Asymptotic Evalution

- we will consider the asymptotic properties when the sample size $\to \infty$. The power of it is the simplification of calculation.

(1) Point Estimation:

① Consistency:

- **Def**: seq of estimators $\{W_n(\vec{X})\}$ is consistent of θ. if $W_n(\vec{X}) \longrightarrow \theta$. in pr.

Thm. If seq of estimator $\{W_n(\vec{X})\}$ for θ. satisfies:

$\lim_{n\to\infty} Var_\theta(W_n(\vec{X})) = \lim_{n\to\infty} Bias_\theta(W_n(\vec{X})) = 0$. Then it's consistent with θ.

Pf: By chebyshev: $P(|W_n-\theta|\geq\varepsilon) \leq \frac{E((W_n-\theta)^2)}{\varepsilon^2}$

Thm (Consistence of MLE)

$X_k \sim f(x|\theta)$. i.i.d. $1\leq k\leq n$. $L(\theta|\vec{x}) = \prod_1^n f(x_i|\theta)$.

If $\hat{\theta}$ is MLE. under regular condition on f. then for $\forall$ τ. conti. fun. $\forall \theta \in \Theta$.

$\tau(\hat{\theta}) \longrightarrow \tau(\theta)$ in pr.

Pf: By $\frac{\log L(\theta|\vec{x})}{n} \longrightarrow E_{\theta_0}(\log f(x|\theta))$. a.e.

note that $\frac{\partial}{\partial\theta}\frac{L(\theta|\vec{x})}{n}\Big|_{\theta=\hat{\theta}} = 0$. $\frac{\partial}{\partial\theta}E_{\theta_0}(\log f(x|\theta))\Big|_{\theta=\theta_0} = 0$

we may hope $\hat{\theta} \xrightarrow{P} \theta_0$. Then by invariant of MLE $\forall \tau(\cdot)$ conti. we have $\tau(\hat{\theta}) \longrightarrow \tau(\theta)$ in pr.

① Efficiency:

· Since different consistent estimator has different asymptotic accuracy, which is related with Var.
We can compare the efficiency of different CE by comparing the speed of convergence of Variance.

· <u>Def</u>: For estimator T_n of $\tau(\theta)$.

i) If $\exists \{k_n\} \subseteq \mathbb{R}$. $\lim_{n} k_n Var(T_n) = \tau^2$. τ^2 is limiting Var.

ii) If $\exists \{k_n\} \subseteq \mathbb{R}$. $k_n(T_n - \tau(\theta)) \xrightarrow{d} N(0, \sigma^2)$. σ^2 is called asymptotic variance.

<u>Remark</u>: The limiting variance is different from asymptotic variance. Since they're different mode of convergence. So they may be not equal.

· <u>Def</u>: W_n is asymptotically efficient for $\tau(\theta)$ if: $W_n \in C_\tau^{\theta}$.

$k_n(W_n - \tau(\theta)) \to N(0, V(\theta))$, $V(\theta) = \dfrac{\tau'(\theta)^2}{E_\theta((\frac{\partial}{\partial\theta}\log f(\vec{X}|\theta))^2)}$

i.e. the Cramér-Rao Lower Bound.

<u>Thm.</u> (Asymptotic efficiency of MLEs)

$X_k \sim f(x|\theta)$, i.i.d. $1 \le k \le n$. $\hat{\theta}$ is MLE of θ. Under regular condition on $f(x|\theta)$. Then, for $\tau(\cdot)$ conti:

$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{d} N(0, V(\theta))$.

<u>Pf</u>: The ideal is from Taylor Series and

$\hat{\theta}$ is zero of $\frac{\partial}{\partial\theta}\log L(\theta|\vec{x}) \triangleq \frac{\partial}{\partial\theta} l(\theta|\vec{x})$

We only need to prove on $\hat{\theta}$.

Since $\tau(\hat{\theta})$ is MLE of $\tau(\theta)$. concerning $L(\tau(\theta)|x)$ which satisfies regular condition, too.

