第二节 点估计

- **主讲内容**
  1. 相合性与渐近有效性
  2. 矩估计
  3. 极大似然估计（MLE）
  4. 稳健估计
- **点估计的优良性**

小样本	大样本
无偏性	相合性
有效性	渐近有效性

- 【定义10.1.1】 称一个估计量序列$W_n = W_n(X_1, \cdots, X_n)$ 是参数θ的一个**相合**或**一致(Consistent)**估计量序列, 如果

$$W_n \xrightarrow{P} \theta, \quad \forall \mathbb{P}_\theta.$$

- 【例10.1.2】 ($\overline{X}$ 的相合性) 设$X_1, X_2, \cdots$ $i.i.d. \sim N(\theta, 1)$，问序列$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ 是否是θ的一个相合估计量序列？

Theorem (定理10.1.3)

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

(i) $\lim_{n\to\infty} Var_\theta(W_n) = 0$,

(ii) $\lim_{n\to\infty} Bias_\theta(W_n) = 0$,

则W_n是参数θ的一个相合估计量序列。

举例说明

- **例2.1** 设$X_1, X_2, \cdots$ $i.i.d. \sim U(0, \theta)$，$\theta > 0$，证明$X_{(n)}$是θ的一个相合估计量。
- **例2.2** 设$X_1, X_2, \cdots$ $i.i.d.$，总体分布c.d.f为$F_\theta(x)$，$\theta > 0$，满足$F_\theta(\theta) = 1$，而$\forall x < \theta$，$F_\theta(x) < 1$。证明$X_{(n)}$ 是θ 的一个相合估计量。

Theorem (定理10.1.5)

设W_n是参数θ的一个相合估计量序列，$a_1, a_2, \cdots$，和$b_1, b_2, \cdots$，是常数序列，满足

(i) $\lim_{n\to\infty} a_n = 1$,

(ii) $\lim_{n\to\infty} b_n = 0$,

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

有效性

- 【定义10.1.7】 对于一个估计量T_n，如果$lim_{n\to\infty} k_n Var(T_n) = \tau^2 < \infty$，其中$\{k_n\}$是一个常数序列，则$\tau^2$称为极限方差（limiting variance）或方差的极限。
- 【例10.1.8】 设$X_1, X_2, \cdots$ $i.i.d.$, $\mathbb{E}X_1 = \mu$, $Var(X_1) = \sigma^2$, 求$\overline{X}_n$的极限方差。

问题 设$\mu \neq 0$，计算$T_n = 1/(\overline{X}_n)$的极限方差。

- 【定义10.1.9】 对于一个估计量T_n，假定$\exists\{k_n\}$常数序列，s.t.

$$k_n\left(T_n - \tau(\theta)\right) \xrightarrow{L} N(0, \sigma^2), \quad as \quad n \to \infty.$$

则参数σ^2称为T_n的渐近方差或T_n的极限分布的方差。

- 【例10.1.10】 分层模型

$$\begin{aligned} Y_n | W_n = \omega_n &\sim N(0, \omega_n + (1-\omega_n)\sigma_n^2), \\ W_n &\sim Bernoulli(p_n). \end{aligned}$$

分别讨论Y_n的极限方差和渐近方差的存在性。

作业1 习题10：1, 并讨论估计量的极限方差和渐近方差.

- 【定义10.1.11】 称一个估计量序列W_n是一个参数$\tau(\theta)$的**渐近有效估计** (Asymptotic Effective Estimators, AEE)，如果

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

而且

$$v(\theta) = \frac{[\tau'(\theta)]^2}{\mathbb{E}_\theta\left\{\left[\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right]^2\right\}}.$$

i.e. W_n的渐近方差达到了Cramér-Rao下界。

- 【定义10.1.16】 如果两个估计量W_n 和V_n满足

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

V_n关于W_n的**渐近相对效率**(Asymptotic Relative Efficiency, ARE)是

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

- 符号约定：总体k阶中心矩$\mu_k = \mathbb{E}(X-\mathbb{E}X)^k$;
  样本k阶中心矩$\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$.
- 总体k阶原点矩$\alpha_k = \mathbb{E}(X^k)$; 样本$k$阶原点矩$\widehat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n}X_i^k$;

Theorem (【2】定理1.9)

