第七章　参数估计

问题提出：为什么要进行参数估计？在实际问题中，总体分布一般是未知的，我们常常事先假定总体分布的类型，再通过取样的方式估计分布中的未知参数.

例 **1.** 考察某随机试验中事件 A 发生的概率. 在一次试验中事件 A 发生的次数服从参数为 p 的两点分布. 这里 p 需要通过样本进行估计.

例 **2.** 一般假定一个城市在单位时间内发生交通事故的次数服从泊松分布 $\pi(\lambda)$，这里的参数 λ 需要通过观察取样来给出.

例 **3.** 假定某种电子元件的寿命服从参数为 λ 的指数分布，这里 λ 为待定参数.

例 **4.** 假定某城市居民的收入服从正态分布 $N(\mu,\sigma^2)$，这里 μ 和 σ^2 的具体值都需要取样估计.

例 **5.** 假定中国 *20-30* 岁人群的身高、体重服从二维正态分布

$$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho),$$

分布中 5 个参数都需要通过抽取样本然后用统计量近似估计.

7.1　点估计

定义 **1.** 设 θ 为总体 X 的待估计参数, 用样本 $X_1,X_2,\cdots,X_n$ 的个统计量 $\widehat{\theta}=\widehat{\theta}(X_1,X_2,\cdots,X_n)$ 来估计 θ,则称 $\widehat{\theta}(X_1,X_2,\cdots,X_n)$ 为的点估计量.对应于样本观测值 $x_1,x_2,\cdots,x_n$ 称 $\widehat{\theta}(x_1,x_2,\cdots,x_n)$ 为的点估计值.

7.1.1　矩估计法

矩估计是基于“替换”的思想给出的一种参数估计方法. 最早由英国统计学家皮尔逊 (K. Pearson) 提出. 基本思想：用同类、同阶的样本矩来估计总体矩.

对总体 X，其 m 阶原点矩为

$$\mu_m := E(X^m).$$

对样本 $X_1, X_2, \cdots, X_n$，其 m 阶样本原点矩为

$$A_m := \frac{1}{n}\sum_{i=1}^{n} X_i^m = \frac{X_1^m + X_2^m + \cdots + X_n^m}{n}.$$

矩估计的理论基础：因为 $X_1, X_2, \cdots, X_n$ 是独立同分布的，故 X_1^m, X_2^m, $\cdots$, X_n^m 也是独立同分布的，且

$$E(X_1^m) = E(X_2^m) = \cdots = E(X_n^m) = \mu_m.$$

由大数定律知，当 n 趋于无穷时，

$$A_m = \frac{1}{n}\sum_{i=1}^{n} X_i^m \xrightarrow{P} \mu_m.$$

矩估计：设总体 X 的分布函数为 $F(x;\ \theta_1, \theta_2, \cdots, \theta_k)$, 其中 θ_1, θ_2, $\cdots$, θ_k 为待定参数. 若 X 的 $1 \sim k$ 阶原点矩都存在，则其为 $\theta_1, \theta_2, \cdots, \theta_k$ 的函数，记为

$$\mu_m = E(X^m) = \mu_m(\theta_1, \theta_2, \cdots, \theta_k),\ m = 1, 2, \ldots, k$$

矩估计（续）：令总体的前 k 阶原点矩分别与同阶的样本原点矩相等，这样就得到了一个 k 元方程组

$$\begin{cases} \mu_1(\theta_1, \theta_2, \cdots, \theta_k) = A_1 \\ \mu_2(\theta_1, \theta_2, \cdots, \theta_k) = A_2 \\ \qquad\cdots\cdots\cdots \\ \mu_k(\theta_1, \theta_2, \cdots, \theta_k) = A_k \end{cases}$$

记方程组的解为 $\widehat{\theta}_1$, $\widehat{\theta}_2$, $\cdots$, $\widehat{\theta}_k$，用它们作为参数 θ_1, θ_2, $\cdots$, θ_k 的估计量. 这样的估计称为矩估计.

例子. 当总体中只有一个参数时，矩估计即是用样本均值估计总体期望.

例 **1.** 在对事件 A 进行 n 次独立重复观测中，得到记录 $X_1, X_2, \cdots, X_n$，其中 X_i 表示第 i 次试验中事件 A 发生的次数. 试估计事件 A 发生的概率 p.

解. $p = \overline{X} = \dfrac{X_1 + X_2 \ldots + X_n}{n}$.

例子. 当总体中有两个或以上的参数时，总体期望与方差的矩估计分别为

$$\widehat{\mu} = \overline{X}, \qquad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

例 **2.** 对正态总体 $N(\mu, \sigma^2)$，两个参数的矩估计即为上式.

