第6章 大样本性质----渐近评价

《统计推断》 第10章

Motivation

- 精确分布难以求出，无法做出评价

- 渐近分布往往比较简单，使得有限样本下无法做到的评价变成了常规。

- 比如：自助法和M-估计

内容

- 点估计
  - 相合性
  - 有效性
  - 渐近相对效率
  - Bootstrap
- 稳健性
- 假设检验的大样本性质
- 区间估计的大样本性质

点估计相合性

统计推断　渐近评价　相合性

点估计的相合性

- 定义：一个估计量序列$W_n = W_n(X_1, \cdots, X_n)$是参数 θ 的一个相合估计量序列(Consistent sequence of estimators), 如果对于 $\forall \epsilon > 0$ 和每个 $\theta \in \Theta$,

$$\lim_{n \to +\infty} P_\theta(|W_n - \theta| < \epsilon) = 1$$

- 注：该定义和依概率收敛的区别：以 θ 为指标的一整族概率结构，不同的θ 相关联的概率结构是不同的。

样本均值的相合性

- 设 $X_1, \cdots, X_n \sim N(\theta, 1)$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_n \sim N(\theta, \frac{1}{n})$$

$$\begin{aligned}
P_\theta(|\overline{X} - \theta| < \epsilon) &= \int_{\theta-\epsilon}^{\theta+\epsilon} (\frac{n}{2\pi})^{1/2} e^{-(n/2)(x-\theta)^2} dx \\
&= \int_{-\epsilon\sqrt{n}}^{\epsilon\sqrt{n}} (\frac{1}{2\pi})^{1/2} e^{-\frac{1}{2}t^2} dt \\
&= P(-\epsilon\sqrt{n} < Z < \epsilon\sqrt{n}) \to 1
\end{aligned}$$

- 因此样本均值是参数的相合估计量序列。

相合性判定定理

- 定理：如果W_n是参数θ的一个估计量序列，如果对于每个$\theta \in \Theta$ 都满足

  $(1).\ \lim\limits_{n\to+\infty} \mathrm{Var}_\theta W_n = 0$

  $(2).\ \lim\limits_{n\to+\infty} \mathrm{Bias}_\theta W_n = 0$

  则W_n是参数θ的一个相合估计量序列.

证明：由切比雪夫不等式及MSE分解即得

$$P_\theta(|W_n - \theta| \geq \epsilon) \leq \frac{E[(W_n-\theta)^2]}{\epsilon^2}$$

$$E_\theta[(W_n-\theta)^2] = \mathrm{Var}_\theta(W_n) + \mathrm{Bias}_\theta^2(W_n)$$

定理

- 定理：如果 W_n 是参数 θ 的一个相合估计量序列，设 $a_1, \cdots, a_n, \cdots$ 和 $b_1, \cdots, b_n, \cdots$ 是常数序列，满足,

$$\lim_{n \to +\infty} a_n = 1, \lim_{n \to +\infty} b_n = 0$$

则序列 $U_n = a_n W + b_n$ 是参数 θ 的一个相合估计量序列。

MLE的相合性

- 定理：设 $X_1, X_2, \cdots \sim f(x|\theta)$, $L(\theta|x)$ 是似然函数，而 $\hat{\theta}$ 表示 θ 的MLE. 设$\tau(\theta)$是 θ 的一个连续函数，那么在一定的正则性条件下，对于每个 $\epsilon > 0$ 和每个$\theta \in \Theta$,有

$$\lim_{n \to +\infty} P_\theta(|\tau(\hat{\theta}) - \tau(\theta)| \geq \epsilon) = 0$$

- 即 $\tau(\hat{\theta})$ 是 $\tau(\theta)$ 的一个相合估计量。

点估计有效性

统计推断 | 渐近评价 | 有效性

有效性

- 相合性考虑的是估计量的渐近精确性，而有效性关心的是估计量的方差。
- 多数情况下方差的极限为0, 因此需要添加一个常数因子 k_n , 得到一个有限的非0极限：

$$\begin{cases} \lim_{n\to+\infty} Var(T_n) = 0 \\ \lim_{n\to+\infty} k_n Var(T_n) = \tau^2 < +\infty \end{cases}$$

