8.3 Limiting distribution of MLE

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta),$$

$$Z = \frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i, \theta)}{\partial \theta}.$$

$$\frac{\partial \ln L(\hat{\theta})}{\partial \theta} = 0.$$

$\Longrightarrow$ MLE.

$$
\begin{aligned}
0 &= \frac{\partial \ln L(\hat{\theta})}{\partial \theta} \\
&\approx \frac{\partial \ln L(\theta)}{\partial \theta} + (\hat{\theta} - \theta)\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}
\end{aligned}
$$

with $\hat{\theta} \sim \theta$.

$$\hat{\theta} - \theta \approx \frac{\partial \ln L(\theta)/\partial\theta}{-\frac{\partial^2 \ln L(\theta)}{\partial\theta^2}}$$

$$\sim \frac{\sum_{i=1}^{n} \frac{\partial \ln f(x_i,\theta)}{\partial\theta}}{nI(\theta)}.$$

$$\sqrt{n}(\hat{\theta} - \theta)$$

$$\sim \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \ln f(x_i, \theta)}{\partial \theta} / I(\theta)$$

$$\sim N(0, \frac{I(\theta)}{I^2(\theta)}) = N(0, \frac{1}{I(\theta)}).$$

$$\hat{\theta} - \theta \sim N(0, 1/(nI(\theta)))$$

$$\mathrm{eff}(\hat{\theta}) \to 1,$$

$\hat{\theta}$ is asymptotically efficient and

$$E(\hat{\theta} - \theta) = b_n(\theta) \to 0 \quad (n \to \infty)$$

$$Var(\hat{\theta}) \propto \frac{1}{nI(\theta)} \to 0 \quad (n \to \infty).$$

$$E(\hat{\theta}-\theta)^2 = b_n^2(\theta) + Var(\hat{\theta}) \to 0$$

as $n \to \infty$.

$$\Longrightarrow P\left(|\hat{\theta}-\theta| \geq \varepsilon\right) \leq \frac{E|\hat{\theta}-\theta|^2}{\varepsilon^2}$$

$$\to 0.$$

$\Longrightarrow \hat{\theta}$ is consistent.

- Cauchy case:

$$0 = \frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{2(x_i - \theta)}{1 + (x_i - \theta)^2}.$$

Let

$$w_{i1} = \frac{2}{1 + (x_i - \hat{\theta}_0)^2}.$$

From

$$0 = \sum_{i=1}^{n} w_{i1}(x_i - \theta),$$

to get

$$\hat{\theta}_1 = \frac{\sum w_{i1} x_i}{\sum w_{i1}}.$$

Let

$$w_{i2} = \frac{2}{1 + (x_i - \hat{\theta}_1)^2}$$

to get

$$\hat{\theta}_2 = \frac{\sum w_{i2} x_i}{\sum w_{i2}},$$

$\cdots$.

- $X_1, \cdots, X_n \sim f(x;\theta_1,\theta_2)$ with $\theta = (\theta_1,\theta_2)'$.

$$I_n(\theta) = nE\left(\frac{\partial \ln f(X,\theta)}{\partial\theta}\right)\left(\frac{\partial \ln f(X,\theta)}{\partial\theta}\right)' = -nE\left(\frac{\partial^2 \ln f(X,\theta)}{\partial\theta\partial\theta'}\right).$$

8.4 Robust M-Estimation

- $X_1, \cdots, X_n$ iid. $\sim$ Cauchy$(\theta, 1)$;

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

$$\ln L(\theta) = -n \ln \pi - \sum_{i=1}^{n} \ln[1 + (x_i - \theta)^2].$$

$$\frac{d\ln L(\theta)}{d\theta} = \sum_{i=1}^{n} \frac{2(x_i - \theta)}{1 + (x_i - \theta)^2} = 0$$

Let

$$w(x - \hat{\theta}_0) = \frac{2}{1 + (x_i - \hat{\theta}_0)^2}$$

with $\hat{\theta}_0 = Med(x_i)$.

$\Longrightarrow$ $\hat{\theta}$ is the MLE of θ.

- $X_i, \cdots, X_n$ **iid.** $\sim f(x-\theta)$

$$\ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i - \theta)$$
$$= -\sum_{i=1}^{n} \rho(x_i - \theta)$$

where $\rho(x) = -\ln f(x)$.

and

$$\frac{d\ln L(\theta)}{d\theta} = \sum_{i=1}^{n} \rho'(x_i - \theta)$$

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

with $\rho'(x) = \psi(x)$.

- Cauchy：

$$\rho(x) = \ln \pi + \ln(1 + x^2)$$

$$\psi(x) = \frac{2x}{1+x^2}, \quad w(x) = \frac{\psi(x)}{x}.$$

- Normal:

$$\rho(x) = \frac{1}{2}\ln(2\pi) + \frac{x^2}{2},$$

with $\psi(x) = x, \quad w(x) = 1,$

$\Longrightarrow \hat{\theta} = \bar{x}.$

- **Laplace:**

$$\rho(x) = \ln 2 + |x|, \quad \psi(x) = \text{sign}(x),$$

with $w(x) = \frac{\text{sign}(x)}{x} = \frac{1}{|x|} \quad (x \neq 0).$

$\Longrightarrow \hat{\theta} = Med\{x_i\}$ since

$$\sum_{i=1}^{n} \text{sign}(x_i - \theta) \approx 0.$$

首都师范大学 数学科学学院 统计学系 Department of Statistics

- For the three distri. The MLE $\hat{\theta}$ for one wouldn't be a good estimator for another. For instance, $\bar{X}$.

