Section for Statistical Theory

TA: Cong Mu

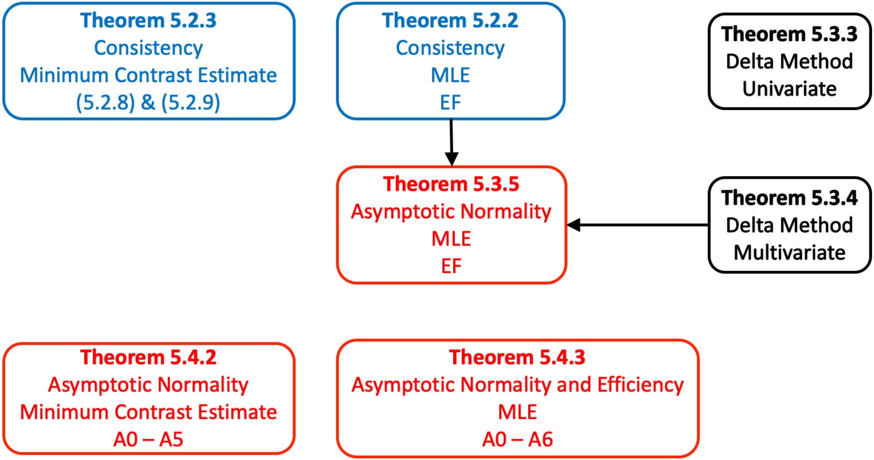
Office Hour: Wednesday 09:30AM - 11:30AM

November 19 & 20, 2020

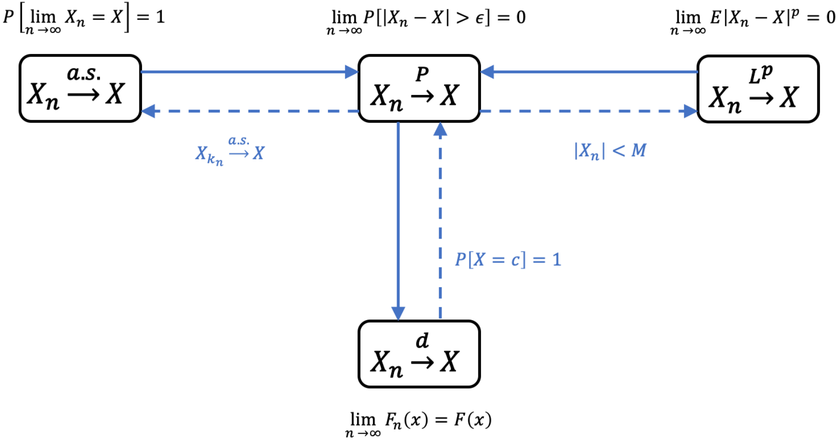
Overview

1. Review
   - Preliminaries
   - Consistency
   - Asymptotic Normality
2. Problems

Overview


Theorem 5.2.3
Consistency
Minimum Contrast Estimate
(5.2.8) & (5.2.9)
Theorem 5.2.2
Consistency
MLE
EF
Theorem 5.3.3
Delta Method
Univariate
Theorem 5.3.5
Asymptotic Normality
MLE
EF
Theorem 5.3.4
Delta Method
Multivariate
Theorem 5.4.2
Asymptotic Normality
Minimum Contrast Estimate
A0 – A5
Theorem 5.4.3
Asymptotic Normality and Efficiency
MLE
A0 – A6

Preliminaries - Convergence of Random Variables



Preliminaries - Convergence of Random Variables

Continuous Mapping Theorem

Let g be a continuous function, then

- $X_n \xrightarrow{d} X \quad \Longrightarrow \quad g(X_n) \xrightarrow{d} g(X)$
- $X_n \xrightarrow{P} X \quad \Longrightarrow \quad g(X_n) \xrightarrow{P} g(X)$
- $X_n \xrightarrow{a.s.} X \quad \Longrightarrow \quad g(X_n) \xrightarrow{a.s.} g(X)$

Slutsky's Theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$ where c is a constant and invertible, then

- $X_n + Y_n \xrightarrow{d} X + c$
- $X_n Y_n \xrightarrow{d} Xc$
- $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$

Consistency - MLE

Theorem 5.2.2. (page 303 in [B&D, 2015])

Suppose $\mathcal{P}$ is a canonical exponential family of rank d generated by $\mathbf{T}$. Let $\boldsymbol{\eta}, \mathcal{E}$ and $A(\cdot)$ correspond to $\mathcal{P}$. Suppose $\mathcal{E}$ is open and $X_1, \cdots, X_n$ are a sample from $P_{\boldsymbol{\eta}} \in \mathcal{P}$. Then

1. (i) $P_{\boldsymbol{\eta}}$ [The MLE $\widehat{\boldsymbol{\eta}}$ exists] $\to 1$.
2. (ii) $\widehat{\boldsymbol{\eta}}$ is consistent, i.e., $\widehat{\boldsymbol{\eta}} \xrightarrow{P} \boldsymbol{\eta}$.