极限方差

- 定义: 对于一个估计量 T_n, 如果

$$\lim_{n\to+\infty} k_n \mathrm{Var} T_n = \tau^2 < +\infty$$

其中 k_n 是一个常数序列，则 τ^2 称之为极限方差(limiting variance)或者方差的极限

渐近方差

- 定义：对于一个估计量 T_n, 如果有依分布收敛

$$k_n(T_n - \tau(\theta)) \to N(0, \sigma^2)$$

则参数 σ^2 称之为估计量 T_n 的渐近方差或者称之为 T_n 的极限分布的方差。

例子：渐近方差和极限方差不相同

- 设有分层模型

$$Y_n|W_n = w_n \sim N(0, w_n + (1-w_n)\sigma_n^2), W_n \sim Bernoulli(p_n)$$

$$Var(Y_n) = p_n + (1-p_n)\sigma_n^2$$

- 只有在$\lim\limits_{n\to+\infty}(1-p_n)\sigma^2 < +\infty$ 时，极限方差存在
- 另一方面由

$$P(Y_n < a) = p_n P(Z < a) + (1-p_n)P(Z < a/\sigma_n)$$

例子：渐近方差和极限方差不相同

- 当 $p_n \to 1, \sigma_n \to +\infty, (1-p_n)\sigma_n^2 \to +\infty$ 时，

$$P(Y_n < a) \to P(Z < a)$$

- 即 $Y_n \to N(0,1)$.
- 此时极限方差趋于无穷，但渐近方差为1。

渐近有效

- 定义：一个估计序列 W_n 关于一个参数 $\tau(\theta)$ 是渐近有效的，如果

$$\begin{cases} \sqrt{n}[W_n - \tau(\theta)] \xrightarrow{L} N(0, v(\theta)) \\ v(\theta) = \dfrac{[\tau'(\theta)]^2}{E_\theta[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2]} \end{cases}$$

MLE的渐近有效性

- 定理6.1.6: 设 $X_1, X_2, \cdots \sim f(x|\theta), \theta$ 的极大似然估计 $\hat{\theta}$, 设 $\tau(\theta)$ 是 θ 的一个连续函数。那么在一定的正则条件下有，

$$\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] \xrightarrow{L} N(0, v(\theta))$$

其中 $v(\theta)$ 是Cramer-Rao下界。也就是说，$\tau(\hat{\theta})$ 是 $\tau(\theta)$ 的一个相合且渐近有效的估计量。

证明大意

- 就对数似然函数Taylor展开

$$l'(\theta|x) = l'(\theta_0|x) + (\theta - \theta_0)l''(\theta_0|x) + \cdots$$

- 取 $\theta = \hat{\theta},\ l'(\hat{\theta}|x) = 0$

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\sqrt{n}\frac{l'(\theta_0|x)}{l''(\theta_0|x)} = \frac{\frac{1}{\sqrt{n}}l'(\theta_0|x)}{-\frac{1}{n}l''(\theta_0|x)}$$

证明大意

- 令 $I(\theta_0) = E[l'(\theta_0|X)]^2 = \dfrac{1}{v(\theta_0)}$ 表示观测的信息数，可以证明

$$\frac{1}{\sqrt{n}} l'(\theta_0|x) \xrightarrow{L} N(0, I(\theta_0))$$

$$-\frac{1}{n} l''(\theta_0|X) \xrightarrow{P} I(\theta_0)$$

- 于是

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{L} N\left(0, \frac{1}{I(\theta_0)}\right)$$

渐近正态蕴含相合性

- 由MLE的渐近有效性

$$\sqrt{n}\frac{W_n-\mu}{\sigma}\xrightarrow{L} Z$$

- 于是

$$W_n-\mu=\frac{\sigma}{\sqrt{n}}(\sqrt{n}\frac{W_n-\mu}{\sigma})\xrightarrow{L}\lim_{n\to}(\frac{\sigma}{\sqrt{n}})Z=0$$

- 依分布收敛到一个点等价于依概率收敛，因此 W_n 是 μ 的一个相合估计量。

MLE相合性：正则性条件

- A1: 观测样本iid
- A2: 参数可识别。即如果 $\theta \neq \theta' \Rightarrow f(x|\theta) \neq f(x|\theta')$
- A3: 各个密度 $f(x|\theta)$ 具有共同的支撑集合，并且密度关于参数可导。
- A4: 参数空间 Ω 包含一个开集 ω, 真实参数 θ_0 为该开集 ω 的一个内点。

MLE的渐近有效性：正则性条件

- A5: 对于每个 $x \in \mathcal{X}$, 密度函数$f(x|\theta)$关于θ 三次可导，其三次导数是θ 的连续函数，并且积分$\int f(x|\theta)dx$ 可以在积分号下微分三次。
- A6: 对任何$\theta_0 \in \Omega$存在一个正数 c和一个函数M(x) (二者都可以依赖于θ_0) 使得,

$$\left|\frac{\partial^3}{\partial\theta^3}\log f(x|\theta)\right| \le M(x), \forall x \in \mathcal{F}, \theta_0 - c < \theta < \theta_0 + c.$$

$$E_{\theta_0}|M(x)| < +\infty$$

正则性条件

- 这些条件是相当一般的，而且几乎在通常情况下都能得到满足
- 有一个条件需要强调：概率密度函数或者概率质量函数的支撑集，也就是似然函数的支撑集必须与参数无关

函数的方差近似

- 如果MLE渐近有效，再利用 Δ 方法

$$\sqrt{n}(Y_n - \theta) \xrightarrow{L} N(0, \sigma^2) => \sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{L} N(0, \sigma^2[g'(\theta)]^2)$$

