Short overview of asymptotic theory

(Wooldridge, chapter 3)

1 Convergence in probability

Let $\{x_n(\omega)\}_{n=1,2,...}$ be a sequence of random variables or vectors and $x(\omega)$ be another random variable/vector of the same dimension. Both x and x_n are defined in the sample space Ω so that $\omega \in \Omega$. Then $x_n \xrightarrow{p} x$ if for any $\epsilon > 0$

$$\lim_{n\to\infty} P\left[|x_n(\omega) - x(\omega)| > \epsilon\right] = 0$$

which means that x_n is arbitrarily close to x for n sufficiently large.
This form of convergence does not actually require that the sequence x_n converges to x in the sense that it may not converge at any point ω for as long as the size of the set at which x_n and x are far from each other decreases to zero as $n \to \infty$.

Example Consider $\omega \in \Omega = [0, 1]$ and define the following sequence of intervals,

$$\begin{array}{llll} I_1 = [0,1] & & & \\ I_2 = [0,1/2) & I_3 = [1/2,1] & & \\ I_4 = [0,1/3) & I_5 = [1/3,2/3) & I_6 = [2/3,1] & \\ I_7 = [0,1/4) & I_8 = [1/4,2/4) & I_9 = [2/4,3/4) & I_{10} = [3/4,1] \\ \ldots & & & \end{array}$$

Now consider the random variable,

$$x_n(\omega) = \mathbf{1}(\omega \in I_n)$$

and the limit, $x = 0$.
Now we notice that the size of the set where $x_n(\omega) \neq 0$ decreases with n and converges to 0. Thus,

$$x_n \xrightarrow{p} 0$$

However, the set at which $x_n(\omega) = 1$ keeps moving and, for all $\omega \in \Omega$, there is an infinite number of indexes, n, at which $x_n(\omega) = 1$. Thus, we can never guarantee that $x_n(\omega)$ converges to x at any point ω.

Slutsky theorem: Let $g : \mathbb{R}^k \to \mathbb{R}^j$ be a continuous function at a point $c \in \mathbb{R}^k$. Let $\{x_n\}_{n=1,2,...}$ be a sequence of random variables such that $x_n \xrightarrow{p} c$. Then

$$g(x_n) \xrightarrow{p} g(c) \quad \text{or} \quad \text{plim}_{n\to\infty} g(x_n) = g\left(\text{plim}_{n\to\infty} x_n\right)$$

2 Convergence in distribution

This is a weaker form of convergence. Let $\{x_n\}_{n=1,2,...}$ be a sequence of random variables or vectors and x be another random variable/vector. Then $x_n \xrightarrow{d} x$ iff

$$P(x_n < a) = F_n(a) \longrightarrow F(a) = P(x < a)$$

at all continuous points of F.

Result: Let $\{x_n\}_{n=1,2,...}$ be a sequence of random variables such that $x_n \xrightarrow{d} x$. If g is a continuous function then $g(x_n) \xrightarrow{d} g(x)$.

Result: Convergence is distribution implies convergence in probability but the reverse is not true.

Example: Consider the sequence of random variables $x_n \sim \mathcal{N}(0, 1)$ and the random variable x : $-x \sim \mathcal{N}(0, 1)$.
Since the normal distribution is symmetric, the density functions of $\mathcal{N}(0, 1)$ and $-\mathcal{N}(0, 1)$ are the same. Thus,

$$P(x_n \leqslant a) = P(-x_n \leqslant a) = P(x \leqslant a)$$

meaning that,

$$x_n \xrightarrow{d} x$$

However, taking $\epsilon > 0$ and $x_n = -x$:

$$\begin{aligned} P(|x_n - x| < \epsilon) &= P(|-x-x| < \epsilon) \\ &= P(2|x| < \epsilon) \\ &= \Phi\left(\frac{\epsilon}{2}\right) - \Phi\left(\frac{-\epsilon}{2}\right) \end{aligned}$$

which does not converge to 1 with n.

3 Weak Law of Large Numbers

Let $\{x_n\}_{n=1,2,\ldots}$ be a sequence of iid random variables such that $E(|x_n|) < \infty$. Then

$$\frac{1}{N}\sum_{n=1}^{N} x_n \quad \xrightarrow{p} \quad E(x_n)$$

which is to say

$$\overline{x}_n \quad \xrightarrow{p} \quad \mu_x$$

4 Central Limit Theorem (Lindeberg-Levy)

Let $\{x_n\}_{n=1,2,\ldots}$ be a sequence of iid random variables/vectors such that $E(x_n) = 0$ and $E(x_n x_n') < \infty$. Then

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N} x_n \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Sigma_x\right)$$

where $\Sigma_x = V(x) = E\left(x_n x_n'\right)$.

