Lecture 24. MLE

Wednesday, December 6, 2017 8:52 AM

Bug in HW 8 Solutions for (d) corrected

$$f'(x) = \frac{f(x+2a_n) - f(x)}{2a_n} + o(a_n)$$

Numerical analysis

$$f'(x) = \frac{f(x+a_n) - f(x-a_n)}{2a_n} + o(a_n^2)$$

The asymptotic IN of Z-estimators framework of Lecture 23 does not work for the median b/c ψ' does not always exists (sign is not differentiable everywhere)

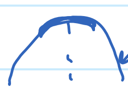
DF a function f is called Lipschitz if

$$|f(x_1) - f(x_2)| \le C \cdot \|x_1 - x_2\|$$

TH Relaxing differentiability requirements

]

- $m_\theta(x)$ is measurable as a function of x and differentiable at θ^*, P-a.s. wrt x
- $m_\theta(x)$ is Lipschitz wrt θ, i.e. $\exists\ C := \dot{m}(x)$
- $\dot{m}(x)$ is measurable and $P\dot{m}^2 < \infty$
- Pm_θ has a 2nd order Taylor expansion at θ^* with the 2nd derivative matrix V_{θ^*} and 1st derivative Pm'_{θ^*}



Pm_θ

θ^*

θ θ^*

θ^*

- $\mathbb{P}_n m_{\hat\theta_n} \geq \sup_\theta \mathbb{P}_n m_\theta - o_p\left(\frac{1}{n}\right)$, i.e.

  $\hat\theta_n$ is an approx. maximizer of $M_n(\theta) = \mathbb{P}_n m_\theta$

- $\hat\theta_n$ is consistent for θ^*

  $$\hat\theta_n \xrightarrow[p]{} \theta^*$$

Then $$\sqrt{n}\,(\hat\theta_n - \theta) = -V_{\theta^*}^{-1} \underbrace{\sqrt{n}\,\mathbb{P}_n m'_{\theta^*}} + o_p(1)$$

In particular, using CLT on

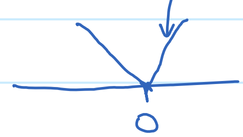
$$\sqrt{n}\left(\hat\theta_n - \theta^*\right) \rightsquigarrow \mathcal{N}\left(0,\ V_{\theta^*}^{-1} P\left(m'_{\theta^*} m'^{T}_{\theta^*}\right) V_{\theta^*}^{-1T}\right)$$

Proof uses empirical processes

$$(\mathbb{P}_n - P) f$$

Example: median

|x|- function



$$M_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} |X_i - \theta|$$

$$\Psi_n(\theta) = M_n'(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \operatorname{sign}(X_i - \theta)$$

$\exists$ pdf, f, exists at θ^* $\Rightarrow$ $\hat\theta_n = \arg\max M_n(\theta)$

$\sqrt{n}\left(\hat\theta_n - \theta^*\right)$ is asymptotically $\mathcal{N}\left(0, \frac{1}{(2f(\theta^*))^2}\right)$

Proof:

$$|m_{\theta_1}(x) - m_{\theta_2}(x)| = \big|\,|x-\theta_1| - |x-\theta_2|\,\big| \leq |\theta_1 - \theta_2| \Rightarrow$$

$$\Rightarrow \dot m_\theta(x) \equiv 1$$

$m'_{\theta^*}(x) = -\operatorname{sign}(x - \theta^*)$ exists a.e. except at $x = \theta^*$

$$M(\theta) = P m_n = \int |x - \theta|\, dF(x) =$$

$$M(\theta) = P m_\theta = \int |x-\theta| \, dF(x) =$$

$\uparrow$ CDF of X

integration by parts

$$= \int_{-\infty}^{\theta} (\theta - x)\, dF(x) + \int_{\theta+}^{\infty} (x-\theta)\, dF(x) =$$

$$= \theta F(\theta) - \theta[1-F(\theta)] - \lim_{a\to\infty} \left[xF(x)\Big|_{-a}^{\theta} - \int_{-a}^{\theta} F(x)dx \right] +$$

$$+ \left[xF(x)\Big|_{\theta+}^{a} - \int_{\theta+}^{a} F(x)dx \right] =$$

$$= -\theta + \lim_{a\to\infty} \left[\int_{-a}^{\theta} F(x)dx - \int_{\theta+}^{a} F(x)\,dx + 2aF(a) \right]$$