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{L} N\left(0, \frac{1}{I(\theta_0)}\right)$$

$$\begin{aligned} Var(h(\hat{\theta})|\theta_0) &\approx \frac{h'(\theta_0)}{I_n(\theta_0)} \\ &= \frac{h'(\theta_0)}{E_{\theta_0}\left(-\frac{\partial^2}{\partial\theta^2}\log L(\theta_0|X)\right)} \\ &\approx \frac{h'(\theta)|_{\theta=\hat{\theta}}}{\left(-\frac{\partial^2}{\partial\theta^2}\log L(\theta|X)\right)|_{\theta=\hat{\theta}}} \end{aligned}$$

例子：近似二项方差(I)

- 设 $X_1, \cdots, X_n \sim \mathrm{Bernoulli}(p)$, 那么p的极大似然估计为 $\hat{p} = \sum_i X_i/n$, 计算得到

$$Var_p(\hat{p}) = \frac{p(1-p)}{n}$$

- 其估计可以如下给出

$$\widehat{Var_p(\hat{p})} = \frac{\hat{p}(1-\hat{p})}{n}$$

例子：近似二项方差(II)

- 另一方面，由前面的函数方差近似公式

$$\widehat{Var_p(\hat{p})} \approx -\frac{1}{-\frac{\partial^2}{\partial p^2}\log L(p|x)_{|p=\hat{p}}}$$

- 对数似然

$$\begin{aligned}\log L(p|x) &= \sum_i X_i \log(p) + (n - \sum X_i) log(1-p) \\ &= n\hat{p}\log(p) + n(1-\hat{p})\log(1-p)\end{aligned}$$

例子：近似二项方差 (III)

$$
\begin{aligned}
\frac{\partial^2}{\partial p^2} \log L(p|x) &= -\frac{\sum_i X_i}{p^2} - \frac{n - \sum_i X_i}{(1-p)^2} \\
&= -\frac{n\hat{p}}{p^2} - \frac{n(1-\hat{p})}{1-p)^2}
\end{aligned}
$$

- 于是

$$
\begin{aligned}
\widehat{Var_p(\hat{p})} &\approx -\frac{1}{-\frac{\partial^2}{\partial p^2} \log L(p|x)_{|p=\hat{p}}} \\
&= \frac{1}{\frac{n\hat{p}}{\hat{p}^2} + \frac{n(1-\hat{p})}{(1-\hat{p})^2}} = \frac{\hat{p}(1-\hat{p})}{n}
\end{aligned}
$$

- 和前面的直接计算结果一致。

例子：近似二项方差 (IV)

- 应用定理6.1.6就可以断言$\hat{p}$ 的渐近有效性，

$$\sqrt{n}(\hat{p}-p) \to N(0, p(1-p))$$

- 再由Slusky定理得到

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}} \to N(0,1)$$

例子：近似二项方差 (V)

- 胜率p/(1-p)的估计量 $\hat{p}/(1-\hat{p})$ 的方差估计

$$\widehat{var}(\frac{\hat{p}}{1-\hat{p}}) = \frac{\left[\frac{\partial}{\partial p}(\frac{p}{1-p})\right]^2\Big|_{p=\hat{p}}}{-\frac{\partial^2}{\partial p^2}\log L(p|x)|_{p=\hat{p}}}$$

$$= \frac{[\frac{(1-p)+p}{(1-p)^2}]^2|_{p=\hat{p}}}{\frac{n}{p(1-p)}|_{p=\hat{p}}}$$

$$= \frac{\hat{p}}{n(1-\hat{p})^3}$$

- 而且这个估计量是渐近有效的

注

- MLE方差近似在多数情况下是成功的；

- 当函数非单调的时候有可能出现问题。因为此时导数将会有符号改变，有可能导致一个被低估的渐近方差。

反例：函数p(1-p)的方差

- 由MLE估计，其方差为： $\hat{p}(1-\hat{p})$
- 另一方面

$$
\begin{aligned}
\widehat{Var}(\hat{p}(1-\hat{p}) &= \frac{\left[\frac{\partial}{\partial p}\left(p(1-p)\right)\right]^2|_{p=\hat{p}}}{-\frac{\partial^2}{\partial p^2}\log L(p|x)|_{p=\hat{p}}} \\
&= \frac{(1-2p)^2|_{p=\hat{p}}}{\frac{n}{p(1-p)}|_{p=\hat{p}}} \\
&= \frac{\hat{p}(1-\hat{p})(1-2\hat{p})^2}{n}
\end{aligned}
$$