$l'(\theta|\vec{x}) = l'(\theta_0|x) + l''(\theta_0|x)(\theta-\theta_0) + o(1)$. (only need $\underline{(\theta-\theta_0)}$")

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

Note that $\frac{1}{n} l''(\theta_0|\vec{x}) \xrightarrow{as} E(\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)|_{\theta_0}) = -E((\frac{\partial}{\partial\theta}\log f)^2)$

$-\frac{1}{\sqrt{n}} l'(\theta_0|x) = -\sqrt{n}(\frac{l'(\theta_0|x)}{n}) \xrightarrow{d} -N(0, E((\frac{\partial}{\partial\theta}\log f(x|\theta)|_{\theta_0})^2))$

$\therefore \sqrt{n}(\hat{\theta}-\theta_0) \longrightarrow N(0, I^{-1}(\theta_0))$. $I(\theta_0) = E((\frac{\partial}{\partial\theta} f(x|\theta))^2)$

Remark: i) Generally. $l'(\tau(\theta)|x) = l'(\tau(\theta_0)|\vec{x}) + l''(\tau(\theta_0)|\vec{x})\tau'(\theta_0)(\theta-\theta_0)$

The similar argument can be applied!

ii) Sometimes we can't solve: $\frac{\partial}{\partial\theta}\log L(\theta|\vec{x}) = 0$.

for MLE $\hat{\theta}$. We can use the asymptotic dist as dist of $\hat{\theta}$. if n is large enough.

③ Calculations and Comparisons.

· Note that we obtain the asympototic variance of MLE. If we unknow θ. then we need to approximate variance: Since $Var(h(\hat{\theta})|\theta) \simeq \dfrac{[h'(\theta)]^2}{-E(\frac{\partial^2}{\partial\theta^2}\log L(\theta|\vec{x}))}$

then we substitute θ by $\hat{\theta}$:

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

Remark: It will fail when $h(\theta)$ isn't monotone. Since it causes sign change. then Var will be underestimation. (Actually. it bases on Cramér

Rao Lower Bound. So it have been underestimated)

· **Def**: If seq of estimators $\{W_n\}$, $\{V_n\}$ satisfies:

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

Then the asymptotic relative efficiency (ARE) is:

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

Remark: i) Asymptotic variance can be used to compare the efficiency. Since every asymptotic estimator will be variance 0 eventually

ii) The smaller the asymptotic variance is, the more efficient it is. So if $ARE(V_n, W_n) > 1$. Then $eff(V_n) > eff(W_n)$. That's why we reverse the position of σ_V^2 and σ_W^2 in the fraction.

iii) ARE can show the proportion of number of samples, if we want to have the same effect on estimation. Since $W_{n_1} \sim AN(0, \frac{\sigma_W^2}{n_1})$, $V_{n_2} \sim AN(0, \frac{\sigma_V^2}{n_2})$

$\therefore \frac{\sigma_W^2}{\sigma_V^2} = \frac{n_1}{n_2}$ if we need: $\frac{\sigma_W^2}{n_1} = \frac{\sigma_V^2}{n_2}$

iv) Since MLE is asymptotic efficient. Another estimator can't hope to beat it. Actually, there exists "Superefficiency":

$X_k \sim N(\theta, 1)$, i.i.d. $1 \le k \le n$, then $CRLB = \frac{1}{n} \triangleq v(\theta)$.

Let $d_n = \begin{cases} \bar{X}, & \text{if } |\bar{X}| > n^{-\frac{1}{4}} \\ a\bar{X}, & \text{if } |\bar{X}| \le n^{-\frac{1}{4}} \end{cases}$ (Choose $a > 0$. it's "Shrinkage")

Then $d_n \xrightarrow{P} \theta$ $(n \to \infty)$. Besides, $\sqrt{n}(d_n - \theta) \xrightarrow{d} N(0, d(\theta))$

$d(\theta) = 1$, if $\theta \neq 0$. $d(\theta) = a^2$ if $\theta = 0$.

Pf: Note that $\bar{X} \xrightarrow{P} \theta$. when $\theta \neq 0$. $P(|\bar{X}| > n^{-\frac{1}{4}}) \to 1 \ (n\to\infty)$

$\therefore d_n \xrightarrow{P} \theta$. $E(d_n) = P(|\bar{X}| > n^{-\frac{1}{4}})\bar{X} + P(|\bar{X}| \le n^{-\frac{1}{4}}) a\bar{X} \to \theta$.

$Var(\sqrt{n}(d_n - \theta)) = n Var(d_n) = n[E(d_n^2) - E(d_n)^2]$

$= n[P(|\bar{X}| > n^{-\frac{1}{4}}) E(\bar{X}^2) + P(|\bar{X}| \le n^{-\frac{1}{4}}) a^2 E(\bar{X}^2) - E(d_n)^2] \to 1$

when $\theta = 0$. $P(|\bar{X}| \le n^{-\frac{1}{4}}) \to 1 \ (n\to\infty)$

$\therefore d_n \xrightarrow{P} a\theta = 0 = \theta$. $E(d_n) \xrightarrow{P} a\theta = 0 = \theta$

$Var(\sqrt{n}(d_n - \theta)) = n Var(d_n) = n E(d_n^2) \longrightarrow a^2 E(\bar{X}^2) = a^2$.

