概率论与数理统计

第六章

韩潇
xhan011@ustc.edu.cn

第六章：　参数点估计

大纲：

- 参数点估计的概念
- 矩估计法
- 最大似然估计
- 优良性准则
- 点估计的大样本理论

6.3 最大似然估计

另一个参数估计法称为最大似然估计(MLE).

例 6.7 设鱼塘养了一塘鱼, 为了估计鱼的数量, 先一网打上 500 尾鱼, 在尾上涂以红色(称其为红尾鱼), 然后放回鱼塘, 几天后, 再打上一网鱼, 在网中的 400 尾鱼中发现有 60 尾红尾鱼. 试估计该鱼塘中鱼的数量.

- ◆假定两次捕捞均是随机的，且两次捕捞期间鱼的数量不变
- ◆设N为鱼塘中鱼的数量，k为第二网中红尾鱼的数量, $p_k(N)$为第二网中恰好有k尾红尾鱼的概率

$$p_k(N) = \frac{\binom{N-500}{400-k}\binom{500}{k}}{\binom{N}{400}}$$

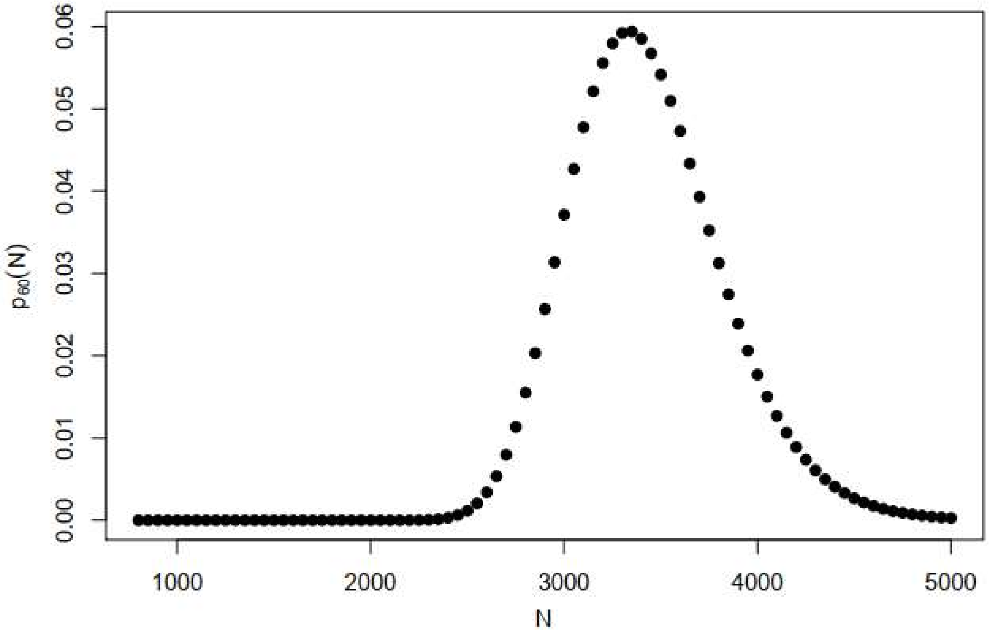

图 6.1: $p_{60}(N)$ 随 N 变化的取值

6.3 最大似然估计

受此启发，引入似然函数的定义

定义6.1 似然函数

设样本$\boldsymbol{X}=(X_1,X_2,\dots,X_n)$有联合概率密度函数或联合概率质量函数

$$f(\boldsymbol{x};\boldsymbol{\theta})=f(\boldsymbol{x};\theta_1,\theta_2,\dots,\theta_k)$$

这里参数$\boldsymbol{\theta}=(\theta_1,\theta_2,\dots,\theta_k)\in\Theta,\boldsymbol{x}=(x_1,x_2,\dots,x_n)$为样本$\boldsymbol{X}$的一个样本值. 当固定$\boldsymbol{x}$时把$f(\boldsymbol{x};\boldsymbol{\theta})$看成为$\boldsymbol{\theta}$的函数，称为似然函数(likelihood function),常记为$L(\boldsymbol{\theta};\boldsymbol{x})$或$L(\boldsymbol{\theta})$.