设$X_1, X_2, \cdots$ $i.i.d. \sim X$，$\mu_4 = \mathbb{E}(X-\mathbb{E}X)^4$存在有限，记$\mathbb{E}(X)=\mu$, $Var(X)=\sigma^2$, 若函数$h(x)$的四阶导数存在且有界，则有

$$\begin{aligned}
\mathbb{E}[h(\overline{X}_n)] &= h(\mu)+\frac{1}{2n}h''(\mu)\sigma^2+O(n^{-2}) \\
Var[h(\overline{X}_n)] &= \frac{1}{n}[h'(\mu)]^2\sigma^2+\frac{1}{2n^2}\{h'(\mu)h''(\mu)\mu_3+[h''(\mu)]^2\sigma^4 \\
&\quad +h'(\mu)h'''(\mu)\sigma^4\}+O(n^{-3})
\end{aligned}$$

- 【例5.5.23】 设随机变量X的期望为$\mathbb{E}_\mu X=\mu\neq 0$，令$g(\mu)=\frac{1}{\mu}$，请给出$g(\mu)$的一个矩估计，并给出其期望和方差的近似估计。

作业2 习题10：4

矩估计的相合性

Theorem (【0】定理3.2.1, 【2】定理2.9)

设$X_1, X_2, \cdots$ $i.i.d. \sim F$，参数$\theta = g(\alpha_1, \cdots, \alpha_k, \mu_1, \cdots, \mu_s)$，这里$\alpha_k, \mu_s$有限。若$g$连续，则矩估计$\hat{\theta}_{MOM} = g(\hat{\alpha}_1, \cdots, \hat{\alpha}_k, \hat{\mu}_1, \cdots, \hat{\mu}_s)$是$\theta$的（强）相合估计。

- 【0】例3.2.9 设$X_1, X_2, \cdots$ $i.i.d. \sim U(\theta_1, \theta_2)$, 分别给出$\theta_1$和$\theta_2$的一个强相合估计。
- 【0】定义3.2.2 设$X_1, X_2, \cdots$ $i.i.d. \sim f(x|\theta), \theta \in \Theta$, $\hat{g}_n(\vec{X})$ 是$g(\theta)$的弱相合矩估计。称$\hat{g}_n(\vec{X})$ 是$g(\theta)$的**相合渐近正态**（Consistent Asymptotic Normal, 简记CAN）估计, 若存在函数$A_n(\theta)$ 和$B_n(\theta)$, 其中$B_n(\theta) > 0$, $\forall \theta \in \Theta$, $s.t.$ 当$n \to \infty$ 时

$$\frac{\hat{g}_n(\vec{X}) - A_n(\theta)}{B_n(\theta)} \xrightarrow{L} N(0, 1).$$

矩估计的渐近正态性

- 在存在连续偏导数条件下，矩估计是相合渐近正态（CAN）估计。

Theorem (【0】定理3.2.2, 【2】定理2.10)

设$X_1, X_2, \cdots$ $i.i.d. \sim X$，总体X的α_{2k}有限。记$\vec{\theta} = (\theta_1, \cdots, \theta_d)^T$, 其中$\theta_i = g_i(\alpha_1, \cdots, \alpha_k)$, 若$g_i$ 关于α_j有连续偏导数, $i = 1, \cdots, d$, $j = 1, \cdots, s$, 则对$\vec{\theta}$的矩估计$\hat{\vec{\theta}}_{MOM} = \left(g_1(\hat{\boldsymbol{\alpha}}), \cdots, g_d(\hat{\boldsymbol{\alpha}})\right)^T$, 其中$\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_k)^T$, 有

$$\sqrt{n}(\hat{\vec{\theta}}_{MOM} - \vec{\theta}) \xrightarrow{L} N_d(\vec{0}, G\Sigma G^T) \tag{1}$$

其中$G = \left(\frac{\partial g_i}{\partial \alpha_j}\right)_{d\times k}$，$\Sigma$是$k \times k$阶矩阵，其$(i, j)$元素为$\alpha_{i+j} - \alpha_i\alpha_j$.