Consistency - Minimum Contrast Estimates

Theorem 5.2.3. (page 304 in [B&D, 2015])

Let $X_1, \cdots, X_n$ be an i.i.d. sample from $P_{\boldsymbol{\theta}}$ where $\boldsymbol{\theta} \in \Theta \subset R^d$. Let $\widehat{\boldsymbol{\theta}}$ be a minimum contrast estimate that minimizes

$$\rho_n(\mathbf{X}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \rho(X_i, \boldsymbol{\theta}),$$

where $D(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0} \rho(X_1, \boldsymbol{\theta})$ is uniquely minimized at $\boldsymbol{\theta}_0$. If

$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \left[\rho(X_i, \boldsymbol{\theta}) - D(\boldsymbol{\theta}_0, \boldsymbol{\theta})\right] \right| \xrightarrow{P_{\boldsymbol{\theta}_0}} 0. \tag{5.2.8}$$

$$\inf_{|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \geqslant \epsilon} D(\boldsymbol{\theta}_0, \boldsymbol{\theta}) > D(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) \qquad \text{for every } \epsilon > 0. \tag{5.2.9}$$

Then $\widehat{\boldsymbol{\theta}}$ is consistent.

Asymptotic Normality - The Delta Method

Theorem 5.3.4. (page 320 in [B&D, 2015])

Suppose $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$ are i.i.d. d vectors with $E|\mathbf{Y}_1|^2 < \infty$, $E[\mathbf{Y}_1] = \mathbf{m}$, $\mathrm{Var}[\mathbf{Y}_1] = \boldsymbol{\Sigma}$ and $\mathbf{h} : \mathcal{O} \to R^p$ where $\mathcal{O} \subset R^d$ is open, $\mathbf{h} = (h_1, \cdots, h_p)$ and $\mathbf{h}$ has a total differential $\left[\mathbf{h}^{(1)}(\mathbf{m})\right]_{ij} = \frac{\partial h_i}{\partial x_j}(\mathbf{m})$. Then

$$\mathbf{h}\left(\bar{\mathbf{Y}}\right) = \mathbf{h}(\mathbf{m}) + \mathbf{h}^{(1)}(\mathbf{m})\left(\bar{\mathbf{Y}} - \mathbf{m}\right) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{5.3.23}$$

$$\sqrt{n}\left[\mathbf{h}\left(\bar{\mathbf{Y}}\right) - \mathbf{h}(\mathbf{m})\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{h}^{(1)}(\mathbf{m})\,\boldsymbol{\Sigma}\left[\mathbf{h}^{(1)}(\mathbf{m})\right]^{\top}\right). \tag{5.3.24}$$

Asymptotic Normality - MLE in EF

Theorem 5.3.5. (page 322-323 in [B&D, 2015])

Suppose $\mathcal{P}$ is a canonical exponential family of rank d generated by $\mathbf{T}$ with $\mathcal{E}$ open. Let $X_1, \cdots, X_n$ be a sample from $P_{\boldsymbol{\eta}} \in \mathcal{P}$ and $\widehat{\boldsymbol{\eta}}$ be defined as the MLE if exists and equal to $\mathbf{c}$ otherwise. Then

(i)

$$\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta} + \frac{1}{n} \sum_{i=1}^{n} \ddot{A}^{-1}(\boldsymbol{\eta}) \left[\mathbf{T}(X_i) - \dot{A}(\boldsymbol{\eta})\right] + o_{P_{\boldsymbol{\eta}}}\left(\frac{1}{\sqrt{n}}\right).$$

(ii)

$$\sqrt{n}\left(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}\right) \xrightarrow{\mathcal{L}_{\boldsymbol{\eta}}} \mathcal{N}_d\left(\mathbf{0}, \ddot{A}^{-1}(\boldsymbol{\eta})\right).$$

Asymptotic Normality - Minimum Contrast Estimates

Theorem 5.4.2. (page 328-329 in [B&D, 2015])

Under **A0-A5** (page 328 in [B&D, 2015]), we have

$$\bar{\theta}_n = \theta(P) + \frac{1}{n}\sum_{i=1}^{n} \widetilde{\psi}(X_i, \theta(P)) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{5.4.22}$$

$$\widetilde{\psi}(X_i, \theta(P)) = \psi(X_i, \theta(P)) \bigg/ \left[-E_P \frac{\partial \psi}{\partial \theta}(X_1, \theta(P))\right]. \tag{5.4.23}$$

