求偏导、并令偏导等于零有

$$\frac{\partial \ln L(\theta,\mu)}{\partial \theta}=\frac{n}{\theta}-\sum_{i=1}^{n}(X_i-\mu)=0 \quad \Rightarrow \quad \theta=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)},$$

另一方面有

$$\frac{\partial \ln L(\theta,\mu)}{\partial \mu}=n\theta=0 \quad \Rightarrow \quad \theta=0,$$

此时无法求解 θ 和 μ 的最大似然估计. 回到似然函数的定义

$$L(\theta,\mu)=\begin{cases}\theta^n e^{-\theta\sum_{i=1}^{n}(X_i-\mu)} & X_1,X_2,\cdots,X_n\geqslant\mu\\ 0 & \text{其它}\end{cases}$$

可以发现 μ 越大似然函数 $L(\theta,\mu)$ 越大, 但须满足 $X_i\geqslant\mu$ $(i\in[n])$. 由此可得最大似然估计

$$\hat{\mu}=\min\{X_1,X_2,\cdots,X_n\},$$

进一步求解可得

$$\hat{\theta}=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\mu})}.$$

9.2 估计量的评价标准

前一节已经讲过不同的点估计方法, 不同的估计方法可能得到不同的估计值, 自然涉及到一个问题: 采用哪一种估计量更好, 或更好的标准是什么呢? 估计量的常用标准: 无偏性, 有效性, 一致性.

9.2.1 无偏性

定义 9.1 设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本, 令 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是 θ 的一个估计量, 若

$$E_{X_1,X_2,\ldots,X_n}\left[\hat{\theta}\right]=E_{X_1,X_2,\ldots,X_n}\left[\hat{\theta}(X_1,X_2,\cdots,X_n)\right]=\theta$$

则称 $\hat{\theta}$ 为 θ 的无偏估计.

无偏估计不要求估计值 $\hat{\theta}$ 在任意情况下都等于 θ, 但在期望的情形下有 $E(\hat{\theta})=\theta$ 成立. 其意义在于无系统性偏差, 无偏性是一种对估计量常见而且重要的标准.

首先看看如下例子:

例 9.8 (样本 k 阶原点矩为总体 k 阶原点矩的无偏估计) 设 $X_1,X_2,\cdots,X_n$ 是总体 X 的样本, 若 $E[X^k]$ 存在, 则 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 是总体 $a_k=E[X^k]$ 的无偏估计.

例 9.9 设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本, 其期望为 μ, 方差为 σ^2, 则: 1) $S_0^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/n$ 是 σ^2 的有偏估计; 2) $S^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)$ 是 σ^2 的无偏估计.

注意 $\hat{\theta}=\hat{\theta}(X_1,X_2,\ldots,X_n)$ 是 θ 的无偏估计, 但并不一定有 $g(\hat{\theta})$ 是 $g(\theta)$ 的无偏估计, 这是因为 $E[\hat{\theta}]=\theta$ 并不能推导出 $E[g(\hat{\theta})]=g(\theta)$. 例如

$$E[\bar{X}]=E[X]=\mu \qquad 但 \qquad E[(\bar{X})^2]\neq\mu^2.$$

例 9.10 设 $X_1,X_2,\cdots,X_n$ 是总体 X 的样本, 以及总体 X 的概率密度为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}} & x\geqslant 0\\ 0 & x<0,\end{cases}$$

证明: $\bar{X}=\sum_{i=1}^n X_i/n$ 和 $n\min\{X_1,X_2,\cdots,X_n\}$ 均是 θ 的无偏估计.

证明 根据期望和指数分布的性质有

$$E[\bar{X}]=E[X]=\theta,$$

由此可知 $\bar{X}$ 是 $E[X]$ 的无偏估计. 设随机变量 $Z=\min\{X_1,X_2,\cdots,X_n\}$, 则有

$$\begin{aligned}F_Z(z) &= \Pr[Z\leqslant z]=1-\Pr[Z>z]\\ &= 1-\Pr[X_1>z]\Pr[X_2>z]\cdots\Pr[X_n>z]\\ &= 1-\prod_{i=1}^n(1-\Pr[X_i\leqslant z])=\begin{cases}0 & z<0\\ 1-e^{-nz/\theta} & z\geqslant 0.\end{cases}\end{aligned}$$

于是当 $z\geqslant 0$ 时有

$$\Pr[Z>z]=1-F_Z(z)=e^{-nz/\theta}.$$

根据期望的性质有

$$E[Z]=\int_0^{+\infty}\Pr[Z>z]dz=\int_0^{+\infty}e^{-nz/\theta}dz=\frac{\theta}{n}.$$

于是有 $\theta=E[nZ]$ 成立.