例 **3.** 设总体 X 服从 $[\theta_1, \theta_2]\,(\theta_1 < \theta_2)$ 上的均匀分布, 即概率密度函数为

$$f(x, \theta_1, \theta_2) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1}, & \theta_1 \leqslant x \leqslant \theta_2, \\ 0, & \text{其他}, \end{cases}$$

其中 θ_1, θ_2 未知 ,$X_1, X_2, \cdots, X_n$ 是一个样本, 求 θ_1, θ_2 的矩估计量.

解. 由条件易得

$$E(X) = \frac{1}{2}(\theta_1 + \theta_2),$$

$$E\left(X^2\right) = D(X) + [E(X)]^2 = \frac{(\theta_2 - \theta_1)^2}{12} + \frac{(\theta_2 + \theta_1)^2}{4}.$$

令

$$\frac{1}{2}(\theta_2 + \theta_1) = A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\frac{(\theta_2 - \theta_1)^2}{12} + \frac{(\theta_2 + \theta_1)^2}{4} = A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

解之, 得 θ_1, θ_2 的矩估计量为

$$\widehat{\theta}_1 = A_1 - \sqrt{3\left(A_2 - A_{\mathrm{L}}^2\right)} = \overline{X} - \sqrt{\frac{3(n-1)}{n}}S,$$

$$\widehat{\theta}_2 = A_1 + \sqrt{3\left(A_2 - A_1^2\right)} = \overline{X} + \sqrt{\frac{3(n-1)}{n}}S.$$

例 **4.** 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本, 并且总体 X 的均值 μ 及方差 σ^2 都存在, 且 $\sigma^2 > 0$. 试求总体的期望 $\mu = E(X)$ 和方差 $\sigma^2 = D(X)$ 的矩估计.

解**.** 由条件易得

$$D(X) = E\left(X^2\right) - [E(X)]^2$$

$$E\left(X^2\right) = \sigma^2 + \mu^2.$$

令

$$\begin{cases} \mu = A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \\ \sigma^2 + \mu^2 = A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2. \end{cases}$$

解上述方程组, 得 μ 和 σ^2 的矩估计量分别为

$$\widehat{\mu} = \overline{X},$$

$$\widehat{\sigma}^2 = A_2 - A_1^2 = \frac{n-1}{n} S^2.$$

例 **5.** 设总体 X 的分布律为

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

其中 $\theta \in (0,1)$. 利用 X 的一组样本值 $(x_1, x_2, x_3) = (-1, 0, -1)$, 求 θ 的矩估计. $\cdot\ \widehat{\theta} = \frac{5}{6}$

即我们估计出来的总体分布为

X	-1	0	1
P	25/36	10/36	1/36

练习 **1.** 设总体 X 的分布律为

X	1	2	3
P	$\frac{\theta}{2}$	$\frac{\theta}{2}$	$1-\theta$

其中 $\theta \in (0,1)$. 现抽取 *4* 个样本观察得

$$x_1 = 2,\ x_2 = 1,\ x_3 = 2,\ x_4 = 3.$$

求 θ 的矩估计.

矩估计的优点

1. 思路直观，简单易行
2. 不需要事先知道总体是什么分布

矩估计的缺点

1. 矩的选取比较随意，在选取不同的组合时，得到的估计量也不同
2. 当总体的分布类型已知时，未充分利用分布所提供的信息.

7.1.2 极大似然估计法

极大似然估计法是在总体的分布类型已知的前提下使用的一种参数估计法. 该方法很早已被高斯、拉普拉斯等数学家使用，其后英国统计学家费歇 (R.A.Fisher) 在 1912–1922 年正式提出该方法，并做分析研究及推广应用.

思路：样本 $X_1,X_2,\cdots,X_n$ 经具体抽样所得的观察值 $x_1,x_2,\cdots,x_n$ 应当是最可能出现的数值.
方法：在已知分布类型的情况下，求参数使得事件

$$\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$$

对应的概率（离散型）或者概率密度（连续型）达到最大值.

定义 **2.** 设 X 为离散型随机变量，称函数

$$f(x)=P\{X=x\},\quad x\in\mathbb{R}$$

为 X 的概率函数.