6.3 最大似然估计

- 对似然函数 $L(\theta;\boldsymbol{x})$可以理解为若$L(\theta;\boldsymbol{y}) = f(\boldsymbol{y};\theta) > L(\theta;\boldsymbol{x}) = f(\boldsymbol{x};\theta)$，则在观测时出现点$\boldsymbol{y}$ 的可能性要大一点

- 用似然函数最大的那个点 $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ 作为估计值，即

$$L(\theta_1^*, \ldots, \theta_k^*; x_1, \ldots, x_n) = \max_{(\theta_1,\ldots,\theta_k)\in\Theta} L(\theta_1, \ldots, \theta_k; x_1, \ldots, x_n)$$

- 这样的估计称为“最大似然估计”或者极大似然估计(Maximum Likelihood Estimate, **MLE**)

- $g(\theta_1^*, \ldots, \theta_k^*)$ 为 $g(\theta_1, \ldots, \theta_k)$ 的最大似然估计

6.3 最大似然估计

- 如果似然函数是严格单调的, 则似然函数的最大值在边界处达到, 从而得到最大似然估计
- 如果似然函数是光滑的, 且样本是简单随机样本, 则似然函数是n个因子(各自分布)的乘积, 可以先取自然对数
- 似然函数达到最大当且仅当对数似然函数$l(\theta) = \ln L(\theta)$达到最大
- 如果对数似然函数关于$(\theta_1, \ldots, \theta_k)$可微，则可通过求解下列方程得到驻点

$$\frac{\partial \ell}{\partial \theta_i} = 0, i = 1, \ldots, k,$$

- 若该方程组在$(\theta_1, \ldots, \theta_k)$定义域内有解，进一步验证是否为最大值点

6.3 最大似然估计

- 似然函数的最大值点可能会在边界上达到，所以要和边界值作比较．当似然函数不可导时，要用定义来求出最大似然估计

- 最大似然估计量也是统计量，也有二重性, 所以最大似然估计量$\hat{\theta}_{\mathrm{L}}(X_1, \ldots, X_n)$是统计量，这方便于研究其概率性质；最大似然估计值$\hat{\theta}_{\mathrm{L}}(x_1, \ldots, x_n)$ 是代入具体样本值后得到的具体数

- 在没有歧义的情况下，我们用最大似然估计来统指参数的最大似然估计量或估计值

6.3 最大似然估计

例 6.8 设 $(x_1,\dots,x_n)$ 是从正态总体 $N(\mu,\sigma^2)$ 中抽取的一个样本, 求 (μ,σ^2) 的最大似然估计.

例 6.9 设 $(x_1,\dots,x_n)$ 是从指数分布总体 $Exp(\lambda)$ 中抽取的一个样本, 求最大似然估计 $\hat{\lambda}_L$.

6.3 最大似然估计

例 6.10 设 $(x_1, \ldots, x_n)$ 是从均匀分布总体 $U(0, \theta)$ 中抽取的一个样本, 求最大似然估计 $\hat{\theta}_L$.

例 6.11 设总体 X 服从 $[\theta - 1/2, \theta + 1/2]$ 上的均匀分布, $\theta > 0$ 为参数, 求参数 θ 的最大似然估计.

6.3 最大似然估计

例 6.12 设 $(x_1, \ldots, x_n)$ 是从二项分布总体 $B(N, p)$ 中抽取的一个样本, 求最大似然估计 $\hat{p}_L$.

例 6.13 设总体分布为柯西分布, 其密度函数为

$$f(x, \theta) = \frac{1}{\pi[1 + (x - \theta)^2]},$$

其中 θ 为参数. 求其最大似然估计.