9.2.1.1 有效性

参数可能存在多个无偏估计, 若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是 θ 的无偏估计, 则可以比较方差

$$\mathrm{Var}(\hat{\theta}_1)=E[(\hat{\theta}_1-\theta)^2] \qquad 和 \qquad \mathrm{Var}(\hat{\theta}_2)=E[(\hat{\theta}_2-\theta)^2].$$

一般而言: 方差越小, 无偏估计越好.

定义 9.2 设 $\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$ 和 $\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$ 是 θ 的两个无偏估计, 若

$$\mathrm{Var}(\hat{\theta}_1)\leqslant\mathrm{Var}(\hat{\theta}_2),$$

则称 θ_1 比 θ_2 有效.

例 9.11　设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, 且 X 的概率密度为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}} & x \geqslant 0\\ 0 & x<0\end{cases},$$

令 $Z=\min\{X_1, X_2, \cdots, X_n\}$, 证明: 当 $n>1$ 时 $\bar{X}=\sum_{i=1}^n X_i/n$ 比 nZ 有效.

证明　根据独立性有

$$\mathrm{Var}(\bar{X})=\frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i)=\frac{\theta^2}{n}.$$

根据例 9.10 可知随机变量 Z 的概率密度为

$$f(z)=\begin{cases}0 & z<0\\ \frac{n}{\theta}e^{-\frac{nz}{\theta}} & z \geqslant 0\end{cases}$$

从而得到

$$\mathrm{Var}(nZ)=n^2\mathrm{Var}(Z)=n^2\frac{\theta^2}{n^2}=\theta^2,$$

因此当 $n \geqslant 1$ 时有 $\mathrm{Var}(\bar{X}) \leqslant \mathrm{Var}(nZ)$ 成立, 故 $\bar{X}$ 比 nZ 有效.

例 9.12　设 $X_1, X_2, \cdots, X_n$ 是总体 X 的样本, 且 $E(X)=\mu$ 和 $\mathrm{Var}(X)=\sigma^2$. 设常数 $c_1, c_2, \cdots, c_n \geqslant 0$ 满足 $\sum_{i=1}^n c_i=1, c_i \neq 1/n$, 求证: $\bar{X}$ 比 $\sum_{i=1}^n c_iX_i$ 有效.

证明　根据样本的独立同分布条件有

$$E\left[\bar{X}\right]=\mu \quad 和 \quad \mathrm{Var}\left(\bar{X}\right)=\sigma^2/n.$$

根据期望的性质有 $E[\sum_{i=1}^n c_iX_i]=\mu$, 进一步有

$$\mathrm{Var}\left(\sum_{i=1}^n c_iX_i\right)=\sum_{i=1}^n c_i^2\mathrm{Var}(X_i)=\sigma^2\sum_{i=1}^n c_i^2 \geqslant \frac{\sigma^2}{n}$$

这里利用不等式 $\sum_{i=1}^n c_i^2/n \geqslant (\sum_{i=1}^n c_i/n)^2=1/n^2$, 所以有 $\mathrm{Var}(\sum_{i=1}^n c_iX_i) \geqslant \mathrm{Var}(\bar{X})$.

下面定义有效统计量:

定理 9.1 (Rao-Crammer 不等式)　设随机变量 X 的概率密度为 $f(x;\theta)$ 或分布函数为 $F(x;\theta)$, 令

$$\mathrm{Var}_0(\theta)=\frac{1}{nE\left[\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right]} \qquad 或 \qquad \mathrm{Var}_0(\theta)=\frac{1}{nE\left[\left(\frac{\partial \ln F(X;\theta)}{\partial\theta}\right)^2\right]},$$

对任意的无偏估计量 $\hat{\theta}$ 有

$$\mathrm{Var}(\hat{\theta}) \geqslant \mathrm{Var}_0(\theta),$$

称 $\mathrm{Var}_0(\theta)$ 为估计量 $\hat{\theta}$ 方差的下界. 当 $\mathrm{Var}(\hat{\theta}) = \mathrm{Var}_0(\theta)$ 时称 $\hat{\theta}$ 为达到方差下界的无偏估计量, 此时 $\hat{\theta}$ 为最有效估计量, 简称有效估计量.

