Asymptotic Analysis in Econometrics

Lecture 1: Stochastic Convergence

Fu Ouyang

April 22, 2018

Outline

Resnick Chapter 5 & 6

Convergence in Distribution

A random vector $X = (X_1, ..., X_k)$ is a vector of random variables. The distribution function of X is the map $x \mapsto P(X \leq x)$. A sequence of random vectors X_n is said to converge in distribution to a random vector X if

$$P(X_n \leq x) \to P(X \leq x)$$

for every x at which $P(X \leq x)$ is continuous. Alternative names are weak convergence and convergence in law, denoted by $X_n \rightsquigarrow X$.

2.11 Implication of weak convergence

Suppose that $X_n \rightsquigarrow X$ for a random vector X with a continuous distribution function. Then

$$\sup_x |P(X_n \leq x) - P(X \leq x)| \to 0$$

Equivalent Definitions of Weak Convergence

2.2 Portmanteau lemma

The following statements are equivalent:

1. $X_n \rightsquigarrow X$.
2. $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for any bounded, continuous function f.
3. $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for any bounded, Lipschitz function f.
4. $\mathbb{E}\exp\left[it^T X_n\right] \to \mathbb{E}\exp\left[it^T X\right], \forall t \in \mathbb{R}^k$. (Lévy's continuity theorem)
5. $t^T X_n \rightsquigarrow t^T X$ for all $t \in \mathbb{R}^k$. (Cramér-Wold device)
6. $\liminf_{n\to\infty} \mathbb{E}f(X_n) \geq \mathbb{E}f(X)$ for all nonnegative, continuous function f.
7. $\liminf_{n\to\infty} P(X_n \in G) \geq P(X \in G)$ for every open set G.
8. $\limsup_{n\to\infty} P(X_n \in F) \leq P(X \in F)$ for every closed set F.
9. $P(X_n \in B) \geq P(X \in B)$ for every Borel set B with $P(X \in \delta B) = 0$.

Continuous Mapping Theorem and Slutsky's Theorem

2.3 Continuous mapping theorem

Let $g : \mathbb{R}^k \mapsto \mathbb{R}^m$ be continuous on a set C such that $\mathbb{P}(X \in C) = 1$.

1. If $X_n \rightsquigarrow X$, then $g(X_n) \rightsquigarrow g(X)$.
2. If $X_n \xrightarrow{p} X$, then $g(X_n) \xrightarrow{p} g(X)$.
3. If $X_n \xrightarrow{a.s.} X$, then $g(X_n) \xrightarrow{a.s.} g(X)$.

2.8 Slutsky's theorem

Let X_n, X and Y_n be random vectors or variables. If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$ for a constant c, then

1. $X_n + Y_n \rightsquigarrow X + c$.
2. $Y_n X_n \rightsquigarrow cX$.
3. $Y_n^{-1} X_n \rightsquigarrow c^{-1} X$ provided $c \neq 0$.

Relations among Different Convergence Modes

2.7 Relations among different convergence modes

Let X_n, X, Y_n and Y be random vectors and c be a constant. Then

1. $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X$.
2. $X_n \xrightarrow{p} X \Rightarrow X_n \rightsquigarrow X$.
3. $X_n \xrightarrow{p} c \Leftrightarrow X_n \rightsquigarrow c$.
4. If $X_n \rightsquigarrow X$ and $d(X_n, Y_n) \to 0$, then $Y_n \rightsquigarrow X$.
5. If $X_n \rightsquigarrow X$ and $Y_n \xrightarrow{p} c$, then $(X_n, Y_n) \rightsquigarrow (X, c)$.
6. $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y \Leftrightarrow (X_n, Y_n) \xrightarrow{p} (X, Y)$.

Prohorov's Theorem

Tightness and uniform tightness

A random vector X is tight if $\forall \varepsilon > 0$, there is a constant $M > 0$ such that

$$P(\|X\| > M) < \varepsilon$$

A set of random vectors $\{X_\alpha : \alpha \in \mathcal{A}\}$ is called uniformly tight (bounded in probability) if

$$\sup_{\alpha \in \mathcal{A}} P(\|X_\alpha\| > M) < \varepsilon$$

2.4 Prohorov's theorem

Let X_n be random vectors in $\mathbb{R}^k$.

1. If $X_n \rightsquigarrow X$ for some X, then $\{X_n : n \in \mathbb{N}\}$ is uniformly tight.
2. If $\{X_n : n \in \mathbb{N}\}$ is uniformly tight, then there exists a subsequence with $X_{n_j} \rightsquigarrow X$ as $j \to \infty$, for some X.

Stochastic O and o Symbols

Big O and small o

1. $X_n = O_p(1) \Leftrightarrow \{X_n\}$ is uniformly tight.
2. $X_n = o_p(1) \Leftrightarrow X_n \xrightarrow{p} 0$.
3. $X_n = O_p(R_n) \Leftrightarrow X_n = Y_n R_n$ and $Y_n = O_p(1)$.
4. $X_n = o_p(R_n) \Leftrightarrow X_n = Y_n R_n$ and $Y_n = o_p(1)$.