- 当 $\hat{p}=1/2$ 时为0, 显然是对方差的低估；非单调是出现这种问题的原因

渐近相对效率

- 定义：如果两个估计量 W_n 和 V_n 满足

$$\begin{cases} \sqrt{n}[W_n - \tau(\theta)] \xrightarrow{L} N(0, \sigma_W^2) \\ \sqrt{n}[V_n - \tau(\theta)] \xrightarrow{L} N(0, \sigma_V^2) \end{cases}$$

那么 V_n 关于 W_n 的渐近相对效率(ARE)是

$$ARE(V_n, W_n) = \frac{\sigma_W^2}{\sigma_V^2}$$

例子：Poisson估计量的ARE (I)

- Poisson(λ)分布，$P(X=0)=e^{-\lambda}$
- 估计一
- 令 $Y_i = I(X_i = 0), Y_i \sim Bernoulli(e^{-\lambda})$

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$E(\hat{\tau}) = e^{-\lambda}, var(\hat{\tau}) = \frac{e^{-\lambda}(1-e^{-\lambda})}{n}$$

例子：Poisson估计量的ARE (II)

- 估计二(MLE)
- Poisson参数的MLE $\hat{\lambda} = \sum_{i=1}^{n} X_i/n,$ 那么 $e^{-\lambda}$的MLE就是 $e^{-\hat{\lambda}}$，有近似

$$E(e^{-\hat{\lambda}}) \approx e^{-\lambda}, Var(e^{-\hat{\lambda}}) \approx \frac{\lambda e^{-2\lambda}}{n}$$

- 所以渐近相对效率

$$ARE(\hat{\tau}, e^{-\hat{\lambda}}) = \frac{\lambda e^{-2\lambda}}{e^{-\lambda}(1-e^{-\lambda})} = \frac{\lambda}{e^{\lambda}-1}$$

- 渐近相对效率在 $\lambda = 0$ 时达到最大值1，而之后快速变小趋于0；

Bootstrap

统计推断 → 渐近评价 → Bootstrap

- Bootstrap方法是上世纪70年代由Efron提出
- 非参数Bootstrap
- 参数Bootstrap

非参数Bootstrap方法

- 设 $X_1, \cdots, X_n \sim F(x)$ 是iid的n个样本, 有放回抽样出相同容量的样本，称之为Bootstrap样本
- 重复多次，从原样本中生成多个Bootstrap样本
- 基于这些bootstrap样本对总体F进行统计推断，称之为非参数Bootstrap方法

估计量的标准差的Bootstrap估计

- 设 $X_1, \cdots, X_n \sim F(x), \theta$ 是待估参数

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

- 设第i 组Bootstrap样本给出的估计为

$$\hat{\theta}_i^* = \hat{\theta}_i^*(X_1^*, X_2^*, \cdots, X_n^*)$$

- 则标准方差估计为

$$\begin{cases} Var_B^*(\hat{\theta}) = \dfrac{1}{B-1}\displaystyle\sum_{i=1}^{B}(\hat{\theta}_i^* - \overline{\hat{\theta}^*})^2 \\ \overline{\hat{\theta}^*} = \dfrac{1}{B}\displaystyle\sum_{i=1}^{B}\hat{\theta}_i^* \end{cases}$$

Bootstrap置信区间 (I)

- 设 $X_1,\cdots,X_n \sim F(x),\theta$ 是待估参数

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

- 可以用Bootstrap方法来求 θ 的1- α 置信区间
- 随机抽取B组Bootstrap样本，

$$X_{i1}^*,X_{i2}^*,\cdots,X_{in}^*,i=1,2,\cdots,B$$

- 相应的估计为

$$\hat{\theta}_i^*=\hat{\theta}_i^*(X_{i1}^*,\cdots,X_{in}^*),i=1,\cdots,B$$

Bootstrap置信区间(II)

- 将它们从小到大排序

$$\hat{\theta}^*_{(1)} \le \hat{\theta}^*_{(2)} \le \cdots \le \hat{\theta}^*_{(n)}$$

- 利用这些估计值的经验分布近似 $\hat{\theta}$ 的分布，求出其近似分位数 $\hat{\theta}^*_{\alpha/2}, \hat{\theta}^*_{1-\alpha/2}$，使得

$$P(\hat{\theta}^*_{\alpha/2} < \hat{\theta}^* < \hat{\theta}^*_{\alpha/2}) = 1-\alpha$$

- 于是近似有

$$P(\hat{\theta}^*_{\alpha/2} < \theta < \hat{\theta}^*_{\alpha/2}) = 1-\alpha$$

Bootstrap置信区间(III)

