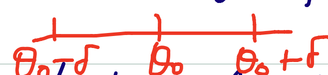
STA 3000 Oct 18/13

Maximum likelihood estimators – summary

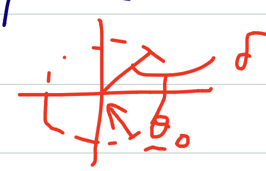
1. Under weak conditions on $f(y;\theta)$, but strong conditions on Θ (finite),

   $\hat{\theta} \xrightarrow{p} \theta_0$ using Jensen's & WLLN

2. Under more smoothness conditions on $f(y;\theta)$, and $\theta_0 \in \omega \subset \Theta$, $\exists$ local max of $\ell(\theta;y)$ in ω, & the one closest to $\theta_0 \xrightarrow{p} \theta_0$



3. This local max is a solution of the score equation $\ell'(\theta;y) = 0$.



   Pf: use TPE ; see also SM p 123

   $\underset{\sim}{\theta}$ $\theta \in \mathbb{R}$

4. Under more smoothness conditions on $f(y;\theta)$, including bdd 3^{rd} derivs,

$$\hat{\theta} - \theta_0 = i^{-1}(\theta_0)\, U(\theta_0) + o_p(1)$$

TPE Theorem 6.3.7

$Y_1, \dots, Y_n$ satisfy A0–A3 & a.a. y, $\frac{\partial}{\partial\theta} f(y;\theta)$ exists in ω. Then ω p 1 as $n \to \infty$, $\frac{\partial}{\partial\theta} \ell(\theta; \underline{y}) = 0$

has a sol$^{\underline{n}}$ $\hat\theta_n(\underline{y})$ s.t. $\hat\theta_n \xrightarrow{p} \theta_0$.

Pf. $\exists$ a s.t. $\theta_0 \pm a \in \omega$. Define

$$S_n = \{ \underline{y} : \ell(\theta_0; \underline{y}) > \ell(\theta_0 - a; \underline{y}) \text{ and } \ell(\theta_0; \underline{y}) > \ell(\theta_0 + a; \underline{y}) \}$$

use Wald pf. to show $P_{\theta_0}(S_n) \to 1$.

$\therefore$ for any $\underline{y} \in S_n$, $\exists$ $\theta_0 - a < \hat\theta_n < \theta_0 + a$ where $\ell(\cdot)$ has a local max., therefore $\ell'(\hat\theta_n) = 0$. Hence for any a suff'tly small, $\exists$ $\hat\theta_n = \hat\theta_n(a)$ s.t $P_{\theta_0}\{|\hat\theta_n(a) - \theta_0| < a\} \to 1$.
Let θ_n^* be the root closest to θ_0 ($\exists$ by continuity of $\ell(\theta)$).
Then $P_{\theta_0}\{|\theta_n^* - \theta_0| < a\} \to 1$.

A0–A3: - Dist$^{\underline{n}}$s P_θ are distinct

$f(y;\theta) = f(y;\theta')$ iff $\theta = \theta'$

- Common ort

- $Y_1, \ldots, Y_n$ iid $f(y; \theta_0)$
- Θ contains open set ω & θ_0 is interior pt. of ω

Vector: $\exists\ \omega \subset \Theta$ (open), w. $\underline{\theta}_0 \in \omega$

Define Q_a sphere centered at $\underline{\theta}_0$, radius a. Want to show

wp $\to 1$, $\ell(\underline{\theta}) < \ell(\underline{\theta}_0)$ $\forall$ θ on surface of Q_a.

$$\frac{1}{n}\ell(\theta) - \frac{1}{n}\ell(\theta_0) = \frac{1}{n}\sum_j (\theta_j - \theta_j^0)\,\ell'(\theta^0; \underline{y}) \qquad S_1$$

$$+ \frac{1}{2n}\sum_{jk} (\theta_j - \theta_j^0)(\theta_k - \theta_k^0)\,\ell_{jk}(\theta_j^0, \underline{y}) \qquad S_2$$

$$+ \frac{1}{6n}\sum_{jkl} (\theta_j - \theta_j^0)(\theta_k - \theta_k^0)(\theta_l - \theta_l^0)\,\ell_{jkl}(\theta^*; \underline{y}) \qquad S_3$$

$$\frac{1}{n}\ell'(\theta; \underline{y}) \xrightarrow{p} 0 \quad ; \quad \frac{1}{n}\ell''(\theta^0; \underline{y}) \xrightarrow{p} -i_1(\underline{\theta}_0)$$

etc.