例 9.13 设 $X_1, X_2, \cdots, X_n$ 为总体 X 的样本, 令总体 X 的密度函数为

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0 \\ 0 & x \leqslant 0, \end{cases}$$

证明: θ 的最大似然估计为有效估计量.

解 首先计算对数似然函数

$$\ln L(\theta) = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^{n} X_i \quad \Rightarrow \quad \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

进一步得到统计量的方差

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{\theta^2}{n}.$$

同时考察

$$\ln f(X;\theta) = -\ln \theta - \frac{X}{\theta}, \quad \frac{\partial \ln f(X;\theta)}{\partial \theta} = -\frac{1}{\theta} + \frac{X}{\theta^2}$$

所以

$$E\left[\frac{\partial \ln f(X;\theta)}{\partial \theta}\right]^2 = E\left[\left(-\frac{1}{\theta} + \frac{X}{\theta^2}\right)^2\right] = \frac{1}{\theta^4} E[(X - E[X])^2] = \frac{1}{\theta^2},$$

从而得到 $\mathrm{Var}_0(X) = \theta^2/n = \mathrm{Var}(\hat{\theta})$, 因此 θ 的最大似然估计是有效估计量.

9.2.1.2 一致性

定义 9.3 设 $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 是 θ 的一个估计量, 若当 $n \to \infty$ 时有 $\hat{\theta}_n \xrightarrow{P} \theta$ 成立, 即对任意 $\epsilon > 0$ 有

$$\lim_{n \to 0} \Pr[|\hat{\theta}_n - \theta| > \epsilon] = 0,$$

则称 $\hat{\theta}_n$ 为 θ 的一致估计量.

估计量的一致性刻画了在足够多样本情形下估计量 $\hat{\theta}$ 能有效逼近真实值 θ, 一致性是对估计的基本要求, 不满足一致性的估计量一般不予考虑. 下面给出满足一致性的充分条件:

定理 9.2 设 $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 是 θ 的一个估计量, 若满足以下两个条件:

$$\lim_{n \to \infty} E\left[\hat{\theta}_n\right] = \theta \quad 和 \quad \lim_{n \to \infty} \mathrm{Var}\left(\hat{\theta}_n\right) = 0,$$

则 $\hat{\theta}_n$ 为 θ 的一致估计量.

证明 根据 $\lim_{n\to\infty} E[\hat{\theta}_n] = \theta$ 知道对任意 $\epsilon > 0$, 存在一个N_0, 当 $n \geqslant N_0$ 有 $|E[\hat{\theta}_n] - \theta| \leqslant \theta/2$, 于是有

$$\lim_{n\to\infty} \Pr\left[|E[\hat{\theta}_n] - \theta| > \epsilon/2\right] = 0.$$

根据 Chebyshev 不等式有

$$\lim_{n\to 0} \Pr\left[\left|\hat{\theta}_n - E[\hat{\theta}_n]\right| > \epsilon/2\right] \leqslant \lim_{n\to 0} \frac{4}{\epsilon}\mathrm{Var}(\hat{\theta}_n) = 0$$

再根据

$$\Pr\left[|\hat{\theta}_n - \theta| > \epsilon\right] \leqslant \Pr\left[\left|\hat{\theta}_n - E[\hat{\theta}_n]\right| > \epsilon/2\right] + \Pr\left[\left|E[\hat{\theta}_n] - \theta\right| > \epsilon/2\right]$$

完成证明.

定理 9.3 设 $\hat{\theta}_{n_1}, \hat{\theta}_{n_2}, \cdots, \hat{\theta}_{n_k}$ 分别为 $\theta_1, \theta_2, \cdots, \theta_k$ 满足一致性的估计量, 对连续函数 $g\colon \mathbb{R}^n \to \mathbb{R}$, 有函数 $\hat{\eta}_n = g(\hat{\theta}_{n_1}, \hat{\theta}_{n_2}, \cdots, \hat{\theta}_{n_k})$ 是 $\eta = g(\theta_1, \theta_2, \cdots, \theta_k)$ 满足一致性的估计量.

根据大数定理可知样本的 k 阶矩是总体 k 阶矩的一致估计量. 矩估计法得到的估计量一般是一致估计量. 最大似然估计量在一定条件下是一致性估计量.