- 取整数k_1, k_2,

$$k_1 = [B \times \alpha/2], k_2 = [B \times (1-\alpha/2)]$$

- 分位数可估计为

$$\hat{\theta}^*_{\alpha/2} = \hat{\theta}^*_{(k_1)}, \hat{\theta}^*_{1-\alpha/2} = \hat{\theta}^*_{(k_2)}$$

- 最后得到 1- α 置信区间

$$(\hat{\theta}^*_{(k_1)}, \hat{\theta}^*_{(k_2)})$$

参数Bootstrap方法

- 设 $X_1, \cdots, X_n \sim f(x|\theta)$, 设 $\hat{\theta}$ 是 θ 的MLE，那么可以抽取样本

$$X_1^*, X_2^*, \cdots, X_n^* \sim f(x|\hat{\theta})$$

- 重复上面的方法可以生成B组样本，按下面的公式估计 $\hat{\theta}$ 的方差

$$Var_B^*(\hat{\theta}) = \frac{1}{B-1} \sum_{i=1}^{B} (\hat{\theta}_i^* - \overline{\hat{\theta}^*})^2$$

稳健性

统计推断 → 渐近评价 → 稳健性

- 均值和中位数
- M-估计量

稳健性提出

- 统计方法要求
  - 在假定模型下应当有一个合理的好(最佳或者接近最佳)效率
  - 稳健性：对于假定模型的微小偏离应该仅引起性能的轻微损失
  - 对模型大一些的偏离也不应该导致灾难性后果

样本均值的稳健性

- 设 $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$. 样本均值 $\overline{X}$ 达到了CR下界，$\overline{X}$ 是均值的最优估计，故满足(1)
- 采用混合模型来考察模型的微小偏离

$$X_i \sim \begin{cases} N(\mu, \sigma^2), & \text{with probability} \quad 1-\delta \\ f(x), & \text{with probability} \quad \delta \end{cases}$$

- 如果f(x) 是Cauchy分布，这时候可得出

$$Var(\overline{X}) = +\infty$$

样本均值的稳健性 (II)

- 考虑要求(3). 考察异常值的影响。比如

$$X_{(n)} = x \to +\infty$$

- 这样的观测值的影响是“灾难性的”

崩溃值

- 定义：设 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ 是容量为n的顺序样本，设T_n是基于这些样本的统计量. T_n 具有崩溃值 $b \in [0,1]$. 如果对于每个 $\epsilon > 0$ 都有

$$\lim_{X_{([(1-b)n])} \to +\infty} T_n < +\infty, \lim_{X_{([(1-b+\epsilon)n])} \to +\infty} T_n = +\infty$$

- 由此定义，样本均值的崩溃值是0, 但中位数的崩溃值是0.5

中位数的渐近正态性 (I)

- 设 $X_1, \cdots, X_n \sim F(x), \mu$ 是总体中位数，M_n是样本中位数. 考虑

$$Y_i = \begin{cases} 1 & \text{if } X_i \le \mu + a/\sqrt{n}, \\ 0 & \text{otherwise.} \end{cases}$$

$$Y_i \sim Bernoulli(n, p_n), p_n = F(\mu + a/\sqrt{n})$$

- 为避免计算上复杂性，不妨设n为奇数，此时有下述随机事件的等价性

$$\{M_n \le \mu + a/\sqrt{n}\} \Leftrightarrow \{\sum_{i=1}^{n} Y_i \ge (n+1)/2\}$$

中位数的渐近正态性 (II)

- 计算得

$$P(\sqrt{n}(M_n-\mu)\le a)=P\left(\frac{\sum_i Y_i-np_n}{\sqrt{np_n(1-p_n)}}\ge\frac{\frac{n+1}{2}-np_n}{\sqrt{np_n(1-p_n)}}\right)$$

- 当 $n\to+\infty, p_n\to F(\mu)=\dfrac{1}{2}$

$$\frac{\frac{n+1}{2}-np_n}{\sqrt{np_n(1-p_n)}}\to -2aF'(\mu)=-2af(\mu)$$

中位数的渐近正态性 (III)

- 计算过程如下

$$\lim_{n\to+\infty}\frac{\frac{n+1}{2}-np_n}{\sqrt{np_n(1-p_n)}}=\lim_{n\to+\infty}\frac{\frac{1}{2}-n(p_n-\frac{1}{2})}{\frac{1}{2}\sqrt{n}}$$

$$=-2\lim_{n\to+\infty}\frac{\int_{\mu}^{\mu+\frac{a}{\sqrt{n}}}f(x)dx}{\frac{1}{\sqrt{n}}}=-2af(\mu)$$

- 由中心极限定理

$$\frac{\sum_i Y_i-np_n}{\sqrt{np_n(1-p_n)}}\xrightarrow{L}N(0,1)$$

中位数的渐近正态性 (IV)