若离散型随机变量 X 的分布律为

$$P\{X=x_i\}=p_i,\quad i=1,2,\cdots$$

则其概率函数为

$$f(x)=\begin{cases}p_i, & x=x_i,\ \text{对某个 } i,\\ 0, & \text{其他}.\end{cases}$$

设离散型总体 X 的概率函数为 $f(x)$，则事件

$$\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$$

对应的概率为

$$\prod_{i=1}^{n} f(x_i)=f(x_1)f(x_2)\cdots f(x_n).$$

设连续型总体 X 的概率密度函数为 $f(x)$，则事件

$$\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$$

对应的联合概率密度为

$$\prod_{i=1}^{n} f(x_i)=f(x_1)f(x_2)\cdots f(x_n).$$

定义. 设连续型（离散型）总体 X 的概率密度函数（概率函数）为

$$f(x;\ \theta_1,\theta_2,\cdots,\theta_k).$$

则样本 $X_1,X_2,\cdots,X_n$ 在观测值 $x_1,x_2,\cdots,x_n$ 处的联合概率密度（概率）为

$$L(x_1,\cdots,x_n;\ \theta_1,\cdots,\theta_k)=\prod_{i=1}^{n} f(x_i;\ \theta_1,\cdots,\theta_k).$$

将观测值 $x_1,\cdots,x_n$ 看成固定的，将 L 看做 $\theta_1,\cdots,\theta_k$ 的函数，则该函数被称为似然函数.

参数的选择：哪一组参数值使得现有的样本观测值出现的可能性最大，哪一组参数就与观察结果最匹配的参数.

定义. 如果似然函数

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

在

$$(\theta_1^*,\theta_2^*,\cdots,\theta_k^*)$$

处达到最大值，则称上述参数为未知参数的极大似然估计.

求似然函数最大值点的方法：

1. 若函数可微，则其在最大值点处的导数（偏导数）为零；
2. 由于 $\ln L$ 与 L 同时达到最大值，且在实际应用中常常比 L 更方便求导，故一般将问题化为求 $\ln L$ 的最大值点.

求极大似然估计的一般方法：

1. 写出似然函数 L；
2. 求似然函数的对数 $\ln L$；
3. 对 $\ln L$ 求导（偏导）并令导数等于零，得到似然方程组；
4. 解方程组得到 $\ln L$ 的驻点，判断该驻点是否最大值点.

例 **6.** 设某电子元件的寿命 T 服从参数为 λ 的指数分布. 测得 n 个元件的失效时间为 $x_1, x_2, \cdots, x_n$, 试求 λ 的极大似然估计值.

解**.** 似然函数为

$$L(\lambda) = \prod_{i=1}^{n} \lambda \mathrm{e}^{-\lambda x_i} = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right),$$

$$\ln[L(\lambda)] = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i.$$

似然方程为

$$\frac{\mathrm{d} \ln[L(\lambda)]}{\mathrm{d}\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

解得

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}},$$

其中 $\overline{x}$ 是样本均值. 因为当 $\widehat{\lambda} < \dfrac{1}{\overline{x}}$ 时,$\dfrac{\mathrm{d} \ln[L(\lambda)]}{\mathrm{d}\lambda} > 0$; 当 $\widehat{\lambda} > \dfrac{1}{\overline{x}}$ 时,$\dfrac{\mathrm{d} \ln[L(\lambda)]}{\mathrm{d}\lambda} < 0$; 故 $\widehat{\lambda} = \frac{1}{\overline{x}}$ 确实是 $L(\lambda)$ 的最大值点. 因此,λ 的极大似然估计值为 $(\overline{x})^{-1}$.

例 **7.** 设总体 X 服从参数为 λ 的泊松分布, $x_1, x_2, \cdots, x_n$ 为一样本取值, 求参数 λ 的极大似然估计.

解. 似然函数为

$$L(\lambda)=\prod_{i=1}^{n} P\left\{X_i=x_i\right\}=\prod_{i=1}^{n}\left(\frac{\lambda^{x_i}}{x_i!}\mathrm{e}^{-1}\right)=\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}\left(x_i!\right)}\mathrm{e}^{-n\lambda},$$

$$\ln[L(\lambda)]=\left(\sum_{i=1}^{n} x_i\right)\ln\lambda-n\lambda-\sum_{i=1}^{n}\ln\left(x_i!\right)$$

故似然方程为

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\{\ln[L(\lambda)]\}=\frac{1}{\lambda}\sum_{i=1}^{n} x_i-n=0.$$

解得 λ 的极大似然估计值为

$$\widehat{\lambda}=\frac{1}{n}\sum_{i=1}^{n} x_i=\overline{x}.$$

极大似然估计量为

$$\widehat{\lambda}=\overline{X}.$$

结论：样本均值是泊松分布的参数的极大似然估计.