Differentiating under the limit, we have

Note: Differentiation is justified if derivatives converge uniformly wrt a. In our situation this is the case because derivatives do not depend on a.

$$\frac{d^2 P m_\theta}{d\theta^2}\Bigg|_{\theta^*} = 2\cdot f(\theta)\Big|_{\theta^*} = V_{\theta^*} \quad ; \quad \frac{dPm_\theta}{d\theta}\Bigg|_{\theta^*} = 2F(\theta)\Big|_{\theta^*} - 1 = 0$$

$\uparrow$ $F(\theta^*) = \frac{1}{2}$

$$P(\dot{m}_{\theta^*}^{\,2}) = P(1) = 1$$

$$\sqrt{n}\,(\hat{\theta}_n - \theta^*) \rightsquigarrow \mathbb{N}\left(0, \frac{1}{(2f(\theta^*))^2}\right)$$

MLE - framework

$$l_n = \sum_{i=1}^{n} \log f_\theta(X_i) \qquad \max_\theta l_n(\theta, \text{data}) \Rightarrow \hat{\theta}_n \text{ MLE}$$

$$M_n(\theta) = \mathbb{P}_n \log \frac{f_\theta}{f_{\theta^*}} = \frac{1}{n}\sum_{i=1}^{n} \Big[\log f_\theta(X_i) - \underbrace{\log f_{\theta^*}(X_i)}_{\text{doesn't depend on } \theta}\Big]$$

$$\text{LLN} \Rightarrow M_n(\theta) \to P \log \frac{f_\theta}{f_{\theta^*}} := M_\theta$$

$$\text{LLN} \Rightarrow M_n(\theta) \to P \log \frac{f_\theta}{f_{\theta^*}} := M_\theta$$

Jensen inequality

$$P \log \frac{f_\theta}{f_{\theta^*}} \leq \log P \frac{f_\theta}{f_{\theta^*}} = \log \int \frac{f_\theta}{f_{\theta^*}} \cdot f_{\theta^*} \, dx = 0$$

$P(\cdot)$ is taken over the X_is $\sim P_{\theta^*}$

$$\left.\begin{array}{l} \text{i.e. } M_\theta \leq 0 \\ \\ M_{\theta^*} = 0 \end{array}\right\} \; M_\theta \text{ is maximized by } \theta^*$$

Note: $-M_n(\theta)$ is called Kullback-Leibler divergence

Model identifiability: $f_\theta(\cdot) = f_{\theta^*}(\cdot) \Leftrightarrow \theta = \theta^*$

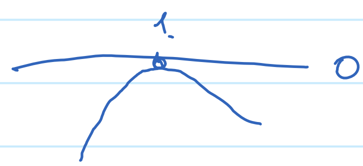
$\Downarrow$

θ^* as $\arg\max M_\theta$ is unique

Proof:

Consider $\varphi(x) = \log x - x + 1$

$\varphi'(x) = \frac{1}{x} - 1$



$\Rightarrow \log x \leq x - 1$ "=" iff $x = 1$

$x \leftarrow \sqrt{\frac{f_\theta}{f_{\theta^*}}}$

$$\log \frac{f_\theta}{f_{\theta^*}} \leq 2 \left(\sqrt{\frac{f_\theta}{f_{\theta^*}}} - 1 \right)$$

Now:

$$P \log \frac{f_\theta}{f_{\theta^*}} = M_\theta \leq 2 \int \left(\sqrt{\frac{f_\theta}{f_{\theta^*}}} - 1 \right) \cdot f_{\theta^*} \, dx =$$

$$= 2\int \sqrt{f_\theta \cdot f_{\theta^*}} - \underbrace{2}_{\int(\sqrt{f_\theta})^2 dx + \int(\sqrt{f_{\theta^*}})^2 dx} =$$

$$= -\int \left(\sqrt{f_\theta} - \sqrt{f_{\theta^*}}\right)^2 dx \leq 0 \Rightarrow$$

w/ it $=0$ iff. $f_\theta \equiv f_{\theta^*}$

b/c of identif. $\Longleftrightarrow$ $\theta = \theta^*$ $\square$

Note: $\frac{1}{2}\int(\sqrt{f_1} - \sqrt{f_2})^2 dx$ is Hellinger distance between f_1 and f_2