- 综上有

$$
\begin{aligned}
P(\sqrt{n}(M_n-\mu)\leq a) &\rightarrow P\left(Z\geq -2af(\mu)\right)\\
&= P\left(Z\leq 2af(\mu)\right), n\rightarrow +\infty
\end{aligned}
$$

- 即

$$\sqrt{n}(M_n-\mu)\xrightarrow{L} N\left(0,\frac{1}{(2f(\mu))^2}\right)$$

$$\sqrt{n}(\overline{X}-\mu)\xrightarrow{L} N(0,\sigma^2),\sigma^2=D(X)$$

中位数和样本均值的渐近相对效率

- 利用上述渐近方差，渐近相对效率为

$$ARE(M, \overline{X}) = 4\sigma^2 f^2(\mu)$$

M-估计

- 回顾

$$\begin{cases} \overline{X} = \underset{a}{\text{Argmin}} \sum_{i=1}^{n} (X_i - a)^2 \\ M_n = \underset{a}{\text{Argmin}} \sum_{i=1}^{n} |X_i - a| \end{cases}$$

- 如何兼顾样本均值和中位数估计

Huber稳健性准则

- 准则函数

$$\sum_{i=1}^{n} \rho(x_i - a)^2$$

$$\rho(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \le k, \\ k|x| - \frac{1}{2}k^2 & \text{if } |x| > k \end{cases}$$

- 参数 k 调节了估计是否接近中位数估计
- 使准则函数 $\sum_{i=1}^{n} \rho(x_i - \theta)$ 达到最小的估计量称为M-估计量

M-估计的渐近分布 (I)

- 令 $\psi(x) = \rho'(x)$, 那么M-估计满足估计方程

$$\sum_{i=1}^{n} \psi(x_i - \theta) = 0$$

- 假设函数$\rho(x)$对称，导函数$\psi(x)$单调递增，则M-估计唯一

M-估计的渐近分布 (II)

- 设 θ_0 为参数真值，在真值处Taylor展开

$$\sum_{i=1}^{n} \psi(x_i - \theta) = \sum_{i=1}^{n} \psi(x_i - \theta_0) + (\theta - \theta_0) \sum_{i=1}^{n} \psi'(x_i - \theta_0) + \cdots$$

- 代入估计M-估计 $\hat{\theta}_M$ 有

$$0 = \sum_{i=1}^{n} \psi(x_i - \theta_0) + (\hat{\theta}_M - \theta_0) \sum_{i=1}^{n} \psi'(x_i - \theta_0) + \cdots$$

M-估计的渐近分布 (III)

- 忽略余项得

$$\sqrt{n}(\hat{\theta}_M - \theta_0) = \frac{-\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(x_i - \theta_0)}{\frac{1}{n}\sum_{i=1}^n \psi'(x_i - \theta_0)}$$

- 如果假设 θ_0 满足$E_{\theta_0}\psi(X-\theta_0)=0$, 于是

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(x_i - \theta_0) = \sqrt{n}[-\frac{1}{n}\sum_{i=1}^n \psi(x_i - \theta_0)] \xrightarrow{L} N(0, E_{\theta_0}\psi(X-\theta_0)^2)$$

M-估计的渐近分布(IV)

- 而根据大数定理有

$$\frac{1}{n}\sum_{i=1}^{n}\psi'(x_i-\theta_0)\xrightarrow{p}E_{\theta_0}\psi'(X-\theta_0)$$

- 上述二式合在一起得到

$$\sqrt{n}(\hat{\theta}_M-\theta_0)\to N\left(0,\frac{E_{\theta_0}\psi(X-\theta_0)^2}{[E_{\theta_0}\psi'(X-\theta_0)]^2}\right)$$

Huber估计量的极限分布(I)

- 设 $X_1, \cdots, X_n \sim F(x-\theta)$, 密度为 $f(x-\theta), f$ 关于0点对称。那么

$$\psi(x) = \begin{cases} x & \text{if } |x| \leq k \\ k & \text{if } x > k \\ -k & \text{if } x < -k \end{cases}$$

- 容易验证

$$E_\theta \psi(X-\theta) = \int_{\theta-k}^{\theta+k} (x-\theta) f(x-\theta) dx$$

$$- k \int_{-\infty}^{\theta-k} f(x-\theta) dx + k \int_{\theta+k}^{+\infty} f(x-\theta) dx = 0$$

Huber估计量的极限分布(II)

