Section 6: Asymptotic theory and identification

ARE 210

October 10, 2017

The section notes are available on the section Github at github.com/johnloeser/are210 in the "section6" folder.

1 Definitions

- **Lindeberg-Levy CLT**: $\{X_i\}_{i=1}^{\infty}$ iid random variables, X_i finite second moment, $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $Y_n = \frac{\overline{X}_n - \mathbf{E}[\overline{X}_n]}{\sqrt{V[\overline{X}_n]}}$, then $Y_n \overset{d}{\to} N(0,1)$
  - $\{X_i\}_{i=1}^{\infty}$ iid random vectors, mean μ and covariance matrix Σ, then $\sqrt{n}(\overline{X}_n - \mu) \overset{d}{\to} N(0, \Sigma)$
- $X_n = O_p(1)$ if $\forall\ \epsilon$, $\exists\ B_\epsilon, N_\epsilon$ such that $P[|X_n| > B_\epsilon] < \epsilon$
  - $\frac{X_n}{a_n} \overset{d}{\to} X \Rightarrow X_n = O_p(a_n)$
  - $\frac{X_n}{a_n} \overset{p}{\to} 0 \Rightarrow X_n = o_p(a_n)$
- **Continuous mapping theorem**: g continuous $\Rightarrow (X_n \overset{p}{\to} X \Rightarrow g(X_n) \overset{p}{\to} g(X))$
  - $g(F_n) \to g(F)$, where F_n is the empirical distribution
- **Slutsky's lemma**: $X_n \overset{d}{\to} X$, $Y_n \overset{d}{\to} c$ (where c is constant), then $X_n + Y_n \overset{d}{\to} X + c$ and $X_n Y_n \overset{d}{\to} Xc$
  - $\sqrt{n}\overline{X}_n / \sqrt{s_n^2} \to N(0,1)$
  - $P\left[\mu \in \left(\overline{X}_n + \frac{\sqrt{s_n^2}}{\sqrt{n}}\Phi^{-1}(\alpha/2), \overline{X}_n + \frac{\sqrt{s_n^2}}{\sqrt{n}}\Phi^{-1}(1-\alpha/2)\right)\right] \to 1 - \alpha$
- **Delta method**: $\sqrt{n}(T_n - \theta) \overset{d}{\to} Y$, g differentiable at θ, then $\sqrt{n}(g(T_n) - g(\theta))) \overset{d}{\to} g'(\theta)Y$
  - $\sqrt{n}(\overline{X}_n - \mu) \overset{d}{\to} N(0, \Sigma)$, then $\sqrt{n}(g(\overline{X}_n) - g(\mu)) \overset{d}{\to} N(0, g'(\mu)\Sigma g'(\mu)^{\mathrm{T}})$
  - $dx_n \overset{p}{\to} 0$, then $g(x + dx_n) = g(x) + g'(x)dx_n + \frac{1}{2}dx_n g''(\theta)dx_n^{\mathrm{T}} + O_p(dx_n^3)$ if g is thrice differentiable at x
- **Identification**
  - Data $\mathbf{X}$, takes values in sample space $\mathcal{X}$, $\mathbf{X}$ has unknown distribution $P \in \mathbf{P}$ family of probability distributions on $\mathcal{X}$, typically assume
    1. $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$
    2. $\mathbf{X} = \{X_i\}_{i=1}^{n}$ iid, so probability distribution of data $\times_{i=1}^{n} P$
  - A **parameter** is a mapping $\nu : \mathbf{P} \to \mathcal{N}$ (i.e. function of distribution of $\mathbf{X}$)

    - $\theta(P) = \arg\max_{b \in \Theta} Q_0(b, P)$
  - The **identified set** $\Theta(P) = \{\theta \in \Theta : \theta = \arg\max_{b \in \Theta} Q_0(b, P)\}$
    - More generally, $\nu(P)$, where we allow ν to be set valued
    - **Point identification** $\Leftrightarrow$ the identified set is a singleton
  - A typical problem:
    1. Show $\Theta(P)$ is a singleton ($P_{\theta_1} = P_{\theta_2} \Rightarrow \theta_1 = \theta_2$) or characterize the set
    2. Construct an estimator, often $\Theta(P_n)$, where P_n is the empirical distribution
    3. Derive its asymptotic properties using implicit function theorem, delta method, ...

2 Practice questions

1) **$\mathbf{O_p}$ and Taylor approximations**: Let $\{X_i\}_{i=1}^{\infty}$ iid, and $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, and let $\mu = \mathbf{E}[X_i]$.

a) Show that $(\overline{X}_n - \mu)^k$ is $O\left(n^{-k/2}\right)$.

b) Show that if $T_n = O(n^k)$, then $T_n = o(n^{k+\epsilon})\ \forall\ \epsilon > 0$.

2) **OLS**: Consider the OLS model - $Y_i = X_i'\beta + \epsilon_i$, with $(Y_i, X_i', \epsilon_i) \sim P$ iid, and assume $\mathbf{E}_P[X'\epsilon] = 0$ and $\mathbf{E}_P[XX']$.

a) Show P is point identified.

b) Suggest an estimator of β.

c) Additionally, assume $\epsilon_i \perp X_i$. Derive the asymptotic distribution of $\hat{\beta}$.

3) **Selection models and potential outcomes**: Consider the potential outcomes model (Y_{i1}, Y_{i0}, D_i), where Y_{i1} and Y_{i0} are binary, and let $Y_i = Y_{i1}D_i + Y_{i0}(1 - D_i)$. Additionally, consider the selection model $Y_i = \mathbf{1}\{(1, D_i)\beta - U_i \geq 0\}$, where $U_i \sim N(0, 1)$, and assume $U_i \perp D_i$. Suppose (Y_i, D_i) are observed.

a) Show that β is identified, and that the two models are observationally equivalent.

b) Show that $\mathbf{E}[Y_{i1}|D_i = 0]$ is not identified. Calculate $\mathbf{E}[Y_{i1}|D_i = 0]$ using the selection model.

c) How can we make the selection model more flexible to accomodate $\mathbf{E}[Y_{i1}|D_i = 0] \neq \mathbf{E}[Y_i|D_i = 1]$?