是θ无偏估计量（他们的均值皆为θ），但是偏离中心的程度可能不同.

最小方差无偏估计准则 Principle of Minimum Variance Unbiased Estimation:,

在无偏估计量中，选择一个具有最小方差的估计量。

方差越小，波动越小，越稳定，越有效

或者称为有效性

例: $X_1, X_2, \ldots, X_n$ 是来自总体 X 的随机样本 $E(X) = \mu$.

样本均值 $\bar{X}$ 可以作为 μ 的一个无偏估计.

X_1 能作为 μ 的无偏估计吗？ YES

$\bar{X}$ 和 X_1哪个更有效?

$$\mathrm{D}(\bar{X}) = \frac{\sigma^2}{n} \leq \mathrm{D}(X_i) = \sigma^2,$$

显然， $\bar{X}$ 更有效。

统计量评价标准：一致性

无偏性、有效性都是在样本容量 n 一定的条件下进行讨论的，然而 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 不仅与样本值有关，而且与样本容量 n 有关，不妨记为 $\hat{\theta}_n$，很自然地，我们认为 n 越大时，$\hat{\theta}_n$ 对 θ 的估计应该越精确.

定义 7.4 如果 $\hat{\theta}_n$ 依概率收敛于 θ，即 $\forall \varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|\hat{\theta}_n - \theta| < \varepsilon\} = 1, \tag{7-8}$$

则称 $\hat{\theta}_n$ 是 θ 的一致估计量(uniform estimator).

由辛钦大数定律可以证明：样本平均数 $\overline{X}$ 是总体均值 μ 的一致 估计量

样本方差 S^2 及二阶样本中心矩 B_2 都是总体方差 σ^2 的一致估计量.

2.矩估计法 The Method of Moment

$X_1, X_2, \dots, X_n$ 是来自总体 X 的简单随机样本

$X \sim F(x; \theta_1, \theta_2, \cdots, \theta_m)$, 这里 $\theta_1, \theta_2, \cdots, \theta_m$ 是需要估计的量

矩估计法的基本思想: 令总体的k 阶矩 $E(X^k)$ **等于**样本的 k 阶矩 A_k

$$\begin{cases} E(X) = A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \\ E(X^2) = A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ \cdots \\ E(X^m) = A_m = \frac{1}{n}\sum_{i=1}^{n} X_i^m \end{cases}$$

方程组的解: $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$;

$\hat{\theta}_k = \hat{\theta}_k(X_1, X_2, \dots, X_n)$ 称为 θ_k 的矩估计量

$\hat{\theta}_k = \hat{\theta}_k(x_1, x_2, \dots, xX_n)$ 称为 θ_k的矩估计值

例: 设总体 X 的概率密度函数为

$$f(x)=\begin{cases}(\alpha+1)x^{\alpha}, & 0<x<1 \\ 0, & \text{其他}\end{cases}$$ 这里 $\alpha>-1$ 未知.

$X_1, X_2, \ldots, X_n$ 是来自总体 X 的随机样本．求 α 的矩估计量.

解析: 含有一个未知数，只需要一阶矩。令 $E(X)=A_1$.

$$E(X)=\int_{-\infty}^{\infty} x f(x) dx=\int_0^1 x(\alpha+1) x^{\alpha} dx=\frac{\alpha+1}{\alpha+2},$$ 由 $A_1=\bar{X}$

得 $$\frac{\alpha+1}{\alpha+2}=\bar{X}$$ 解得 $$\alpha=\frac{2\bar{X}-1}{1-\bar{X}}$$

故 α 的矩估计量为 $$\hat{\alpha}=\frac{2\bar{X}-1}{1-\bar{X}}$$

例: 设总体 $X \sim N(\mu,\sigma^2)$, 其中 μ,σ^2 未知。$X_1,X_2,\dots,X_n$ 为来自总体 X 的样本. 求 μ 和 σ^2 的矩估计量.

解析: 含有2个未知参数，需要一阶矩、二阶矩.

$$E(X)=A_1=\bar{X}.$$

$$E(X^2)=A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2.$$

$$E(X)=\mu. \qquad E(X^2)=D(X)+E^2(X)=\sigma^2+\mu^2.$$

得 $$\begin{cases}\mu=\bar{X}\\ \sigma^2+\mu^2=\dfrac{n-1}{n}S^2+\bar{X}^2\end{cases}$$

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

$$A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2 \qquad =\frac{(n-1)S^2+n\bar{X}^2}{n}$$

解得 $$\begin{cases}\hat{\mu}=\bar{X}\\ \hat{\sigma}^2=\dfrac{n-1}{n}S^2\end{cases}$$

例 7.3 在某班期末数学考试成绩中随机抽取 9 人的成绩，结果如表 7-1 所示. 试求该班数学成绩的平均分数、标准差的矩估计值.