Choose $a^2 < 1$. then $ARE(d_n, MLE) = \frac{1}{a^2} > 1$.

④ Boostrap Standard Error:

- Suppose we resample B times. $\hat{\theta}_i^*$ is the estimator from the i^{th} resample of common size n. from n original samples $\{X_k\}_1^n$. $\bar{\hat{\theta}}^* = \sum_1^B \frac{\hat{\theta}_i^*}{B}$. Then: (Nonparametric case)

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

Note that $Var_B^* = \frac{\sum_1^B Var_{Bi}^*}{B-1} \longrightarrow Var^*(\hat{\theta}) \ (B\to\infty)$.

$Var^*(\hat{\theta}) = \frac{1}{n^n - 1} \sum_1^{n^n} (\theta_i^* - \frac{\sum_1^{n^n} \theta_i^*}{n^n}) \longrightarrow Var(\hat{\theta})$. $(n\to\infty)$, by LLN.

($\theta_i^* = \hat{\theta}^*(X_{i_1}, X_{i_2} \cdots X_{i_n})$. $\{i_k\} = \{k\}_1^n$. from n original samples $\{X_k\}_1^n$

So there're $n \cdot n \cdots n = n^n$ different kinds for $\theta_i^* \neq \theta_j^*$. $i \neq j$)

Remark: i) The boostrap open has second-order accuracy.

ii) For parametric case. by estimate $\hat{\theta}_0$ from original samples. Suppose $X_k^* \sim f(x|\hat{\theta}_0)$. i.i.d. $1 \le k \le n$. use the estimator in ③!

(2) Robustness:

· We have evaluated the performance of estimators when the underlying model is correct.

If the model has some error. We will give up some optimality for exchange with "robustness".

Interpretion of robustness:

i) have a good efficiency under assumed model.

ii) small deviation from model should impair the performance slightly.

iii) larger deviation will not cause catastrophy.

① Median and Mean:

· Def: $\{X_{(k)}\}_1^n$ is order statistic. T_n is statistic based on it. T_n has break number "b". If:

$$\lim_{\substack{X_{([1-b]n)} \\ \to \infty}} T_n < \infty . \quad \lim_{\substack{X_{([1-b-\varepsilon]n)} \\ \to \infty}} T_n = \infty . \ \forall \varepsilon > 0 . \ (0 < b < 1)$$

Remark: i) Mean has break number "0". And Median has break number "0.5".

ii) In comparison of efficiency:

ARE(Median, Mean) ↑ if tail of dist ↑

② Criteria of estimator:

· Many of estimators are results from maximize or minimize some criteria, which have good properties of optimality. ($\sum(X_i-a)^2 \to \bar{X}$, $\sum|X_i-a| \to X_{\frac{1}{2}}$)

Next, we will introduce a criteria whose minimum will result in an estimator with good robustness properties.

i) <u>Huber's estimator</u>:

$\hat{a}$ is the desired estimator which minimizes:

$$\sum_{i=1}^{n} \rho(X_i - a), \quad \rho(x) = \begin{cases} \frac{1}{2}x^2, & |x| \le k \\ k|x| - \frac{1}{2}k^2, & |x| \ge k \end{cases} \quad \text{(it's conti)}$$

<u>Remark</u>: It makes a compromise between mean and median. Note that the mean criteria is square ($\bar{X}$ minimize $\sum(X_i-a)^2$). The median criteria is absolute value ($X_{\frac{1}{2}}$ minimize $\sum|X_i-a_i|$)

$\Rightarrow$ Since square has too much weight on tail. So we replace absolute value with it! (Break number states mean has more sensitivity!)

"k" is the turning number determing the estimator generated from the criteria act more like mean or median.

ii) <u>M-estimator</u>:

· For general function ρ, we call the estimator minimize $\sum_{1}^{n} \rho(X_i - \theta) = f(\theta)$ M-estimator.

<u>e.g.</u> If $\rho = -\ln(\theta|X)$, then it's maximal-likelihood-type.