注 若各个参数分量$\theta_i = g_i(\alpha_1, \mu_2, \cdots, \mu_r)$，这里$\mu_r$有限, g_i 关于$\alpha_1, \mu_2, \cdots, \mu_r$ 均有连续偏导数, 则矩估计

$$\hat{\vec{\theta}}_{MOM} = \left(g_1(\hat{\alpha}_1, \hat{\boldsymbol{\mu}}), \cdots, g_d(\hat{\alpha}_1, \hat{\boldsymbol{\mu}})\right)^T$$

同样是$\vec{\theta}$ 的CAN估计，渐近正态结果(1)中Σ的元素σ_{ij} 具体参考【0】定理3.2.3.

- **例2.3** 设$X_1, X_2, \cdots$ $i.i.d. \sim X$，总体X的α_4有限。求变异系数$\nu = \frac{\sigma}{\mu}$的矩估计并讨论它的渐近正态性。(参考【0】例3.2.10)

作业3 【0】习题3：1, 并讨论S^2的渐近正态性。

Theorem (定理10.1.6)

设$X_1, X_2, \cdots$ $i.i.d. \sim f(x|\theta)$，满足$\mathbb{E}_\theta|\log f(X|\theta')| < \infty$, $\forall \theta, \theta' \in \Theta$. 记$\hat{\theta}_n$为$\theta$的$MLE$, 并设$\tau(\theta)$ 是θ 的一个连续函数. 那么当$f(x|\theta)$满足如下正则条件$A1 - A4$ 时，对于每个$\epsilon > 0$和每个$\theta \in \Theta$，有

$$\lim_{n\to\infty} \mathbb{P}(|\tau(\hat{\theta}_n) - \tau(\theta)| \geq \epsilon) = 0.$$

即，$\tau(\hat{\theta}_n)$是参数$\tau(\theta)$的一个相合估计量。

条件A

- A1 $X_1, \cdots, X_n$ $i.i.d. \sim f(x|\theta)$;
- A2 如果$\theta \neq \theta'$，则$f(x|\theta) \neq f(x|\theta')$;
- A3 各个密度$f(x|\theta)$有共同的支撑集，并且$f(x|\theta)$关于θ可导；
- A4 参数空间Θ包含一个开集$\mathcal{O}$，真实参数值θ_0为该开集的一个内点。

MLE的渐近有效性

A5 对于每个$x \in \mathcal{X}$，密度$f(x|\theta)$关于θ二阶连续可导；

A6 对任意$\theta_0 \in \Theta$，存在一个正数c和一个函数$M(x)$（二者皆可依赖θ_0），使得对于所有$x \in \mathcal{X}$, $\theta_0 - c < \theta < \theta_0 + c$ 成立

$$\left|\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right| \le M(x), \text{以及} \mathbb{E}_{\theta_0}|M(X)| < \infty.$$

Theorem (定理10.1.12, 【0】定理3.3.2)

设$X_1, X_2, \cdots\ i.i.d. \sim f(x|\theta)$，满足$\mathbb{E}_\theta|\log f(X|\theta')| < \infty$, $\forall\theta, \theta' \in \Theta$, 并记$\hat{\theta}_n$ 为θ 的MLE. 设$\tau(\theta)$是θ 的一个连续函数，且在θ处可微, $\tau'(\theta) \neq 0$, 那么当$f(x|\theta)$满足正则条件$A1 - A6$ 时，有

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

即，$\tau(\hat{\theta}_n)$是参数$\tau(\theta)$的一个相合且渐近有效的估计量。（未完待续）

- **定理续** 其中
$$v(\theta)=\frac{|\tau'(\theta)|^2}{\mathrm{I}(\theta)}.$$
这里$\mathrm{I}(\theta)$为总体Fisher信息量，
$$\mathrm{I}(\theta)=\mathbb{E}_\theta\Big|\frac{\partial}{\partial}\ln f(X|\theta)\Big|^2.$$
- **【定义10.1.11】** 设T_n是$\tau(\theta)$的相合估计，$\tau(\theta)$可微，$\mathrm{I}(\theta)$存在有限，且
$$\sqrt{n}[T_n-\tau(\theta)]\xrightarrow{L}N(0,v(\theta)),$$
则称如下$e(\theta,T_n)$为T_n的**渐近效率**:
$$e(\theta,T_n)=\frac{[\tau'(\theta)]^2}{\mathrm{I}(\theta)v(\theta)}.$$
如果$e(\theta,T_n)=1$, 则称T_n是$\tau(\theta)$的**渐近有效估计**。