表 7-1

序号	1	2	3	4	5	6	7	8	9
分数	94	89	85	78	75	71	65	63	55

解析: 假设成绩 X 服从正态分布

$$\mu = E(X), \sigma^2 = D(X)$$

$$\overline{X} = \frac{1}{9}\sum_{i=1}^{9} x_i = \frac{1}{9}(94 + 89 + \cdots + 55) = 75;$$

$$S^2 = \frac{1}{9-1}\left(\sum_{i=1}^{9} X_i^2 - n\bar{X}^2\right) = 165.75$$

由 $\begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \dfrac{n-1}{n} S^2 \end{cases}$ 得矩估计值分别为 $\begin{cases} \hat{\mu} = 75 \\ \hat{\sigma} = \sqrt{\dfrac{8}{9}} S \approx 12.1381 \end{cases}$

3.最大（极大）似然估计 Maximum Likelihood Estimation

The method of maximum likelihood was first used by Sir Ronald Fisher (a geneticist and statistician) in 1922s, for finding estimator of a unknown parameter.

However, the method originated in the works of Gauss and Bernoulli.

Most statisticians recommend this method, at least when the sample size is large, since the resulting estimators have certain desirable(可取的） efficiency properties

定义: $X_1, X_2, \dots, X_n$ 为来自总体 X 的随机样本。X 的概率密度或者概率质量函数为$f(x,\theta)$，那么，$X_1, X_2, \dots, X_n$ 的联合密度或者质量函数（离散随机变量的分布律）为

$$f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta) \triangleq L(\theta)$$

又称为似然函数， likelihood function

如果 $\hat{\theta}_L$ 能使得似然函数$L(\theta)$ 达到最大（或极大）值，则称为θ的极大似然估计值，记为 $\hat{\theta}_L$.

注： $L(\theta)$ 达到最大（或极大）值时，意味着事件 $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ 的概率达到最大

14. 设总体 X 的概率分布如表 7-6 所示，其中 $\theta, 0<\theta<\frac{1}{2}$ 是未知参数，利用总体的如下样本值 3，1，3，0，3，1，2，3，求 θ 的矩估计值和最大似然估计值. 数学一 (2002 研考)

表 7-6

X	0	1	2	3
p	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

解析: $X_1=3, X_2=1, X_3=3, X_4=0, X_5=3,$
$X_6=1, X_7=2, X_8=3$ 的概率为

$$L(\theta)=(1-2\theta)^4[2\theta(1-\theta)]^2(\theta^2)(\theta^2)$$

如何求该函数的最大值点？ 导数为0. 但是 $L(\theta)$ 的导数表达式复杂.

$\ln L(\theta)$ 和 $L(\theta)$ 具有相同的极大值点.

$$\ln L(\theta)=4\ln(1-2\theta)+2\ln\big(2\theta(1-\theta)\big)+4\ln\theta$$
$$=4\ln(1-2\theta)+2\ln 2+2\ln\theta+2\ln(1-\theta)+4\ln\theta$$

令 $\frac{d\ln L(\theta)}{d\theta}=4\frac{-2}{1-2\theta}+2\frac{1}{\theta}+2\frac{-1}{1-\theta}+4\frac{1}{\theta}=0$

化简， $12\theta^2-14\theta+3=0$ 解得， $\theta_1=\frac{7-\sqrt{13}}{12}$ $\theta_2=\frac{7+\sqrt{13}}{12}$(舍掉)

极大似然估计法一般步骤 (一个未知参数 θ)

1. 获得样本值 $x_1, x_2, \ldots, x_n$

2. 构造极大似然估计函数 $L(\theta) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$

3. 计算 $\ln L(\theta)$, 令 $\frac{d}{d\theta}\ln L(\theta) = 0$ 求解 $\hat{\theta}_L$

4. $\hat{\theta}_L$即为 θ 的极大似然估计值.