例 **8.** 设 $x_1,x_2,\cdots,x_n$ 是来自总体 $N\left(\mu,\sigma^2\right)$ 的样本取值, 其中 μ,σ^2 是未知参数, 试求 μ 和 σ^2 的极大似然估计 $\left(|\mu|<+\infty,\sigma^2>0\right)$.

解. 似然函数为

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

$$\ln\left[L\left(\mu,\sigma^2\right)\right]=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2.$$

似然方程组为

$$\left\{\begin{array}{l}\frac{\partial}{\partial\mu}\left\{\ln\left[L\left(\mu,\sigma^2\right)\right]\right\}=\frac{1}{\sigma^2}\sum\limits_{i=1}^{n}\left(x_i-\mu\right)=0,\\ \frac{\partial}{\partial\sigma^2}\left\{\ln\left[L\left(\mu,\sigma^2\right)\right]\right\}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum\limits_{i=1}^{n}\left(x_i-\mu\right)^2=0.\end{array}\right.$$

解方程组, 得 μ 和 σ^2 的极大似然估计值为

$$\left\{\begin{array}{l}\widehat{\mu}=\frac{1}{n}\sum_{i=1}^{n} x_i=\overline{x},\\ \widehat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2=\frac{n-1}{n}s^2.\end{array}\right.$$

μ 和 σ^2 的极大似然估计量为

$$\widehat{\mu}=\overline{X},\quad \widehat{\sigma}^2=\frac{n-1}{n}S^2.$$

例 **9.** 设总体 X 的分布律为

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

其中 $\theta\in(0,1)$. 利用 X 的一组样本值 $(x_1,x_2,x_3)=(-1,0,-1)$, 求 θ 的极大似然估计.

$\theta^*=\frac{5}{6}$

练习 **2.** 设总体 X 的分布律为

X	1	2	3
P	$\frac{\theta}{2}$	$\frac{\theta}{2}$	$1-\theta$

其中 $\theta\in(0,1)$. 现抽取 *4* 个样本观察得

$$x_1=2,\ x_2=1,\ x_3=2,\ x_4=3.$$

求 θ 的极大似然估计.

如果似然函数不可导，则只能用定义求极大似然估计.

例 **10.** 设总体 X 服从均匀分布 $U[0,\theta],(\theta>0)$, 即密度函数为

$$f(x)=\begin{cases}\frac{1}{\theta}, & x\in[0,\theta],\\ 0, & \text{其他}.\end{cases}$$

$X_1,X_2,\cdots,X_n$ 为一样本, 求 θ 的极大似然估计.

解**.** 似然函数为

$$L(\theta)=\begin{cases}\frac{1}{\theta^n}, & x_1,\cdots,x_n\in[0,\theta],\\ 0, & \text{其他}.\end{cases}$$

显然, $L(\theta)$ 的最大值应该在 $x_1,x_2,\cdots,x_n\in[0,\theta]$ 时取得. 而 $x_1,x_2,\cdots,x_n\in[0,\theta]$ 等价于

$$0\leqslant\min\{x_1,x_2,\cdots,x_n\}\leqslant\max\{x_1,x_2,\cdots,x_n\}\leqslant\theta.$$

另一方面, $L(\theta)=\theta^{-n}$ 是关于 θ 的单週递減函数, 而 θ 的最小值为 $\max\{x_1,x_2,\cdots,x_n\}$, 故 $L(\theta)$ 在 $\theta=\max\{x_1,x_2,\cdots,x_n\}$ 取得最大值，即所求的极大似然估计值为

$$\widehat{\theta}=\max\{x_1,x_2,\cdots,x_n\}.$$

极大似然估计量为

$$\widehat{\theta}=\max\{X_1,X_2,\cdots,X_n\}.$$

7.2 估计量的评选标准

假设 θ 为总体分布的参数，设

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

是 θ 的一个估计．如果对 θ 的一切可能取值，都有

$$E(\widehat{\theta})=\theta,$$

则称 $\widehat{\theta}$ 是 θ 的无偏估计(unbiased estimator). 称 $E(\widehat{\theta})-\theta$ 为估计量 $\widehat{\theta}(X_1,X_2,\cdots,X_n)$ 的系统误差.