注1 满足Cramer-Rao 不等式正则条件的分布族，例如指数族，定理条件一般都满足（注意二阶导存在、连续、绝对可积性）。但非Cramer-Rao正则族也可能成立相合与渐近正态性。

- 【2】例2.28 设$X_1, \cdots, X_n$ $i.i.d. \sim \frac{1}{2}e^{-|x-\theta|}$, $x \in \mathbb{R}$, $\theta \in \mathbb{R}$, 则$\hat{\theta}_{MLE} = m_{1/2}$中位数，满足当$n \to \infty$时，

$$\sqrt{n}(m_{1/2} - \theta) \xrightarrow{L} N(0, 1).$$

注2 如果条件A不满足，则MLE的相合性和渐近正态性并不一定成立。

- 【2】例2.29 设$X_1, \cdots, X_n$是来自两点分布的一个简单样本，

$$\mathbb{P}(X_1 = 1) = 1 - \mathbb{P}(X_1 = 0) = \begin{cases} \theta, & \theta\text{为有理数} \\ 1-\theta, & \theta\text{为无理数} \end{cases}$$

则由$\hat{\theta}_{MLE} = \overline{X}_n$ 可得

$$\hat{\theta}_{MLE} \xrightarrow{a.s.} \begin{cases} \theta, & \theta\text{为有理数} \\ 1-\theta, & \theta\text{为无理数} \end{cases}$$

i.e. $\hat{\theta}_{MLE}$不是θ的相合估计。

- **例2.4** 设$X_1, \cdots, X_n$ $i.i.d. \sim U(0, \theta), \theta > 0$, 则$\hat{\theta}_{MLE} = X_{(n)}$，满足$n(\theta - X_n) \xrightarrow{L} Exp(\theta^{-1})$, i.e. $\hat{\theta}_{MLE}$ 的渐近分布非正态。

一 若分布族满足条件$A1-A6$，则可计算大样本的近似方差。

- MLE渐近有效$\Rightarrow$ C-R下界估计MLE方差。两步估计(*)

$$
\begin{aligned}
Var_\theta\{h(\hat{\theta})\} &\approx \frac{[h'(\theta)]^2}{n\mathrm{I}(\theta)} = \frac{[h'(\theta)]^2}{n\mathbb{E}_\theta\left(-\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right)} \\
&\approx \frac{[h'(\theta)]^2|_{\theta=\hat{\theta}}}{-n\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)|_{\theta=\hat{\theta}}}
\end{aligned}
$$

分母：观测信息数$\mathrm{I}_n(\hat{\theta})$.

- 【例10.1.14】 设$X_1, X_2, \cdots$ $i.i.d. \sim Bernoulli(p)$, 记$\hat{p}=\overline{X}_n$, 分别求胜算估计$g(\hat{p})=\frac{\hat{p}}{1-\hat{p}}$和方差估计$h(\hat{p})=\hat{p}(1-\hat{p})$ 这两个统计量的方差的两步估计。

二 比较估计量（的渐近方差）– 渐近相对效率（ARE）

- 【例10.1.17】 $X_1, \cdots, X_n \ i.i.d. \sim Poisson(\lambda)$,

$$\tau = h(\lambda) = \mathbb{P}_\lambda(X_1 = 0) = e^{-\lambda},$$

求τ的如下两个估计量$\hat{\tau}_1$和$\hat{\tau}_2$ 的ARE:

1. 估计量一：$\hat{\tau}_1 = \frac{1}{n}\sum_{i=1}^n Y_i$, 这里$Y_i = I(X_i = 0)$;
2. 估计量二：$\hat{\tau}_2 = e^{-\hat{\lambda}}$ (MLE), 其中$\hat{\lambda} = \overline{X}$.

- 【例10.1.18】 $X_1, \cdots, X_n \ i.i.d. \sim Gamma(\alpha, \beta)$, p.d.f

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}.$$

求$\mu = \alpha\beta$的$\hat{\mu}_{MLE}$关于其矩估计$\hat{\mu}_{MOM}$的ARE。

作业4 习题10：3, 并讨论估计量$\hat{\theta}_{MLE}$的渐近正态性.

作业4 【0】习题3：5, 并求$\hat{\theta}_1$关于$\hat{\theta}_2$的ARE.