Hence

$$\sqrt{n}\,[\bar{\theta}_n - \theta(P)] \xrightarrow{\mathcal{L}_P} \mathcal{N}\left(0, \sigma^2(\psi, P)\right),$$

$$\sigma^2(\psi, P) = \frac{E_P \psi^2(X_1, \theta(P))}{\left[E_P \frac{\partial \psi}{\partial \theta}(X_1, \theta(P))\right]^2}. \tag{5.4.24}$$

Asymptotic Normality - MLE

Theorem 5.4.3. (page 331 in [B&D, 2015])

If **A0-A6** (page 328 and 330 in [B&D, 2015]) apply to $\rho(x, \theta) = -\ell(x, \theta)$ and $P = P_\theta$, then the MLE $\widehat{\theta}_n$ satisfies

$$\widehat{\theta}_n = \theta + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{I(\theta)} \frac{\partial \ell}{\partial \theta}(X_1, \theta) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{5.4.33}$$

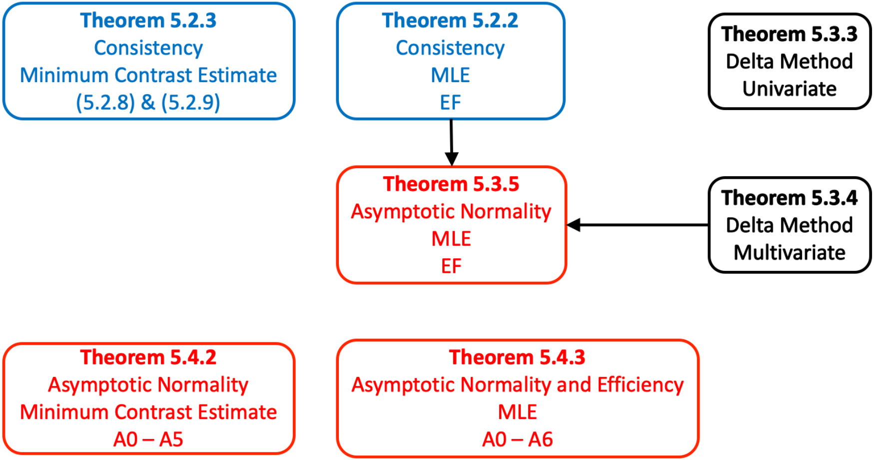
$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{L}_\theta} \mathcal{N}\left(0, I^{-1}(\theta)\right). \tag{5.4.34}$$

Furthermore, if $\bar{\theta}_n$ is a minimum contrast estimate whose corresponding ρ and ψ satisfy **A0-A6**, then

$$\sigma^2(\psi, P_\theta) \geqslant \frac{1}{I(\theta)} \tag{5.4.34}$$

with equality iff $\psi = a(\theta) \frac{\partial \ell}{\partial \theta}$ for some $a \neq 0$.

Summary



Theorem 5.2.3
Consistency
Minimum Contrast Estimate
(5.2.8) & (5.2.9)
Theorem 5.2.2
Consistency
MLE
EF
Theorem 5.3.3
Delta Method
Univariate
Theorem 5.3.5
Asymptotic Normality
MLE
EF
Theorem 5.3.4
Delta Method
Multivariate
Theorem 5.4.2
Asymptotic Normality
Minimum Contrast Estimate
A0 – A5
Theorem 5.4.3
Asymptotic Normality and Efficiency
MLE
A0 – A6

Problems - HW9

Markov's Inequality

$$P[X \geq \epsilon] \leq \frac{E[X]}{\epsilon}$$

$X \geq 0$
$\epsilon > 0$

Chebyshev's Inequality

$$P[|X - E[X]| \geq \epsilon] \leq \frac{Var[X]}{\epsilon^2}$$

$E[X] < \infty$
$Var[X] < \infty$
$\epsilon > 0$

Hoeffding's Inequality

$$P[|\bar{X} - E[\bar{X}]| \geq \epsilon] \leq 2exp\left[-\frac{2n^2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right]$$

$$P[|S_n - E[S_n]| \geq \epsilon] \leq 2exp\left[-\frac{2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right]$$

$X_1, \ldots, X_n \quad i.i.d.$
$a_i \leq X_i \leq b_i$

$$S_n = \sum_{i=1}^{n} X_i$$

$$\bar{X} = \frac{1}{n} S_n$$

References

Bickel, Peter J., and Kjell A. Doksum. (2015) Mathematical statistics: basic ideas and selected topics, volume I. CRC Press.

Thanks for listening!