6.4 优良性准则

- 总体均值，可以用样本均值来估计，也可构造加权平均$\sum_{i=1}^{n} w_{\mathrm{i}} X_i$ 来估计，其中$\sum_{i=1}^{n} w_{\mathrm{i}} = 1$，$w_{\mathrm{i}} \geq 0$. 哪个更好？

- 若$\hat{\theta}_1, \hat{\theta}_2$都可以估计$\theta$，自然的想法是看估计量与真值的差

- 由于样本是随机变量，不能根据一个样本的表现来决定$\hat{\theta}_1, \hat{\theta}_2$的好坏，要考虑“整体性能”

  - 估计量是否具有某个特性，如“无偏性”

  - 比较估计量的某种数量指标，小者为优，如“均方误差”

6.4 优良性准则

定义 6.2 偏差与无偏性

设 $\hat{g}(X_1,...,X_n)$ 是 $g(\theta_1,...,\theta_k)$ 的一个估计量, 则称

$$\mathbb{E}_\theta \hat{g}(X_1,...,X_n) - g(\theta_1,...,\theta_k)$$

为估计量 $\hat{g}$ 的偏差. 如果对任一可能的 $(\theta_1,...,\theta_k) \in \Theta$, 都有

$$\mathbb{E}_{\theta_1,...,\theta_k} \hat{g}(X_1,...,X_n) = g(\theta_1,...,\theta_k), \tag{6.5}$$

则称 $\hat{g}$ 是 $g(\theta_1,...,\theta_k)$ 的一个无偏估计量.

- E_θ，$\mathrm{E}_{\theta_{1,...,}\theta_k}$是指期望在给定$\theta = (\theta_{1,...,}\theta_k)$下计算的, 换句话说，分布函数F(x; $\theta_{1,...,}\theta_k$)固定. 无偏性要求(6.5)对所有的$\theta \in \Theta$都成立.
- 无偏性就是没有系统误差

6.4 优良性准则

例 6.14 设总体分布的期望为 μ, 由样本可以构造两个估计量:$\hat{\theta}_1 = \overline{X}, \hat{\theta}_2 = \sum_{i=1}^{n} \omega_i X_i$, 其中 $\omega_i, i = 1, \ldots, n$ 为一组权. 问两个估计量是否为无偏估计?

例 6.15 设总体分布的方差为 σ^2, 证明样本方差是 σ^2 的无偏估计.

- 例6. 15说明矩估计m_2不是方差σ^2的无偏估计
- 从无偏性标准看，矩估计有时候不是一个很好的估计
- 相比于最大似然估计，矩估计不需要知道总体的分布，更便于使用

6.4 优良性准则

- 样本方差中，表面上看分子$\sum_{i=1}^{n}(X_i-\bar{X})^2$是n个变量的平方和，但由等式$\sum_{i=1}^{n}(X_i-\bar{X})=0$容易看出他们线性相关，非独立

- 真正能独立变化的变量只有n-1个，我们称独立变量的个数为“自由度”，所以在样本方差中的分母是自由度

- 为什么会减少？可以这样理解：原有n个i.i.d.样本$X_1,\dots,X_n$，因为均值μ未知，用$\bar{X}$去估计它，等于加了一个约束。因此，$\sum_{i=1}^{n}(X_i-\bar{X})^2$ 只有n-1个能独立变化的分量

6.4 优良性准则

例 6.16 设总体分布服从 $U(0,\theta)$, 证明 θ 的最大似然估计 $\hat{\theta}_L = X_{(n)}$ 不是 θ 的无偏估计.

设 $\hat{\theta}(X_1,\dots,X_n)$ 是 θ 的一个点估计, 如果

$$\lim_{n\to\infty} \mathbb{E}\hat{\theta}(X_1,\dots,X_n) = \theta,\ \forall\,\theta \in \Theta,$$

则称 $\hat{\theta}(X_1,\dots,X_n)$ 是 θ 的一个渐近无偏估计.

命题:样本矩是总体矩的渐近无偏估计

6.4 优良性准则

例 6.17 设样本 $X \sim B(N,p)$, N 已知而 p 未知. 令 $g(p) = 1/p$, 则参数 $g(p)$ 的无偏估计不存在.