2.12 Rules of calculus with O and o symbols

Let R be a function on $\mathbb{R}^k$ with $R(0) = 0$ and $X_n = o_p(1)$ be a sequence of random vectors with values in the domain of R. Then, $\forall p > 0$,

1. If $R(h) = O(\|h\|^p)$ as $h \to 0$, then $R(X_n) = O_p(\|X_n\|^p)$.
2. If $R(h) = o(\|h\|^p)$ as $h \to 0$, then $R(X_n) = o_p(\|X_n\|^p)$.

Outline

Differentiability

Let $\phi : \mathbb{R}^k \to \mathbb{R}^m$. For $h \in \mathbb{R}^k$, denote

$$\phi'_\theta = \phi'(\theta) \text{ and } \phi'_\theta(h) = \phi'(\theta)\, h$$

ϕ is differentiable at θ if there is a linear map $\phi'_\theta : \mathbb{R}^k \to \mathbb{R}^m$ such that

$$\phi(\theta + h) - \phi(\theta) = \phi'_\theta(h) + o(\|h\|)\, , \text{ as } h \to 0$$

where

$$\phi'_\theta = \begin{pmatrix} \frac{\partial \phi_1}{\partial x_1}(\theta) & \cdots & \frac{\partial \phi_1}{\partial x_k}(\theta) \\ \vdots & \ddots & \vdots \\ \frac{\partial \phi_m}{\partial x_1}(\theta) & \cdots & \frac{\partial \phi_m}{\partial x_k}(\theta) \end{pmatrix}$$

A sufficient condition for ϕ to be differentiable at θ is that all $\partial \phi_j(x) / \partial x_i$ exist for x in a neighborhood of θ and are continuous at θ. ϕ is said to be continuously differentiable if ϕ'_θ is continuous on θ. ϕ'_θ is called gradient when $m = 1$.

Delta Method

3.1 Delta method

Suppose $\phi : \mathbb{D}_\phi \subset \mathbb{R}^k \to \mathbb{R}^m$ is differentiable at θ. Let T_n be random vectors taking their values in $\mathbb{D}_\phi$. If $r_n \left(T_n - \theta\right) \rightsquigarrow T$ for $r_n \to \infty$, then

$$r_n \left(\phi\left(T_n\right) - \phi\left(\theta\right)\right) \rightsquigarrow \phi'_\theta \left(T\right)$$

Moreover, $r_n \left(\phi\left(T_n\right) - \phi\left(\theta\right)\right) - \phi'_\theta \left(r_n \left(T_n - \theta\right)\right) \xrightarrow{p} 0.$

3.8 Uniform delta method

Let $\phi : \mathbb{D}_\phi \subset \mathbb{R}^k \to \mathbb{R}^m$ be continuously differentiable in a neighborhood of θ and T_n be random vectors taking their values in $\mathbb{D}_\phi$. If $\theta_n \to \theta$ and $r_n \left(T_n - \theta_n\right) \rightsquigarrow T$ for $r_n \to \infty$, then

$$r_n \left(\phi\left(T_n\right) - \phi\left(\theta_n\right)\right) \rightsquigarrow \phi'_\theta \left(T\right)$$

Moreover, $r_n \left(\phi\left(T_n\right) - \phi\left(\theta_n\right)\right) - \phi'_\theta \left(r_n \left(T_n - \theta_n\right)\right) \xrightarrow{p} 0.$

Application: Variance Stabilizing Transformations

Suppose that $\sqrt{n}(T_n - \theta) \rightsquigarrow N(0, \sigma^2(\theta))$. Provided that it is consistent, some $\sigma^2(\widehat{\theta})$ can be used in place of unknown $\sigma^2(\theta)$ to conduct various inference.

An alternative approach is called variance-stabilizing transformation, which may lead to a better approximation.

If $\phi(\theta) \equiv \int \sigma^{-1}(\theta) d\theta$ is differentiable, then

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) \rightsquigarrow N(0, \phi'(\theta)^2 \sigma^2(\theta)) = N(0, 1)$$

If ϕ is well-defined and strictly monotone, one can use this approach to derive the asymptotic confidence interval for θ.

Higer-Order Expansions

In the one-dimensional case, Taylor's expansion of $\phi(T_n)$ has the form:

$$\phi(T_n) = \phi(\theta) + (T_n - \theta)\,\phi'(\theta) + \frac{1}{2}(T_n - \theta)^2\,\phi''(\theta) + \cdots$$

When $\phi'(\theta) = 0$, the rate and limiting distribution of $\phi(T_n)$ is determined by the quadratic term:

$$2r_n^2\left[\phi(T_n) - \phi(\theta)\right] = r_n^2 (T_n - \theta)^2\,\phi''(\theta) + \cdots$$

Need to carefully determine the rate if T_n is higher-dimensional:

$$\begin{aligned} \phi(T_n) - \phi(\theta) &= \sum_{k=1}^{K} \frac{\partial \phi}{\partial x_k}(\theta)(T_{nk} - \theta_k) \,| \\ &\quad + \frac{1}{2} \sum_{k=1}^{K} \sum_{l=1}^{K} \frac{\partial^2 \phi}{\partial x_k \partial x_l}(\theta)(T_{nk} - \theta_k)(T_{nl} - \theta_l) \end{aligned}$$

Outline

Resnick Chapter 7

Outline

Resnick Chapter 9

Outline

- lim sup & lim inf
- inequalities