如果有 k 个未知的参数 $\theta_1, \theta_2, \dots, \theta_k$:

1. 构造极大似然函数

$$L(\theta_1, \theta_2, \dots, \theta_k) = f(x_1; \theta_1, \theta_2, \dots, \theta_k) f(x_2; \theta_1, \theta_2, \dots, \theta_k) \cdots f(x_n; \theta_1, \theta_2, \dots, \theta_k)$$

2.
$$\begin{cases} \dfrac{\partial \ln L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_1} = 0 \\ \dfrac{\partial \ln L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_2} = 0 \\ \cdots \\ \dfrac{\partial \ln L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_k} = 0 \end{cases}$$

例 $X_1, X_2, \ldots, X_n$ 为来自总体 X 的随机样本，X 的概率密度函数为

$$f(x;\theta) = \begin{cases} (1-\theta)x^{-\theta}, 0 < x < 1 \\ 0, \qquad\qquad 其他, \end{cases}$$

求 θ 的极大似然估计量。

解：θ 的极大似然函数是

$$L(\theta) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta) = (1-\theta)x_1^{-\theta}\,(1-\theta)x_2^{-\theta}\cdots(1-\theta)x_n^{-\theta}$$

$$= (1-\theta)^n x_1^{-\theta} x_2^{-\theta} \cdots x_n^{-\theta}$$

$$\ln L(\theta) = n\ln(1-\theta) - \theta\ln x_1 - \theta\ln x_2 - \cdots - \theta\ln x_n$$

令 $\frac{d}{d\theta}\ln L(\theta) = \frac{-n}{1-\theta} - \ln x_1 - \ln x_2 - \cdots - \ln x_n = 0$，得

$$\hat{\theta}_L = 1 + \frac{n}{\sum_{i=1}^n \ln x_i}$$

θ 的极大似然估计量是

$$\hat{\theta}_L = 1 + \frac{n}{\sum_{i=1}^n \ln X_i}$$

用 X_i 代替 x_i

例 7.7 设 $x_1, x_2, \cdots, x_n$ 为来自正态总体 $N(\mu, \sigma^2)$ 的观测值，试求总体未知参数 μ, σ^2 的最大似然估计.

解 因正态总体为连续型，其密度函数为 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$， 故似然函数为

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\},$$

$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$. 故对数似然方程组为

$$\begin{cases} \dfrac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu) = 0, \\ \dfrac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(x_i-\mu)^2 = 0. \end{cases}$$

解得 $\hat{\mu}_L = \bar{x}$, $\hat{\sigma}_L^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

所以 $\hat{\mu}_L = \bar{X}$

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

例: $X_1, X_2, \dots, X_n$ 为来自总体X服从泊松分布的随机样本， 参数为 λ. 给定一组样本观测值 $x_1, x_2, \dots, x_n$， 求λ 的极大似然估计值。

解: 泊松分布的分布律为 $$P(X = x_i) = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

故似然函数为 $$L(\lambda) = \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} e^{-\lambda} \cdots \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} = \frac{\lambda^{x_1 + x_2 + \cdots + x_n}}{x_1! \, x_2! \cdots x_n!} e^{-n\lambda}$$

$$\ln L(\lambda) = (x_1 + x_2 + \cdots + x_n) \ln \lambda - n\lambda - \ln(x_1! \, x_2! \cdots x_n!)$$

令 $\frac{d}{d\lambda} \ln L(\lambda) = (x_1 + x_2 + \cdots + x_n) \frac{1}{\lambda} - n = 0$ 得

λ 的极大似然估计值

$$\hat{\lambda}_L = \frac{1}{n} \sum_{i=1}^{n} x_i$$

λ 的极大似然估计量

$$\hat{\lambda}_L = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$$

例: 设总体 X 服从 $[0, \ \theta]$ 上的均匀分布，$X_1, X_2, \dots, X_n$ 为随机样本， 样本值分别为 $x_1, x_2, \dots, x_n$， 求 θ 的极大似然估计值和矩估计值

解:

（1） X的概率密度函数为 $f(x;\theta) = \begin{cases} \frac{1}{\theta}, 0 < x < \theta \\ 0, \quad \text{其他}, \end{cases}$

故似然函数为 $L(\theta) = \frac{1}{\theta} \ \frac{1}{\theta} \ \cdots \ \frac{1}{\theta} = \frac{1}{\theta^n}$,

在给定样本值 $x_1, x_2, \dots, x_n$ 条件下 $L(\theta)$ =何时达到最大值?

注意到 $x_1 \le \theta, x_2 \le \theta, \dots, x_n \le \theta$

显然 $\hat{\theta}_L = \max\limits_{1 \le i \le n}\{x_i\}$ 极大似然估计量为 $\hat{\theta}_L = \max\limits_{1 \le i \le n}\{X_i\}$

（2）矩估计 $E(X) = A_1.$ 即 $\frac{\theta}{2} = \bar{X}$ 得矩估计量 $\hat{\theta} = 2\bar{X}$