Check: $\ell'(\hat{\theta}) = 0 = \ell'(\theta) + (\hat{\theta} - \theta)\ell''(\theta) + R_n$

$$\Rightarrow \quad \frac{-\ell'(\theta) - R_n}{\ell''(\theta)} = (\hat{\theta} - \theta)$$

$$\Rightarrow \quad \frac{+\boxed{\frac{1}{\sqrt{n}}}\left(\boxed{\ell'(\theta)} + R_n\right)}{-\frac{1}{n}\ell''(\theta)} = \sqrt{n}\,(\hat{\theta} - \theta)$$

$$\sqrt{n}\,(\hat{\theta} - \theta) = \frac{\frac{1}{\sqrt{n}}\{U(\theta) + R_n\}}{i_1(\theta)} \cdot \frac{i_1(\theta)}{-\frac{1}{n}\ell''(\theta)}$$

$$= \left\{\frac{\frac{1}{\sqrt{n}}U(\theta)}{i_1(\theta)} + \frac{R_n}{\sqrt{n}\,i_1(\theta)}\right\}(1 + o_p(1))$$

$$= \frac{\frac{1}{\sqrt{n}}U(\theta)}{i_1(\theta)} + \frac{R_n}{\sqrt{n}\,i_1(\theta)} + \{\ \}\,o_p(1)$$

$$\frac{1}{\sqrt{n}} R_n = \frac{1}{2}(\hat{\theta} - \theta)^2 \ell'''(\theta_n^*)\frac{1}{\sqrt{n}}$$

$$= \frac{1}{2} n(\hat{\theta} - \theta)^2 \cdot \frac{\frac{1}{n}\ell'''(\theta_n^*)}{i_3(\theta)} \cdot i_3(\theta)\frac{1}{\sqrt{n}}$$

$$= O_p(1)\,(1 + o_p(1))\,O(1)\frac{1}{\sqrt{n}}$$

ntbc

$$= O_p\left(\frac{1}{\sqrt{n}}\right) = o_p(1)$$

$$\underline{\sqrt{n}(\hat{\theta}-\theta)} = \frac{\boxed{\frac{1}{\sqrt{n}}U(\theta)}}{i_1(\theta)} + o_p(1) + \{\ \}o_p(1)$$

CLT

$$O_p(1) + o_p(1)$$

$$\hat{\theta}-\theta = \frac{U(\theta)}{n i_1(\theta)} + o_p(1)$$

$$= \frac{U(\theta)}{i_+(\theta)} + o_p(1)$$

Vector version $\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta} = i_+^{-1}(\underset{\sim}{\theta})\underset{\sim}{U}(\underset{\sim}{\theta}) + o_p(1)$

$$\ell'(\hat{\theta}) = \underset{\sim}{0} = \ell'(\underset{\sim}{\theta}) + \ell''(\theta)(\hat{\theta}-\theta) + R_n$$

$p \times 1$ $\quad p \times p$ $\quad p \times 1$

$$\underbrace{i_+^{-1}(\theta)\{-\ell''(\theta)\}}_{1+o_p(1)}(\hat{\theta}-\theta) \doteq i_+^{-1}(\theta)U(\theta)$$

$$\frac{1}{6}\left[\sum_{jkl}(\hat{\theta}_i-\theta_i)(\hat{\theta}_j-\theta_j)(\hat{\theta}_k-\theta_k)\times \frac{\partial^3\ell(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right]$$

$$R_n : (\hat{\theta}-\theta)^T \ell'''(\theta_n^*)(\hat{\theta}-\theta)$$

$1 \times p$ $\quad p \times p \times p$ $\quad p \times 1$

$(1 \times p \times p)\cdot(p \times 1) \rightarrow p \times p \times p \times 1 = p \times 1$