- An estimator that is fairly good for a wide variety of distri. is called a robust estimator.

- The solution $\hat{\theta}$ from the equation:

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

is called robust M-estimators (thought of as MLE).

- Need to select function $\psi(\cdot)$ that is good (robust M-estimator) for some distri. class.

Home Page

Title Page

◀◀ ▶▶

◀ ▶

Go Back

Full Screen

Close

Quit

- Huber used $\psi(\cdot)$ (combining Normal and Laplace), i.e.

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

and $$w(x) = \begin{cases} 1, & |x| \le k \\ \frac{k}{|x|}, & |x| > k \end{cases}$$

Then

$$\rho_1(x) = \begin{cases} \frac{x^2}{2}, & |x| \le k \\ k|x| - \frac{k^2}{2}, & |x| > k \end{cases}$$

- If

$$f(x) = ce^{-\rho_1(x)}$$

$\Longrightarrow$ M-estimator is MLE.

- If sample changes as lX_i,

$$X_i \longrightarrow lX_i,$$
$$\text{Med}\{X_i\} \longrightarrow l\text{Med}\{X_i\}.$$

From

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

$\hat{\theta}(x_1, \cdots, x_n) \Rightarrow l\hat{\theta}(x_1, \cdots, x_n).$

首都师范大学 数学科学学院 统计学系 Department of Statistics

We need to solve the equation:

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - \theta}{d}\right) = 0,$$

where $d = \frac{\text{Med}\{|x_i - \text{Med}\{x_i\}|\}}{0.6745}$.

- If $X_1, \cdots, X_n$ iid.$\sim N(\theta, \sigma^2)$, then

$$E(d) \to \sigma \quad (n \to \infty)$$

and

$$d \to \sigma \ \ a.s.(n \to \infty).$$

• How to choose k?

Under the normal case, we want most x_i to satisfy

$$\frac{|x_i - \theta|}{d} \leq k,$$

Since for all x_i, then

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - \theta}{d}\right) = \sum \frac{x_i - \theta}{d} = 0,$$

$$\Rightarrow \hat{\theta} = \bar{X},$$

$d \longrightarrow \sigma$, we choose

$k = 1.5$ or 2.

- Solve equation. Let $\hat{\theta}_0 = \text{Med}\{x_i\}$, by Taylor expansion:

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - \hat{\theta}_0}{d}\right) + (\theta - \hat{\theta}_0) \sum_{i=1}^{n} \psi'\left(\frac{x_i - \hat{\theta}_0}{d}\left(-\frac{1}{d}\right)\right) \approx 0,$$

we have

$$\hat{\theta}_1 \approx \hat{\theta}_0 + \frac{d \sum_{i=1}^{n} \psi\left(\frac{x_i - \hat{\theta}_0}{d}\right)}{\sum_{i=1}^{n} \psi'\left(\frac{x_i - \hat{\theta}_0}{d}\right)}.$$

- 1-step M-esti., by iteration, 2-step M-esti., 3-step M-esti, $\cdots$, then get $\hat{\theta}$.

- For Huber's $\psi(\cdot)$, we compute $\sum_{i=1}^{n} \psi'\left(\frac{x_i-\hat{\theta}_0}{d}\right)$ easily.

- Asymptotic distri. (σ known). From

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - \theta}{d}\right) = 0,$$

we get

$$\hat{\theta} - \theta \approx \frac{\sigma \sum_{i=1}^{n} \psi\left(\frac{x_i - \theta}{\sigma}\right)}{\sum_{i=1}^{n} \psi'\left(\frac{x_i - \theta}{d}\right)}$$

- If $X - \theta$ is symmetric, $\psi(\cdot)$ is odd, then

$$E\left[\psi\left(\frac{X-\theta}{\sigma}\right)\right] = 0,$$

and

$$\begin{aligned} &Var\left[\psi\left(\frac{X-\theta}{\sigma}\right)\right] \\ = \ & E\left[\psi\left(\frac{x-\theta}{\sigma}\right)\right]^2. \end{aligned}$$

$\Longrightarrow$

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

$$\sim N\left(0, \frac{\sigma^2 E\psi^2\left(\frac{x-\theta}{\sigma}\right)}{\left(E\psi'\left(\frac{x-\theta}{\sigma}\right)\right)^2}\right)$$

$$= N(0, v^2).$$

where

$$v^2 = \frac{\sigma^2 E\psi^2\left(\frac{x-\theta}{\sigma}\right)}{\left[E\psi'\left(\frac{x-\theta}{\sigma}\right)\right]^2};$$

Take estimator:

$$\hat{v}^2 = \frac{d^2\frac{1}{n}\sum_{i=1}^{n}\psi^2\left(\frac{x_i-\hat{\theta}_k}{d}\right)}{\left[\frac{1}{n}\sum_{i=1}^{n}\psi'\left(\frac{x_i-\hat{\theta}_k}{d}\right)\right]^2}.$$

- $\hat{\theta}_k$ is called the k-step M-estimator of θ.

The 95% CI of θ is

$$\hat{\theta}_k \pm 1.96\hat{v}.$$