多维参数情形

- 【2】定义2.8 设$X_1, X_2, \cdots$ $i.i.d. \sim X$, p.d.f 为$f(x|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$, $k \in \mathbb{N}^+$. 记$\ell(\boldsymbol{\theta}|X) = \log f(X|\boldsymbol{\theta})$，假设$\forall \boldsymbol{\theta} \in \Theta$, 如下随机向量有定义
$$S_{\boldsymbol{\theta}}(X) = \left(\frac{\partial \ell(\boldsymbol{\theta}|X)}{\partial \theta_1}, \cdots, \frac{\partial \ell(\boldsymbol{\theta}|X)}{\partial \theta_k} \right)^T,$$
且满足$\mathbb{E}_{\boldsymbol{\theta}} S_{\boldsymbol{\theta}}(X) = 0$, $\mathbb{E}_{\boldsymbol{\theta}} \|S_{\boldsymbol{\theta}}(X)\|^2 < \infty$, $\forall \boldsymbol{\theta} \in \Theta$. 则称$S_{\boldsymbol{\theta}}(X)$的协方差阵
$$\mathrm{I}(\boldsymbol{\theta}) = Var_{\boldsymbol{\theta}}(S_{\boldsymbol{\theta}}(X)) = \mathbb{E}_{\boldsymbol{\theta}}[S_{\boldsymbol{\theta}}(\mathbf{X}) S_{\boldsymbol{\theta}}^T(X)]$$
为总体X的Fisher信息矩阵，简称Fisher信息。

- 多元参数情形下MLE的相合性与渐近正态性：条件同定理10.1.12，关于θ的高阶可导性改为关于$\theta_i, i=1,\cdots,k$ 的高阶偏导，则

Theorem (多元参数MLE的CAN)

假设*Fisher*信息矩阵$I(\boldsymbol{\theta})$正定，记$\hat{\boldsymbol{\theta}}_n$是参数$\boldsymbol{\theta}$的*MLE*，则$\hat{\boldsymbol{\theta}}_n$是参数$\boldsymbol{\theta}$的一个相合估计量，且

$$\begin{aligned}\sqrt{n}[\tau(\hat{\boldsymbol{\theta}}_n)-\tau(\boldsymbol{\theta})] &= \frac{1}{\sqrt{n}}[\nabla\tau(\boldsymbol{\theta})]I^{-1}(\boldsymbol{\theta})\sum_{i=1}^{n}S_{\boldsymbol{\theta}}(X_i)+o_P(1)\\ &\xrightarrow{L} N\left(0,[\nabla\tau]I^{-1}(\boldsymbol{\theta})[\nabla\tau]^T\right)\end{aligned}$$

这里$\nabla\tau=\nabla\tau(\boldsymbol{\theta})=\left(\frac{\partial\tau(\boldsymbol{\theta})}{\partial\theta_1},\cdots,\frac{\partial\tau(\boldsymbol{\theta})}{\partial\theta_k}\right)$.

- 给定评价标准，最佳估计量是在假定模型正确的前提下给出。如果模型不正确，则不能保证最佳性。
- 假定模型有小的或中等偏离情形⇒解决方案：稳健估计量（Robust Estimator）
- 稳健性定义：
  1. 在假定模型下具有一个合理的（最佳或接近最佳）好效率；
  2. 在上述最佳意义下，对于假定模型的微小偏离仅引起性能的轻微损伤；
  3. 对于模型的大一些的偏离也不会导致灾难性后果。

均值和中位数

- 【例10.2.1】（样本均值的稳健性） 假定模型$X_1, \cdots, X_n\ i.i.d. \sim N(\mu, \sigma^2)$, 此时$Var(\overline{X}_n) = \frac{\sigma^2}{n}$ ⇔C-R 下界. i.e. $\overline{X}$ 为μ 的UMVUE，满足稳健性条件1。
- 实际模型（微小偏离）– δ污染模型（δ-contamination model），

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

其中$f(x)$是某个其它分布的概率密度/质量函数。假设其均值为θ，方差τ^2，则

$$Var(\overline{X}_n) = (1-\delta)\frac{\sigma^2}{n} + \delta\frac{\tau^2}{n} + \frac{\delta(1-\delta)}{n}(\theta-\mu)^2.$$