Vector version

$$\ell(\underset{\sim}{\theta}) = \ell(\hat{\underset{\sim}{\theta}}) - (\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}})\ell'(\hat{\theta})$$

$$- \frac{1}{2}(\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}})^T \ell''(\hat{\underset{\sim}{\theta}})(\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}})$$

$$- \frac{1}{6}\sum_{r,s,t}(\theta_r - \hat{\theta}_r)(\theta_s - \hat{\theta}_s)(\theta_t - \hat{\theta}_t)\frac{\partial^3 \ell(\theta)}{\partial\theta_r\partial\theta_s\partial\theta_t}\bigg|_{\theta=\hat{\theta}}$$

$$2\{\ell(\hat{\theta}) - \ell(\theta)\} = (\hat{\theta} - \theta)^T i_+(\theta)(\hat{\theta} - \theta) + o_p(1)$$

$$\xrightarrow{d} \chi^2_q \qquad\qquad \xrightarrow{d} \chi^2_q$$

$$X_n \xrightarrow{d} X \qquad a_n \xrightarrow{p} 0 \Rightarrow X_n + a_n \xrightarrow{d} X$$

etc.

Approximations   scalar

$$\sqrt{n}\,(\hat{\theta}-\theta)\, i_1^{1/2}(\theta) \xrightarrow{d} N(0,1)$$

$$(\hat{\theta}-\theta)\, i^{1/2}(\theta) \mathrel{\dot\sim} N(0,1)$$

To this order of approx $\dot{=}$,

$$(\hat{\theta}-\theta)\, i^{1/2}(\hat{\theta}),\quad (\hat{\theta}-\theta)\, j^{1/2}(\hat{\theta}),\quad (\hat{\theta}-\theta)\, j^{1/2}(\theta)$$

also $\dot\sim N(0,1)$

Also $u(\theta)/i^{1/2}(\theta) \mathrel{\dot\sim} N(0,1)$ and

$$u(\theta)/j^{1/2}(\hat{\theta}),\ \text{etc.} \mathrel{\dot\sim} N(0,1)$$

And

$$r(\theta) = \pm\sqrt{2\{\ell(\hat{\theta})-\ell(\theta)\}} \mathrel{\dot\sim} N(0,1)$$

$$Y \sim \text{Bin}(n, \theta) \qquad \binom{n}{y} \theta^{y} (1-\theta)^{n-y}$$

$$\hat{\theta} = \frac{Y}{n}$$

$$\ell(\theta) = y \ln \theta + (n-y) \ln (1-\theta)$$

$$\ell'(\theta) = \frac{y}{\theta} - \frac{n-y}{1-\theta}$$

$$\ell''(\theta) = -\frac{y}{\theta^{2}} - \frac{(n-y)}{(1-\theta)^{2}}$$

$$\ell'(\hat{\theta}) = 0 \qquad y(1-\hat{\theta}) - (n-y)\hat{\theta} = 0$$

$$y - y\hat{\theta} - n\hat{\theta} + y\hat{\theta} = 0 \qquad \hat{\theta} = \frac{y}{n}$$

$$i(\theta) = E\{-\ell''(\theta)\} = \frac{n\cancel{\theta}}{\theta^{\cancel{2}}} + \frac{n\cancel{(1-\theta)}}{(1-\theta)^{\cancel{2}}}$$

$$= n\left(\frac{1}{\theta(1-\theta)}\right)$$

$$(\hat{\theta}-\theta)\, i^{1/2}(\theta) \;\dot{\sim}\; N(0,1)$$

$$\left(\frac{y}{n}-\theta\right) \cdot \sqrt{n}\, \frac{1}{\sqrt{\theta(1-\theta)}} \;\dot{\sim}\; N(0,1) \text{ using } i(\theta)$$

$$(\hat{\theta}-\theta)\, j^{1/2}(\hat{\theta}) \;\dot{\sim}\; N(0,1)$$

$$\left(\frac{y}{n}-\theta\right) \cdot \sqrt{n}\, \frac{1}{\sqrt{\frac{y}{n}\left(1-\frac{y}{n}\right)}} \;\dot{\sim}\; N(0,1) \quad j(\hat{\theta})$$

$$CI_{2} \quad \frac{y}{n} = \hat{p} \pm 1.96 \frac{\hat{p}(1-\hat{p})}{\sqrt{n}}$$

CI_1 inverting $(*)$

Bickel & Doksum

$\hat{\theta} - \theta = i^{-1}(\theta) U(\theta) + o_p(1)$

Using vector version

$Y_1, \ldots, Y_n$ iid

$$(\underline{\hat{\theta}} - \underline{\theta}) \dot{\sim} N_2(0, i_+^{-1}(\theta))$$