例 **1.** 设 $X_1,X_2,\cdots,X_n$ 为抽自均值为 μ 的总体的样本，以下三个估计量都是 μ 的无偏估计：

$$\widehat{\mu}_1=X_1,\qquad \widehat{\mu}_2=\frac{X_1+X_2}{2},$$

$$\widehat{\mu}_3=\frac{X_1+X_2+X_{n-1}+X_n}{4}\ \ (\text{当 } n\geq 4 \text{ 时}).$$

而以下两个估计量都不是 μ 的无偏估计：

$$\widehat{\mu}_4=2X_1,\qquad \widehat{\mu}_5=\frac{X_1+X_2}{3}.$$

例 **2.** 设总体 X 服从均匀分布 $U[0,\theta](\theta>0)$, 即概率密度为

$$f_X(x)=\begin{cases}\dfrac{1}{\theta}, & x\in[0,\theta],\\ 0, & \text{其他}.\end{cases}$$

$X_1,\cdots,X_n$ 为一样本, 试问 $T_1=2\overline{X}$ 和 $T_2=X_{(n)}=\max\{X_1,\cdots,X_n\}$ 是否为 θ 的无偏估计?

解. 由 $E(X)=\dfrac{\theta}{2}$ 知

$$E(T_1)=E\left(\frac{2}{n}\sum_{i=1}^{n}X_i\right)=\frac{2}{n}\sum_{i=1}^{n}E(X_i)=\theta,$$

可见 T_1 是 θ 的无偏估计.

为计算 $E(T_2)$, 先求 T_2 的分布

$$\begin{aligned}F_{X(n)}(x)&=P\left\{X_{(n)}\leqslant x\right\}=P\left\{\max\left\{X_1,\cdots,X_n\right\}\leqslant x\right\}\\&=P\left\{X_1\leqslant x,X_2\leqslant x,\cdots,X_n\leqslant x\right\}\\&=P\left\{X_1\leqslant x\right\}\cdot\cdots\cdot P\left\{X_n\leqslant x\right\}\\&=\left[F_X(x)\right]^n=\begin{cases}\frac{x^n}{\theta^n}, & x\in[0,\theta],\\0, & \text{其他},\end{cases}\end{aligned}$$

故 T_2 的概率密度为

$$f(x,\theta)=\begin{cases}\frac{nx^{n-1}}{\theta^n}, & 0\leqslant x\leqslant\theta,\\0, & \text{其他},\end{cases}$$

从而

$$E(T_2)=\int_0^\theta xf(x,\theta)\mathrm{d}x=\int_0^\theta\frac{nx^n}{\theta^n}\,\mathrm{d}x=\frac{n}{n+1}\theta.$$

可见 T_2 不是 θ 的无偏估计.

定理 **1.** 设总体 X 的期望和方差分别为

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

从总体取一组样本 $X_1,X_2,\cdots,X_n$，则样本均值 $\overline{X}$ 与样本方差 S^2 分别是 μ 和 σ^2 的无偏估计，即

$$E(\overline{X})=\mu,\qquad E(S^2)=\sigma^2.$$

作为对比，估计量

$$\widehat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

不是总体方差的无偏估计.

说明：如果 $\widehat{\theta}$ 是参数 θ 的一个估计，我们通常用 $g(\widehat{\theta})$ 去估计 $g(\theta)$. 但是，即使 $\widehat{\theta}$ 是 θ 的无偏估计，$g(\widehat{\theta})$ 也不一定是 $g(\theta)$ 的无偏估计.

例 **3.** 样本标准差 S 不是总体标准差 σ 的无偏估计.

定义 **1.** 设 $\widehat{\theta}$ 是参数 θ 的估计量，称

$$MSE(\widehat{\theta}) := E[(\widehat{\theta}-\theta)^2]$$

为 $\widehat{\theta}$ 的均方误差*(mean squared error)*.

均方误差满足以下等式

$$MSE(\widehat{\theta}) = D(\widehat{\theta}) + [E(\widehat{\theta})-\theta]^2.$$

当 $\widehat{\theta}$ 为 θ 的无偏估计量时，有

$$MSE(\widehat{\theta}) = D(\widehat{\theta}).$$

估计量的比较：在估计量的选取中，

1. 无偏估计量优于有偏估计量；
2. 在无偏估计量中，方差越小的越好（有效性准则）.

设 $\widehat{\theta}$ 是参数 θ 的一个估计量，若对于任意 $\theta \in \Theta$，当 $n \to \infty$ 时，$\widehat{\theta}$ 依概率收敛于 θ，则称 $\widehat{\theta}$ 是 θ 的相合估计量. 即若对任意的 $\theta \in \Theta$ 都满足：对任意的 $\epsilon > 0$，有 $\lim\limits_{n\to\infty} P\{|\widehat{\theta}-\theta| < \epsilon\} = 1$ 则称 $\widehat{\theta}$ 是 θ 的相合估计量.

估计三准则：无偏性、有效性、相合性.