- 如果$\theta \approx \mu$ 且$\tau \approx \sigma$，则$\overline{X}_n$接近最佳，满足稳健性条件2。
- 糟糕情形：$f(x)$为Cauchy分布p.d.f, 此时$Var(\overline{X}_n) = \infty$。

- 稳健性条件3：考虑异常观测值的出现会不会导致灾难性后果？
- 【定义10.2.2】 设T_n是一个基于顺序样本$X_{(1)}, \cdots, X_{(n)}$的统计量，称T_n具有崩溃值（breakdown value）b, $0 \le b \le 1$, 如果对于每一个$\varepsilon > 0$，都有

$$\lim_{X_{([(1-b)n])} \to \infty} T_n < \infty \quad \text{和} \quad \lim_{X_{([(1-(b+\varepsilon))n])} \to \infty} T_n = \infty$$

i.e. b为趋于无穷大的样本值占整个样本的比例。

注1 样本均值的崩溃值$b = 0$，即任何比例的样本值趋于无穷都会导致$\overline{X}_n$也趋于无穷，影响是灾难性的。

注2 样本中位数在这种变化下不变，崩溃值$b = 50\%$，对极端观测值不敏感。也即，样本中位数相对样本均值，在稳健性方面有所改善，但在其它方面是否有所损失？

- 最佳性比较准则：样本中位数对样本均值的渐近相对效率（ARE）
- 【例10.2.1（续）】设$f(x)$为$N(\mu,\tau^2)$的p.d.f, 则

$$\begin{aligned}\sqrt{n}(\overline{X}_n-\mu) &\xrightarrow{L} N(0,\sigma_1^2),\\ \sqrt{n}(m_{1/2}-\mu) &\xrightarrow{L} N(0,\sigma_2^2),\end{aligned}$$

其中

$$\begin{aligned}\sigma_1^2 &= (1-\delta)\sigma^2+\delta\tau^2\\ \sigma_2^2 &= \frac{\pi}{2}\left(\frac{1-\delta}{\sigma}+\frac{\delta}{\tau}\right)^{-2}\end{aligned}$$

则$\mathrm{ARE}(m_{1/2},\overline{X}_n)=\frac{\sigma_1^2}{\sigma_2^2}$.

1. 当τ接近σ或δ很小时，$\mathrm{ARE}(m_{1/2},\overline{X}_n)<1$，即$\overline{X}_n$优于$m_{1/2}$；
2. 当$\tau=4\sigma$, $\delta=0.05$ 或$\tau=10\sigma$, $\delta=0.01$ 时，$\mathrm{ARE}(m_{1/2},\overline{X}_n)>1$，即$m_{1/2}$优于$\overline{X}_n$.

具体参考【2】例2.44.

注 样本中位数对样本均值的ARE，当分布的尾部越重则得到的ARE越大。也即，在重尾分布情况下，用中位数性能会改善。参考例10.2.4。

M-估计量

- 我们使用的很多统计量是最小化一个特别的准则的结果。例如令$\rho(\theta|x) = -\log f(x|\theta)$, 则

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \prod_{i=1}^{n} f(x_i|\theta) = \arg\min_{\theta} \sum_{i=1}^{n} \rho(\theta|x_i).$$

- M-估计是MLE的延申（Huber,1954）.
- 【2】定义2.19 设$X_1, \cdots, X_n$ 是来自某总体的一个样本，$\rho(\theta|x)$为一选定非负函数，若估计量$\hat{\theta} = \hat{\theta}(\mathbf{X})$满足

$$\sum_{i=1}^{n} \rho(\hat{\theta}|X_i) = \min_{\theta} \sum_{i=1}^{n} \rho(\theta|X_i)$$

则称$\hat{\theta}$为θ的一个M-估计量（M-estimator）。
- 若$\rho(\theta|x)$关于θ可微，记$\psi(\theta|x) = \frac{\partial\rho(\theta|x)}{\partial\theta}$，则M-估计为

$$\sum_{i=1}^{n} \psi(\theta|X_i) = 0$$

的解。

- **例2.5** 分别令准则函数$\rho_1(x) = x^2$, $\rho_2(x) = |x|$, 则
  1. $\overline{X}_n = \arg\min_\theta \sum_{i=1}^n \rho_1(X_i - \theta)$,
  2. $m_{1/2} = \arg\min_\theta \sum_{i=1}^n \rho_2(X_i - \theta)$,
- 【例10.2.5】 Huber估计量。令