Then we say

$$\overline{x}_N \quad \dot{\sim} \quad \mathcal{N}\left(0, \frac{1}{N}\Sigma_x\right)$$

5 Asymptotic properties of the estimators

Consistency: Let $\left\{\widehat{\theta}_N\right\}_{N=1,2,\ldots}$ be a sequence of estimators of the $k * 1$ vector of coefficients, θ, where N represents the sample size. If $\widehat{\theta}_N \xrightarrow{p} \theta$ for any value of θ then it is said that $\widehat{\theta}_N$ is a consistent estimator of θ.

Asymptotic normality: Let $\left\{\widehat{\theta}_N\right\}_{N=1,2,\ldots}$ be a sequence of estimators of the $k*1$ vector of coefficients, θ, where N represents the sample size. If $\sqrt{N}\left(\widehat{\theta}_N-\theta\right)\stackrel{d}{\longrightarrow}\mathcal{N}(0,\Sigma)$ where Σ is positive semi-definite, then it is said that $\widehat{\theta}_N$ is a asymptotically normally distributed with Σ being the asymptotic variance of $\sqrt{N}\left(\widehat{\theta}_N-\theta\right)$.
Notice that $\widehat{\theta}_N$ will not in general be normally distributed or have a variance-covariance matrix Σ/N. However, we treat it as such: $\widehat{\theta}_N\sim\mathcal{N}(\theta,\Sigma/N)$.

Asymptotic efficiency: Let $\widehat{\theta}_N$ and $\widetilde{\theta}_N$ be two alternative asymptotically normally distributed estimators of θ with respective variance-covariance matrices Σ and Λ. $\widehat{\theta}_N$ is asymptotically more efficient than $\widetilde{\theta}_N$ iff $\Lambda-\Sigma$ is positive semi-definite.

6 Example: the OLS estimator

Consider the model,

$$y_i = x_i\beta+u_i \tag{1}$$

where x_i is $1\times k$ and β is $k\times 1$. The OLS assumptions are,

OLS1. $E(x_i'u_i)=0$ (u is uncorrelated with each regressor)

OLS2. $E(x_i'x_i)=M_{xx}$ has rank k (is positive definite)

We can rearrange equation (1) using (OLS1) and (OLS2) to obtain

$$\beta=E(x_i'x_i)^{-1}E(x_i'y_i)$$

Analogy principle: Turn population moments into their sample counterparts.
In this case, we replace $E(x_i'x_i)$ by $N^{-1}\sum x_i'x_i$ and $E(x_i'y_i)$ by $N^{-1}\sum x_i'y_i$ to obtain:

$$\begin{aligned}\beta^{OLS} &= \left(\frac{1}{N}\sum_{i=1}^N x_i'x_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N x_i'y_i\right)\\ &= (X'X)^{-1}X'Y\end{aligned}$$

6.1 Consistency of OLS under OLS1 and OLS2

The OLS estimator can be re-written as,

$$\beta^{OLS} = \beta + \left(\frac{1}{N}\sum_{i=1}^{N} x_i' x_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} x_i' u_i\right)$$

From the LLN we have

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} x_i' x_i = E\left(x_i' x_i\right) = M_{xx}$$

which by OLS2 is positive definite, and thus invertible. The LLN also implies that

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} x_i' u_i = E\left(x_i' u_i\right)$$

which by OLS1 equals zero.

Thus,

$$\text{plim}_{N\to\infty}\beta^{OLS} = \beta$$

which is to say that OLS consistently estimates β.

6.2 Asymptotic distribution of OLS

We can write,

$$\sqrt{N}\left(\beta^{OLS} - \beta\right) = \left(\frac{1}{N}\sum_{i=1}^{N} x_i' x_i\right)^{-1} \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i' u_i\right)$$

The first term on the rhs converges in probability to M_{xx}^{-1} (OLS2). The second term is the sum of a sequence $\{x_i' u_i\}_{i=1,2,...}$ of iid random variables with zero mean and finite variance. Thus, from the CLT,

$$N^{-1/2}\sum_{i=1}^{N} x_i' u_i \overset{d}{\longrightarrow} \mathcal{N}(0,B)$$

where $B = E\left(u_i^2 x_i' x_i\right)$.

Thus,

$$\sqrt{N}\left(\beta^{OLS} - \beta\right) \overset{p}{\longrightarrow} M_{xx}^{-1} N^{-1/2}\sum_{i=1}^{N} x_i' u_i \overset{d}{\longrightarrow} \mathcal{N}(0,C)$$

where $C = M_{xx}^{-1} B M_{xx}^{-1}$.

Consider the additional homoscedasticity assumption,

OLS3: $E\left(u_i^2 x_i' x_i\right) = \sigma_u^2 M_{xx}$

Under OLS3,

$$\begin{aligned} C &= \sigma_u^2 M_{xx}^{-1} M_{xx} M_{xx}^{-1} \\ &= \sigma_u^2 M_{xx}^{-1} \end{aligned}$$