证 采用反证法: 若不然, $g(p)$ 有无偏估计 $\hat{g}(X)$. 由于 X 只取 $0,1,\ldots,N$ 这些值, 令 $\hat{g}(X)$ 的取值用 $\hat{g}(i) = a_i$ 表示, $i = 0,1,\ldots,N$. 由 $\hat{g}(X)$ 的无偏性, 应有

$$\mathbb{E}_p(\hat{g}(X)) = \sum_{i=0}^{N} a_i \binom{N}{i} p^i (1-p)^{N-i} = 1/p, \quad 0 < p < 1.$$

于是有

$$\sum_{i=0}^{N} a_i \binom{N}{i} p^{i+1} (1-p)^{N-i} - 1 = 0, \quad 0 < p < 1$$

但上式左端是 p 的 $N+1$ 次多项式, 它最多在 $(0,1)$ 区间有 $N+1$ 个实根, 可无偏性要求对 $(0,1)$ 中的任一实数 p 上式都成立. 这个矛盾说明 $g(p) = 1/p$ 无的偏估计不存在. ■

由此，无偏估计不一定存在

6.4.2 最小方差无偏估计

- 有多种参数估计的方法，因此需要优良性准则衡量估计的好坏
- 直观上我们希望一个优良的估计量 偏差要小，波动也不能太大
- 可以用 $MSE_\theta(\hat{\theta}) = \mathbb{E}_\theta[\hat{\theta}(X_1, \ldots, X_n) - \theta]^2$ 来衡量, 越小越好.

6.4.2 最小方差无偏估计

- 注意到均方误差有如下性质

$$MSE_\theta(\hat{\theta}) = \mathrm{Var}_\theta(\hat{\theta}) + [\mathbb{E}_\theta(\hat{\theta} - \theta)]^2$$

- 第一部分衡量估计量的波动，第二部分衡量偏差，显然无偏估计有如下性质

$$MSE_\theta(\hat{\theta}) = \mathrm{Var}_\theta(\hat{\theta})$$

6.4.2 最小方差无偏估计

定义 6.3 有效性

设 $\hat{\theta}_1, \hat{\theta}_2$ 都是总体参数 θ 的无偏估计, 方差存在, 若

$$\mathrm{Var}_\theta(\hat{\theta}_1) \leq \mathrm{Var}_\theta(\hat{\theta}_2), \forall \theta \in \Theta \tag{6.6}$$

且至少存在一个 θ, 使上式不等号成立, 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效. ♣

6.4.2 最小方差无偏估计

例 6.18 设总体分布的均值为 μ, 方差为 σ^2. 现在用 $\hat{\theta}_1 = \overline{X}$ 和 $\hat{\theta}_2 = \sum_{i=1}^{n} \omega_i X_i$ 来估计 μ, 哪个更有效? 其中 $\sum_{i=1}^{n} \omega_i = 1, \omega_i \geq 0$.

例 6.19 设总体分布服从 $U(0,\theta)$, 试比较 θ 的矩估计 $\hat{\theta}_M = 2\overline{X}$ 和修正的最大似然估计 $\hat{\theta}_L = ((n+1)/n)X_{(n)}$ 的有效性.

是否存在这样一个无偏估计, 它在所有无偏估计中具有最小的方差?

定义 6.4 最小方差无偏估计

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

$$\mathrm{Var}_\theta(\hat{\theta}) \le \mathrm{Var}_\theta(\hat{\theta}_1), \quad \forall \theta \in \Theta,$$

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) ♣

6.4.3 克拉默-拉奥方差下界

记 $\hat{g}(\mathbf{X})$ 为待估参数 $g(\theta)$ 的任一方差有限的无偏估计，令

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

利用柯西-施瓦茨不等式，我们有Cov(X,Y)≤Var(X)Var(Y)，可知

$$Cov^2(\hat{g}(\mathbf{X}),S(\mathbf{X},\theta))\leq \mathrm{Var}(S(\mathbf{X},\theta))\cdot\mathrm{Var}(\hat{g}(\mathbf{X}))$$

我们假设一些正则性条件：对一切可能的$\theta\in\Theta,\{x:f(x,\theta)>0\}$ 与θ无关，对θ求导与积分号可交换，涉及到的函数可导和存在等