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

我们称使$\sum_{i=1}^n \rho(x_i - \theta)$达到最小的$\theta$的估计量为Huber估计量。
- 举例说明，数据

$$x = -1.28, -0.96, -0.46, -0.44, -0.26, -0.21, -0.063, 0.39, 3, 6, 9$$

$\overline{x} = 1.33$，中位数$x_{(6)} = -0.21$，Huber估计量

k	0	1	2	3	4	5	6	8	10
估计值	-0.21	0.03	-0.04	0.29	0.41	0.52	0.87	0.97	1.33

参考表10.2.1 。

相合性与渐近正态性

Theorem (相合性，【2】定理2.19)

设$X_1,\cdots,X_n$ 是来自$F(x)$的一个样本，$\psi(x)$在$(-\infty,\infty)$上非降，并设θ是如下方程的唯一解：$\lambda(t)=:\mathbb{E}\psi(X-t)=0$. 假定方程$\sum_{i=1}^n\psi(X_i-\theta)=0$ 对一切n和$X_1,\cdots,X_n$有解，以$\hat\theta_n$记它的一个解，则$\hat\theta_n\xrightarrow{a.s.}\theta$.

Theorem (渐近正态性，【2】定理2.20)

设【2】定理2.19条件满足，并设存在$\epsilon>0$,使在$(\theta-\epsilon,\theta+\epsilon)$内有: (1) $\lambda(\theta)$连续；(2) $\sigma^2(t)=:Var[\psi(X-t)]>0$存在有限且连续, 则

$$\sqrt{n}\lambda(\hat\theta_n)/\sigma(\theta)\xrightarrow{L}N(0,1).$$

进一步，若$\lambda'(\theta)$存在且不为0，则

$$\sqrt{n}\lambda'(\theta)(\hat\theta_n-\theta)/\sigma(\theta)\xrightarrow{L}N(0,1).$$

- **推论** 在【2】定理2.20 条件下，M-估计量具有如下渐近正态性：

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N\left(0, \frac{\mathbb{E}_\theta[\psi(X-\theta)]^2}{[\mathbb{E}_\theta\psi'(X-\theta)]^2}\right)$$

- 【例10.2.6】 Huber估计量的渐近分布。设$X_1, \cdots, X_n$ $i.i.d. \sim f(x-\theta)$, 其中f关于0对称，则由(2)得到

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

计算可得，Huber估计量的渐近正态分布其均值为θ，方差

$$\frac{\int_{-k}^{k} y^2 f(y)dy + k^2\mathbb{P}_\theta(|X|>k)}{[\mathbb{P}_{\theta=0}(|X|\leq k)]^2}。$$

- 几种不同分布考察Huber估计量的ARE。
- 【例10.2.5】 Huber估计量的ARE,k=1.5.

	正态分布	罗辑斯蒂克分布	双指数分布
与均值比较	0.96	1.08	1.37
与中位数比较	1.51	1.31	0.68

作业5 习题10: 29(b)

渐近相对效率

- 为了得到稳健性，M-估计$\hat{\theta}_M$在效率方面可能放弃了什么?
- 记总体分布的对数似然$\ell(\theta|x) = \log f(x-\theta)$，则

$$\mathbb{E}_\theta \psi'(X-\theta) = -\mathbb{E}_\theta\Big[\frac{\partial}{\partial\theta}\psi(x-\theta)\Big] = \mathbb{E}_\theta\big[\psi(x-\theta)\ell'(\theta|X)\big].$$

- 比较M-估计量和MLE的渐近方差

$$ARE(\hat{\theta}_M, \hat{\theta}_{MLE}) = \frac{\big[\mathbb{E}_\theta\big[\psi(x-\theta)\ell'(\theta|X)\big]\big]^2}{\mathbb{E}_\theta[\psi(x-\theta)]^2\mathbb{E}_\theta[\ell'(\theta|X)]^2} \leq 1.$$

- 一个M-估计量的效率总是比MLE低，只有当ψ和ℓ'成比例时（$\psi = \ell'$，$\hat{\theta}_M = \hat{\theta}_{MLE}$），它的效率才能与MLE相匹敌。