$(\psi, \lambda) \in \mathbb{R} \times \mathbb{R}$

$$\begin{pmatrix} \hat{\psi} - \psi \\ \hat{\lambda} - \lambda \end{pmatrix} \dot{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} i^{\psi\psi}(\theta) & i^{\psi\lambda}(\theta) \\ i^{\lambda\psi}(\theta) & i^{\lambda\lambda}(\theta) \end{pmatrix}\right]$$

1) $\hat{\psi} - \psi \dot{\sim} N(0, i^{\psi\psi}(\theta))$

$\hat{\lambda} - \lambda \dot{\sim} N(0, i^{\lambda\lambda}(\theta))$

$\hat{\beta}_1$ s.e. $\hat{\beta}_1$ z

$\hat{\beta}_2$ s.e

$\vdots$ $\vdots$

$\hat{\beta}_p$ s.e.

$$(\hat{\psi} - \psi)\{i^{\psi\psi}(\hat{\theta})\}^{1/2} \dot{\sim} N(0,1)$$

$$\hat{\psi} \pm 1.96 \{i^{\psi\psi}(\hat{\theta})\}^{-1/2} \sim 95\% \text{ CI}$$

j $\hat{\theta}$

$$r_e = (\hat{\psi} - \psi)\{j^{\psi\psi}(\hat{\psi}, \hat{\lambda})\}^{1/2} \dot{\sim} N(0,1)$$

$$\begin{bmatrix} i_{\psi\psi}(\theta) & i_{\psi\lambda}(\theta) \\ i_{\lambda\psi}(\theta) & i_{\lambda\lambda}(\theta) \end{bmatrix}^{-1} = \frac{1}{i_{\psi\psi}\, i_{\lambda\lambda} - i_{\psi\lambda}^2} \begin{bmatrix} i_{\lambda\lambda} & -i_{\psi\lambda} \\ -i_{\lambda\psi} & i_{\psi\psi} \end{bmatrix}$$

$$i_{\psi\lambda} = i_{\lambda\psi}$$

$$E\, \frac{\partial^2 l}{\partial\lambda\,\partial\psi}$$

$$= \begin{bmatrix} \frac{i_{\lambda\lambda}}{\cdots} & — \\ — & — \end{bmatrix}$$

$$\begin{bmatrix} i^{\psi\psi} & i^{\psi\lambda} \\ i^{\lambda\psi} & i^{\lambda\lambda} \end{bmatrix} = \begin{bmatrix} \frac{1}{i_{\psi\psi} - i_{\psi\lambda}^2\, i_{\lambda\lambda}} & — \\ — & — \end{bmatrix}$$

In general

$$i^{\psi\psi} = i_{\psi\psi} - i_{\psi\lambda}\, i_{\lambda\lambda}^{-1}\, i_{\lambda\psi}$$

2) $\begin{pmatrix} U_\psi \\ U_\lambda \end{pmatrix} \dot{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} i_{\psi\psi} & i_{\psi\lambda} \\ i_{\lambda\psi} & i_{\lambda\lambda} \end{pmatrix} \right)$

$$U_\psi(\psi, \lambda)\{i_{\psi\psi}(\theta)\}^{-\frac{1}{2}} \xrightarrow{d} N(0,1)$$

$$r_u = U_\psi(\psi, \hat{\lambda}_\psi)\{j_{\psi\psi}(\hat{\psi}, \hat{\lambda})\}^{-\frac{1}{2}}$$

3) $w(\theta) = 2\{\ell(\hat{\underset{\sim}{\theta}}) - \ell(\theta)\} \xrightarrow{d} \chi^2_2$

$\{\underset{\sim}{\theta} : w(\theta) < c_\alpha\}$ 95% c. reg. for $\underset{\sim}{\theta}$

1)' $(\hat{\underline{\theta}} - \underline{\theta})^T i_+(\underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})$ also

$U(\underline{\theta})^T i_+^{-1}(\theta) U(\underline{\theta})$ als

Example $Y_1, \dots, Y_n$ iid $N(\mu, \sigma^2)$

$$f(\underset{\sim}{y}; \underset{\sim}{\theta}) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} e^{-\frac{1}{2\sigma^2}\sum(y_i - \mu)^2}$$