Let $\psi = \ell'$. Note that the estimator is the zero of $\sum_{i=1}^{n} \psi(x_i-\theta)$ i.e. the solution of $\sum_i^n \psi(x_i-\theta)=0$. Denote $\hat{\theta}_m$

Note that by Taylor expansion:

$$\sum \psi(x_i-\theta) = \sum \psi(x_i-\theta_0) + \sum \psi'(x_i-\theta_0)(\theta-\theta_0) + o(1).$$

Let $\theta = \hat{\theta}_m$ $\therefore$ $0 = \sum \psi(x_i-\theta_0) + \sum \psi'(x_i-\theta_0)(\hat{\theta}_m-\theta_0) + o(1)$

$$\sqrt{n}(\hat{\theta}_m-\theta_0) = \frac{-\frac{1}{\sqrt{n}}\sum \psi(x_i-\theta_0)}{\frac{1}{n}\sum \psi'(x_i-\theta_0)} + o(1)$$

$$\frac{1}{n}\sum \psi'(x_i-\theta_0) \xrightarrow{P} E_{\theta_0}(\psi'(X-\theta_0)),\quad -\frac{1}{\sqrt{n}}\sum \psi(x_i-\theta_0) \xrightarrow{d} N(0, E_{\theta_0}(\psi^2(X-\theta_0)))$$

$$\therefore \sqrt{n}(\hat{\theta}_m-\theta_0) \xrightarrow{d} N(0, E_{\theta_0}(\psi^2(X-\theta_0))/E^2_{\theta_0}(\psi'(X-\theta_0)))$$

e.g. Asymptotic variance of Huber's estimator is:

$$\int_{-k}^{k} x^2 f(x)\,dx + 2k^2\int_k^\infty f(x)\,dx \Big/ [P_{\theta_0}(|X|\le k)]^2 \text{ when } X \sim f(x-\theta).$$

Remark: Under regular condition:

$$E_{\theta_0}(\psi'(X-\theta_0)) = \int \psi'(x-\theta_0) f(x-\theta_0)\,dx = -\int [\tfrac{\partial}{\partial\theta_0}\psi(x-\theta_0)] f(x-\theta_0)\,dx$$

$$\because \frac{\partial}{\partial\theta_0}\int \psi(x-\theta_0) f(x-\theta_0) = 0 \quad \therefore -\int \frac{\partial}{\partial\theta_0}[\psi(x-\theta_0)] f(x-\theta_0) = \int \psi(x-\theta_0)\frac{\partial}{\partial\theta} f(x-\theta_0)\,dx$$

$$\text{i.e. } E_{\theta_0}(\psi'(X-\theta_0)) = E_{\theta_0}(\psi(X-\theta_0)\tfrac{\partial}{\partial\theta}\log f(X-\theta_0))$$

(e.g. if $\psi = \frac{\partial}{\partial\theta}\log f(X-\theta_0)$, then it's familiar)

$$\Rightarrow ARE(\hat{\theta}_m, \hat{\theta}) = \frac{[E_{\theta_0}(\psi(X-\theta_0)\,\ell'(\theta|X))]^2}{E^2_{\theta_0}(\psi(X-\theta_0))\,E^2_{\theta_0}(\ell'(\theta|X))} \le 1.$$

where $\hat{\theta}$ is from MLE type

$\therefore$ M-estimator is alway less efficient than MLE. But it's more robust!

(3) Hypothesis Testing:

① Asymptotic Distributions of LRTs:

· In LRTs. If we can't explicitly write out $\lambda(\vec{x})$. then we can consider a asymptotic answer.

Thm: For $H_0: \theta = \theta_0$ v.s. $H_1 = \theta \neq \theta_0$. $X_k \sim f(x|\theta)$. i.i.d $1 \le k \le n$. satisfies regular condition. Then under H_0:

$-2\log\lambda(\vec{x}) \longrightarrow \chi_1^2$ in dist. $(n \to \infty)$

pf. Denote $\hat{\theta}$ is MLE of θ. $l(\theta|\vec{x}) = \log L(\theta|\vec{x})$

$$-2\log\lambda(\vec{x}) = -2l(\theta_0|\vec{x}) + 2l(\hat{\theta}|\vec{x}) \quad \cdots ①$$

$$l(\theta|x) = l(\hat{\theta}|x) + l'(\hat{\theta}|x)(\theta - \hat{\theta}) + \frac{l''(\hat{\theta}|\vec{x})}{2}(\theta - \hat{\theta})^2 + o(1)$$

$$= l(\hat{\theta}|X) + \frac{l''(\hat{\theta}|\vec{x})}{2}(\theta - \hat{\theta})^2. \quad (\hat{\theta} \text{ is MLE})$$

Let $\theta = \theta_0$. From ①:

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

$$\frac{-l''(\hat{\theta}|\vec{x})}{n} \xrightarrow{P} E_{\theta_0}(-l''(\hat{\theta}|\vec{x})) = E_{\theta_0}((l'(\hat{\theta}|\vec{x}))^2) = I(\theta_0)$$

$$\sqrt{n}(\theta_0 - \hat{\theta}) \xrightarrow{d} N(0, 1/I(\theta_0))$$

$$\therefore -2\log\lambda(\vec{x}) \xrightarrow{d} Z^2 = \chi_1^2. \quad (n \to \infty)$$

Extension: $X_k \sim f(x|\theta)$. i.i.d. $1 \le k \le n$. under regular condition

$H_0: \vec{\theta} = (\theta_1, \cdots \theta_n) \in \Theta_0$ v.s. $H_1: \vec{\theta} \in \Theta_0^c$. Then.