例 9.14 设$X_1, X_2, \cdots, X_n$是来自总体X的样本, 以及总体 X 的密度函数为

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}} & x > 0 \\ 0 & x < 0 \end{cases},$$

则样本均值 $X_n = \sum_{i=1}^n X_i/n$ 为 θ 的无偏、有效、一致估计量.

由前面的例子可知估计的无偏性和有效性, 一致性可根据 $E[X_n] = \theta$ 以及

$$\lim_{n\to\infty} Var(\bar{X}) = \lim_{n\to\infty} \frac{\theta^2}{n} = 0.$$

例 9.15 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim U(0,\theta)$ 的样本, 证明: θ 的最大似然估计量是一致估计量.

证明 根据前面的例题可知 θ 的最大似然估计为 $\hat{\theta}_n = \max(X_1, X_2, \cdots, X_n)$. 设随机变量 $Z = \max(X_1, X_2, \cdots, X_n)$, 则由 Z 的分布函数

$$F_Z(z) = \Pr[Z \leqslant z] = \Pr[\max(X_1, X_2, \cdots, X_n) \leqslant z] = \prod_{i=1}^n \Pr[X_i \leqslant z] = \begin{cases} 1 & z > \theta \\ (\frac{z}{\theta})^n & z \in [0,\theta] \\ 0 & z < 0. \end{cases}$$

由此得到当 $z \in [0,\theta]$ 时随机变量 Z 的密度函数 $f_Z(z) = nz^{n-1}/\theta^n$, 进一步有

$$E\left[\hat{\theta}_n\right] = E[Z] = \int_0^\theta \frac{nz^n}{\theta^n} dz = \frac{n}{n+1}\theta,$$

因此 $\hat{\theta}$ 是 θ 的有偏估计. 另一方面有

$$E\left[Z^2\right] = \int_0^\theta \frac{nz^{n+1}}{\theta^n} dz = \frac{n}{n+2}\theta^2,$$

从而得到

$$\mathrm{Var}\left(\hat{\theta}_n\right) = \mathrm{Var}(Z) = E[Z^2] - (E[Z])^2 = \frac{n}{n+2}\theta^2 - (\frac{n\theta}{n+1})^2 = \frac{n}{(n+1)^2(n+2)}\theta^2,$$

于是有

$$\lim_{n\to\infty} E[\hat{\theta}_n] = \theta \quad \text{和} \quad \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n) = 0,$$

由此可得 $\hat{\theta}$ 是 θ 的有偏、但一致估计量.

9.3 区间估计

区间估计问题: 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, θ 为总体 X 的分布函数 $F(x,\theta)$ 的未知参数, 根据样本估计 θ 的范围 $(\hat{\theta}_1, \hat{\theta}_2)$, 其中 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$, 使得以较大的概率保证有 $\theta \in (\hat{\theta}_1, \hat{\theta}_2)$ 成立. 具体而言, 对任意给定 $\alpha \in (0,1)$, 有

$$\Pr\left[\hat{\theta}_1(X_1, X_2, \cdots, X_n) < \theta < \hat{\theta}_2(X_1, X_2, \cdots, X_n)\right] \geqslant 1 - \alpha.$$

定义 9.4 (置信区间与置信度) 设 $X_1, X_2 \cdots, X_n$ 是来自总体 X 的样本, 总体 X 的分布函数含未知参数 θ, 找出统计量 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$ $(\hat{\theta}_1 < \hat{\theta}_2)$, 使得

$$\Pr\left[\hat{\theta}_1 < \theta < \hat{\theta}_2\right] \geqslant 1 - \alpha$$

成立, 则称 $1-\alpha$ 为置信度, $[\hat{\theta}_1, \hat{\theta}_2]$ 为 θ 的置信度为 $1-\alpha$ 的置信区间.

注意: 置信区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 是随机区间, $1-\alpha$ 为该区间包含 θ 的概率/可靠程度. 若 $\alpha = 0.05$, 则置信度为 95%. 通常采用 95% 的置信度, 有时也可 99% 或 90%等. 说明:

i) $\hat{\theta}_2 - \hat{\theta}_1$ 反映了估计精度, 长度越小精度越大.

ii) α 反映了估计的可靠度, α 越小可靠度越高.

iii) 给定 α, 区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 的选取并不唯一确定, 通常选长度最小的一个区间.

置信区间的求解方法: **枢轴变量法**.