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum(y_i - \mu)^2$$

$$\ell_\mu = -\frac{1}{2\sigma^2} 2 \sum(y_i - \mu)(-1)$$

$$= \frac{n\bar{y} - n\mu}{\sigma^2} \qquad \hat{\mu} = \bar{y}$$

$$\ell_{\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(y_i - \mu)^2$$

$$0 = -\frac{n\hat{\sigma}^2}{2\hat{\sigma}^4} + \frac{1}{2\hat{\sigma}^4}\sum(y_i - \bar{y})^2$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum(y_i - \bar{y})^2$$

m.l.e.
not unbiased

$$\ell_{\mu\mu} = -\frac{n}{\sigma^2} \qquad \ell_{\mu\sigma^2} = -\frac{n(\bar{y}-\mu)}{\sigma^4}$$

$$\ell_{\sigma^2\sigma^2} = \frac{n}{2\sigma^4} - \frac{2}{2\sigma^6}\sum(y_i-\mu)^2$$

$$j(\hat{\underset{\sim}{\theta}}) = \begin{bmatrix} \frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{bmatrix} \qquad j(\underset{\sim}{\theta}) \overset{?}{=} \begin{bmatrix} \ \end{bmatrix}$$

$$i(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & n/2\sigma^4 \end{bmatrix}$$

$$\begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma}^2 - \sigma^2 \end{pmatrix} \overset{\cdot}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}\right)$$

$\bar{y} \overset{\cdot}{\sim} N(\mu, \frac{\sigma^2}{n})$ in fact exact

$\hat{\sigma}^2 \overset{\cdot}{\sim} N(\sigma^2, \frac{2\sigma^4}{n})$ not exact

(exact $\propto$ to χ^2_{n-1})

$$\underline{l_p(\mu)} = l(\mu, \hat{\sigma}^2_\mu)$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(y_i-\mu)^2$$

$$\boxed{\hat{\sigma}^2_\mu = \frac{\sum(y_i-\mu)^2}{n}} \quad \text{(ntbc)}$$

$\pm$

$$\sqrt{2\{l_p(\hat{\mu}) - l_p(\mu)\}} \sim N(0,1)$$

$$l(\mu, \hat{\sigma}^2_\mu) = -\frac{n}{2}\log\hat{\sigma}^2_\mu - \frac{1}{2\hat{\sigma}^2_\mu}\sum(y_i-\mu)^2$$

$$l_p(\mu) = -\frac{n}{2}\log\left(\frac{\sum(y_i-\mu)^2}{n}\right) - \frac{n}{2}$$

$$= +\frac{n}{2}\log n - \frac{n}{2}\log\sum(y_i-\mu)^2 - \frac{n}{2}$$

$$\hat{\mu} = \bar{y}$$

$$2\{\ell_p(\hat{\mu}) - \ell_p(\mu)\}$$

$$= 2\left\{\frac{n}{2}\log\sum(y_i - \mu)^2 - \frac{n}{2}\log\sum(y_i - \bar{y})^2\right\}$$

$$= n\log\hat{\sigma}_\mu^2 - n\log\hat{\sigma}^2$$

$$= n\log\left(\frac{\hat{\sigma}_\mu^2}{\hat{\sigma}^2}\right) = n\log\left(\frac{\hat{\sigma}_\mu}{\hat{\sigma}}\right)^2$$

$$= 2n\log(\hat{\sigma}_\mu / \hat{\sigma})$$

$$\pm\sqrt{2\{\ell_p(\hat{\mu}) - \ell_p(\mu)\}} = \pm\sqrt{2n\log(\hat{\sigma}_\mu / \hat{\sigma})}$$

$$\hat{\sigma}_\mu^2 = \frac{1}{n}\sum(y_i - \mu)^2$$

$$= \frac{1}{n}\left\{\sum(y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right\}$$

$$= \hat{\sigma}^2 + (\bar{y} - \mu)^2$$

$$\frac{\hat{\sigma}_\mu^2}{\hat{\sigma}^2} = 1 + \frac{(\bar{y}-\mu)^2}{\hat{\sigma}^2} = 1 + \frac{(\bar{y}-\mu)^2}{s^2/n} \cdot \left(\frac{s^2/n}{\hat{\sigma}^2}\right)$$

$$= 1 + T^2 \cdot \left(\frac{n-1}{n}\right) \; ? \; \left(\frac{n}{n-1}\right)$$

$$\uparrow$$