If $\vec{\theta} \in \Theta_0$. we have: $-2\log\lambda(\vec{x}) \xrightarrow{d} \chi_k^2$. where

$k = \dim(\Theta_0 \cup \Theta_0^c) - \dim \Theta_0$.

② Other Large Sample Test:

- For other test, if $W_n = W(X_1 \cdots X_n)$, the estimator of θ. Denote σ_n^2 is the variance of W_n, then:

$$\frac{W_n - \theta}{\sigma_n} \xrightarrow{d} N(0,1).$$

In some case, σ_n contains unknown parameters, then we will replase S_n with σ_n, where $\frac{S_n}{\sigma_n} \to 1$.
(usually retain the form, but replace the unknown para γ by its estimator $\hat{\gamma}$. e.g. $Var = \frac{\sqrt{p(1-p)}}{n} \to \frac{\sqrt{\hat{p}(1-\hat{p})}}{n}$)

Then we call $Z_n = \frac{W_n - \theta_0}{S_n}$, wald test.

Remark: generalized Wald statistic:

$$Z_{GW} = \sqrt{n}\,\frac{\hat{\theta}_M - \theta_0}{\sqrt{\widehat{Var}_{\theta_0}(\hat{\theta}_M)}}, \quad \hat{\theta}_M \text{ is a M-estimator}$$

$\widehat{Var}_{\theta_0}(\hat{\theta}_M)$ can be any consistent estimator.

Score Test:

$$Z_S = \frac{S(\theta_0)}{\sqrt{I_n(\theta_0)}}, \text{ where } S_n(\theta) = \frac{\partial}{\partial\theta}\log L(\theta|\vec{X}), \ \vec{X} = (X_1 \cdots X_n)$$

we know $E_\theta(S_n(\theta)) = 0$. $Var_\theta(S_n(\theta)) = I_n(\theta)$

$\therefore Z_S \xrightarrow{d} N(0,1)$ under $H_0: \theta = \theta_0$.

Generalized: $Z_{GS} = \sqrt{n}\,\frac{\hat{\theta}_M - \theta_0}{\sqrt{Var_{\theta_0}(\hat{\theta}_M)}}$. $\hat{\theta}_M$ is M-estimator.

(4) Interval Estimation:

· Now we explore approximate and asymptotic form of confidence interval.

① Approximate maximal likelihood interval:

i) Note that: $\frac{h(\hat\theta) - h(\theta)}{\sqrt{\hat{Var}(h(\hat\theta)|\theta)}} \to N(0,1)$.

where $\hat\theta$ is MLE. $\hat{Var}(h(\hat\theta)|\theta) = \frac{h'(\theta)^2}{-\frac{\partial^2}{\partial\theta^2}\log L(\theta|\vec{x})}\Big|_{\theta=\hat\theta}$.

Then we obtain approx. "$1-\alpha$" CI:

$h(\theta) \in [h(\hat\theta) - Z_{\frac{\alpha}{2}}\sqrt{\hat{Var}(h(\hat\theta)|\theta)}, h(\hat\theta) + Z_{\frac{\alpha}{2}}\sqrt{\hat{Var}(h(\hat\theta)|\theta)}]$

Remark: We may not use $Var(h(\hat\theta)|\theta)$ to replace $\hat{Var}$ for accuracy. Since it may need to solve a complicated equation for the interval!

ii) Note that score statistic is also applicable. It will give a better interval with optimal properties: $Q(\vec{X}|\theta) = \frac{\frac{\partial}{\partial\theta} l(\theta|\vec{x})}{E(-l''(\theta|\vec{x}))} \to N(0,1)$

Remark: It provide the shortest interval in a certain class. But it will be complicated!

iii) By LRTs : use $-2\log\lambda(\vec{x}) \to \chi_1^2$.

we can also solve a CI !

② Other Large Sample Intervals:

· Consider the form (Wald-type):

$$\frac{W-\theta}{V} \to N(0,1). \quad W, V \text{ are statistic. as } n \to \infty.$$

Remark: Sometimes we will replace V with some known parameters for reducing variability.