则我们可以推出

6.4.3 克拉默-拉奥方差下界

$$\mathbb{E}_\theta S(\mathbf{X},\theta) = 0,$$

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

$$\langle\hat{g}(\mathbf{X}), S(\mathbf{X},\theta)\rangle = g'(\theta)$$

因此我们得到如下命题

命题 6.1 克拉默-拉奥方差下界

对 $g(\theta)$ 的任一无偏估计 $\hat{g}(\mathbf{X})$, 在正则条件下有

$$\mathrm{Var}_\theta(\hat{g}(\mathbf{X})) \geq (g'(\theta))^2[nI(\theta)]^{-1}. \tag{6.8}$$

这就是克拉默-拉奥方差下界.

6.4.3 克拉默-拉奥方差下界

例 6.22 设 $(X_1, \ldots, X_n)$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的一个样本, 其中 σ^2 已知. 对 μ 作点估计.

例 6.23 设总体为指数分布 $Exp(\lambda)$, 问 λ^{-1} 的无偏估计 $\overline{X}$ 是否为 MVUE?

6.5 点估计的大样本理论

- 当样本量趋向于无穷时点估计量的性质称为大样本性质，估计量在样本量固定时的性质称为小样本性质.
- 关于估计量或一般统计量的大样本性质的讨论构成了数理统计学的一个非常重要的部分，称为大样本统计理论.
- 大样本理论直到今天依然是统计研究的核心工作之一

定义 6.5 相合性

设 $\hat{\theta}(X_1, \ldots, X_n)$ 是参数 θ 的一个点估计，如果样本量 $n \to \infty$ 时有

$$\hat{\theta}(X_1, \ldots, X_n) \xrightarrow{P} \theta, \tag{6.9}$$

则称 $\hat{\theta}(X_1, \ldots, X_n)$ 是 θ 的一个 (弱) 相合估计量(consistent estimator). ♣

6.5 点估计的大样本理论

相合性是对一个估计量的基本要求. 如果一个估计量没有相合性, 则无论样本量多大, 我们也不能把未知参数估计到任意精度. 这种估计量显然是不可取的.

例 6.26 证明 $m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 是总体方差 σ^2 的相合估计.

6.5 点估计的大样本理论

- 可以证明，矩估计是总体矩的相合估计．一般而言，最大似然估计也是待估参数的相合估计.

- 点估计的大样本性质中另一个重要的准则是其分布极限的特点，称为渐近正态性

定义 6.6 渐近正态性

设 $\hat{\theta}(X_1, \ldots, X_n)$ 是参数 θ 的一个点估计, 设它的方差存在, 记

$$\mathrm{Var}_\theta[\hat{\theta}(X_1, \ldots, X_n)] = \sigma_n^2(\theta),$$

如果样本量 $n \to \infty$ 时有

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\hat{\theta}(X_1, \ldots, X_n) - \theta}{\sigma_n(\theta)} \leq x\right) = \Phi(x), \forall\, x \in \mathbb{R}, \tag{6.10}$$

则称估计量 $\hat{\theta}(X_1, \ldots, X_n)$ 有渐近正态性. ♣

6.5 点估计的大样本理论

- 渐近正态性提供了估计量的一个近似分布，利用它我们才能完成相关的统计推断.

- 在一般条件下,矩估计和最大似然估计都会有渐近正态性,例如总体均值μ的点估计$\bar{X}_{\mathrm{n}}$,由林德伯格-莱维中心极限定理，我们有

$$\sqrt{n}(\overline{X}_n - \mu)/\sigma \xrightarrow{\mathcal{L}} N(0,1)$$

例 6.28 设 $X_1, \ldots, X_n$ 为来自总体 F 的一组简单随机样本, 设总体 F 期望和方差分别为 μ 和 σ^2. 则易知样本均值 $\bar{X}$ 和样本方差 S_n^2 分别为 μ 和 σ^2 的相合估计, 且有

$$\sqrt{n}(\overline{X}_n - \mu)/S_n \xrightarrow{\mathcal{L}} N(0,1).$$


Thank
you!