- 方差计算

$$E_\theta \psi'(X-\theta) = \int_{\theta-k}^{\theta+k} f(x-\theta)dx = P_0(|X| \le k)$$

$$\begin{aligned} E_\theta \psi(X-\theta)^2 = &\int_{\theta-k}^{\theta+k} (x-\theta)^2 f(x-\theta)dx + k^2 \int_{\theta+k}^{+\infty} f(x-\theta)dx \\ &+ k^2 \int_{-\infty}^{\theta-k} f(x-\theta)dx \\ &= \int_{-k}^{k} x^2 f(x)dx + 2k^2 \int_{k}^{+\infty} f(x)dx \end{aligned}$$

- 于是渐近方差为

$$\frac{\int_{-k}^{k} x^2 f(x)dx + 2k^2 P_0(|X| > k)}{[P_0(|X| \le k)]^2}$$

M-估计和MLE的比较(I)

- 因为

$$\begin{cases} \sqrt{n}(\hat{\theta}_M - \theta_0) \to N\left(0, \dfrac{E_{\theta_0}\psi(X-\theta_0)^2}{[E_{\theta_0}\psi'(X-\theta_0)]^2}\right) \\ \sqrt{n}(\hat{\theta}_{MLE} - \theta_0) \to N\left(0, \dfrac{1}{E_{\theta_0} l'(\theta|X)^2}\right) \end{cases}$$

- 由渐近相对效率的定义

$$ARE(\hat{\theta}_M, \hat{\theta}_{MLE}) = \frac{[E_{\theta_0}\psi'(X-\theta_0)]^2}{E_{\theta_0}\psi(X-\theta_0)^2 E_{\theta_0} l'(\theta|X)^2}$$

M-估计和MLE的比较(II)

- 注意到

$$E_\theta \psi'(X-\theta) = -E_\theta\left[\frac{d}{d\theta}\psi(X-\theta)\right] = -\int [\frac{d}{d\theta}\psi(X-\theta)] f(x-\theta)dx$$

$$0 = \frac{d}{d\theta}\int \psi(x-\theta)f(x-\theta)dx$$

$$= \int \left[\frac{d}{d\theta}\psi(x-\theta)\right] f(x-\theta)dx + \int \psi(x-\theta)\left[\frac{d}{d\theta}f(x-\theta)\right]dx$$

- 即

$$-\int \left[\frac{d}{d\theta}\psi(x-\theta)\right] f(x-\theta)dx = \int \psi(x-\theta)\left[\frac{d}{d\theta}f(x-\theta)\right]dx$$

$$= \int \psi(x-\theta)\left[\frac{d}{d\theta}\log f(x-\theta)\right] f(x-\theta)dx$$

M-估计和MLE的比较(III)

- 即

$$E_\theta \psi'(X-\theta) = -E_\theta\left[\frac{d}{d\theta}\psi(X-\theta)\right] = E_\theta\left[\psi(X-\theta)l'(\theta|x)\right]$$

- 从而渐近相对效率

$$ARE(\hat{\theta}_M, \hat{\theta}_{MLE}) = \frac{[E_\theta \psi(X-\theta)l'(\theta|X)]^2}{E_\theta \psi(X-\theta)^2 E_\theta l'(\theta|X)^2}$$

- 由Schwartz不等式得

$$ARE(\hat{\theta}_M, \hat{\theta}_{MLE}) \leq 1$$

假设检验的大样本性质

统计推断 › 渐近评价 › 假设检验

- LRT的渐近分布
- Wald检验
- Score检验

简单H_0下LRT的渐近分布

- 定理：设$X_1, \cdots, X_n \sim f(x|\theta), \hat{\theta}$ 是MLE, 且密度函数f满足通常的正则性条件(保证MLE的有效性)。考虑检验问题$H_0 : \theta = \theta_0 \Leftrightarrow H_1 : \theta \neq \theta_0,$ 那么在H_0下，当$n \to +\infty$时

$$-2 \log \lambda(X) \to \chi_1^2$$

$$\lambda(x) = \frac{\sup_{\Theta_0} L(\theta|x)}{\sup_{\Theta} L(\theta|x)}$$

证明思路

- 证明：在 $\hat{\theta}$ 点邻域内进行Taylor展开对数似然函数

$$l(\theta|x) = l(\hat{\theta}|x) + l'(\hat{\theta}|x)(\hat{\theta}-\theta) + l''(\hat{\theta}|x)(\hat{\theta}-\theta)^2/2 + \cdots$$

- 将真值 θ_0 代入上式得

$$-2\log\lambda(x) = -2l(\theta_0|x) + 2l(\hat{\theta}|x) \approx \frac{(\theta-\theta_0)^2}{-l''(\hat{\theta}|x)}$$

- 上式中分母即为观察信息数 $\hat{I}_n(\hat{\theta})$, 依概率收敛到 $I(\theta_0)$, 于是有极大似然估计的有效性定理和Slutsky定理得结论

复杂H_0下LRT的渐近分布

- 定理：设$X_1, \cdots, X_n \sim f(x|\theta)$, 当 $\theta \in \Theta_0$ 时 $f(x|\theta)$ 满足通常的正则性条件(保证MLE有效)，那么在 $H_0 : \theta \in \Theta_0$ 下，当 $n \to +\infty$ 时，

$$-2 \log \lambda(X) \xrightarrow{L} \chi_v^2$$

其中v是 $\theta \in \Theta$ 指明的自由参数和 $\theta \in \Theta_0$ 指明的自由参数之差

自由度的计算

- 若全空间参数 $\Theta \subset R^q$, 包含了中的一个开子集
- 若零假设参数 $\Theta_0 \subset R^p$, 包含了中的一个开子集
- 那么自由参数为

$$v = q - p$$

Wald检验

- 若

$$Z_n = \frac{W_n - \theta}{S_n} \xrightarrow{L} N(0,1)$$

- 则可以利用Z_n来构造检验拒绝域

Wald检验

- 对于检验问题: $H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$
- 考虑统计量

$$Z_n = \frac{W_n - \theta_0}{S_n}$$

- 在H_0下，Z_n依分布收敛到N(0,1)
- 于是拒绝域为：$R = \{x : |Z_n| > z_{\alpha/2}\}$
- 该检验的真实水平为 α

Wald检验

- 当 $\theta \neq \theta_0$ 时(即在被择假设下)

$$Z_n = \frac{W_n - \theta_0}{S_n} = \frac{W_n - \theta}{S_n} + \frac{\theta - \theta_0}{S_n}$$

- 由于通常 $S_n \to 0$, 于是 $Z_n \to \infty$
- 于是

$$P_\theta(R) = P_\theta(|Z_n| > z_{\alpha/2}) \to 1, n \to +\infty$$

- 即此检验是一个真实水平为 α , 渐近功效为1的检验。

从MLE构造Wald检验

- 若W_n是参数 θ的MLE，由前面的讨论W_n的标准误差可以估计为

$$\frac{1}{\sqrt{I_n(W_n)}} \approx \frac{1}{\sqrt{\widehat{I}_n(W_n)}}$$

- 其中（观测信息数）

$$\widehat{I}_n(W_n) = -\frac{\partial}{\partial\theta^2}\log L(\theta|x)|_{\theta=W_n}$$

Score检验

- 令

$$S(\theta) = \frac{\partial}{\partial\theta}\log f(X|\theta) = \frac{\partial}{\partial\theta}\log L(\theta|X)$$

- 有

$$\begin{cases} E_\theta S(\theta) = 0 \\ Var_\theta S(\theta) = E_\theta(\frac{\partial}{\partial\theta}\log L(\theta|X)^2 \\ \qquad\qquad\qquad = -E_\theta(\frac{\partial^2}{\partial\theta^2}\log L(\theta|X) = I_n(\theta) \end{cases}$$

Score检验

- 对于检验问题：$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$
- 于是构建统计量

$$Z_S = \frac{S(\theta_0)}{\sqrt{I_n(\theta_0)}}$$

- 当H_0成立时，$Z_S \xrightarrow{L} N(0,1)$
- 于是可以构造拒绝域

$$R = \{x : |Z_S| > z_{\alpha/2}\}$$

区间估计的大样本性质

统计推断 > 渐近评价 > 区间估计

- 基于MLE的区间估计
- 基于Wald检验的区间估计

基于MLE的近似区间估计

- 设 $X_1, \cdots, X_n \sim f(x|\theta), \hat{\theta}$ 是 θ 的MLE, $\hat{\theta}$ 的任一函数 $h(\hat{\theta})$ 的方差可以由下式近似

$$\widehat{Var}(h(\hat{\theta})|\theta) \approx \frac{[h'(\theta)]^2|_{\theta=\hat{\theta}}}{-\frac{\partial^2}{\partial\theta^2}\log L(\theta|x)|_{\theta=\hat{\theta}}}$$

- 而且

$$\frac{h(\hat{\theta}) - h(\theta)}{\sqrt{\widehat{Var}(h(\hat{\theta})|\theta)}} \xrightarrow{L} N(0,1), n \to +\infty$$

基于MLE的近似区间估计

- 于是我们可以给出近似的置信区间

$$h(\hat{\theta}) - Z_{\alpha/2}\sqrt{\widehat{Var}(h(\hat{\theta})|\theta)} \leq h(\theta) \leq h(\hat{\theta}) + Z_{\alpha/2}\sqrt{\widehat{Var}(h(\hat{\theta})|\theta)}$$

基于Wald检验的区间估计

- 若

$$Z_n = \frac{W_n - \theta}{S_n} \xrightarrow{L} N(0,1)$$

- 则区间可以由下式给出

$$W_n - Z_{\alpha/2} S_n \le \theta \le W_n + Z_{\alpha/2} S_n$$