Convergence in probability

- We denote convergance in probability as

$$\text{plim}_{n\to\infty}(a_n) = a$$

or

$$a_n \xrightarrow{p} a$$

- **Informal definition**: A sequence of random variables converge in probability to a if for large n probability that the difference between a_n and a is large converge to 0.

Covergence of estimators

- Estimator $\widehat{\boldsymbol{\theta}}$ is unbiased if

$$\mathrm{E}\left(\widehat{\boldsymbol{\theta}}\right) = \boldsymbol{\theta}$$

- Estimator $\widehat{\boldsymbol{\theta}}$ is consistent if

$$\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$$

Law of Large Numbers (LLN)

- In large samplest, for randomly drawn sample, the average of the observations is converging to expected value of the variable.

- **Formaly (simplest version):** If y_i are iid with $\mathrm{E}\left(y_i\right) = \boldsymbol{\mu}_y$ and $\mathrm{Var}\left(y_i\right) < \infty$ for $i = 1, \ldots, n$

$$n^{-1} \sum_{i=1}^{n} y_i \xrightarrow{p} \mu_y$$

Convergance in distribution

- We write it as

$$a_n \xrightarrow{D} a$$

where a random variable with distribution $F(\bullet)$,

- **Informal definition:** A sequence of random variables converge a_n converge in distribution to $F(\bullet)$ if for large n distribution a_n becomes close to distribution $F(\bullet)$.

Central Limit Theorem (CLT)

- **Limit theorem:** Central Limit Theorem. If $\boldsymbol{a}_i$ are iid with $\mathrm{E}\left(\boldsymbol{a}_i\right) = \boldsymbol{\mu}$ and $Var\left(\boldsymbol{a}_i\right) = \boldsymbol{B} < \infty$ for $g = 1, \ldots, G$ then

$$\sqrt{n} \sum_{i=1}^{n} \left(\overline{\boldsymbol{a}}_i - \boldsymbol{\mu}\right) \xrightarrow{D} N\left(0, \boldsymbol{B}\right)$$

where $\overline{\boldsymbol{a}}_i = \frac{\sum_{i=1}^{n} a_i}{n}$.

Squared root consistent estimators

- $\sqrt{n}$ consistent estimator (Central Limit Theorem)

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \xrightarrow{D} N\left(0, \boldsymbol{V}\right)$$

- Asymptotic variance $Avar\left(\widehat{\boldsymbol{\theta}}\right) = \boldsymbol{V}/\,n$

- Aproximate distribution of $\widehat{\boldsymbol{\theta}}$: $N\left(\boldsymbol{\theta}, \boldsymbol{V}/\,n\right)$

- $\widehat{\boldsymbol{\theta}}$ more efficient than $\widetilde{\boldsymbol{\theta}}$ if $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \xrightarrow{D} N\left(0, \boldsymbol{V}_1\right)$ and $\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \xrightarrow{D} N\left(0, \boldsymbol{V}_2\right)$ and $\boldsymbol{V}_2 - \boldsymbol{V}_1$ positive semidefinite.

Expected values vs. probability limits

- For random variables a and b and random iid sequences a_n, b_n.

Expected values	Probability limits (for $n \to \infty$)	Theorem
$\mathrm{E}\,(ab) \neq E\,(a)\,E\,(b)$	$\mathrm{plim}\,(a_n b_n) = \mathrm{plim}\,(a_n)\,\mathrm{plim}\,(b_n)$	Cramer
$\mathrm{E}\left(\frac{a}{b}\right) \neq \mathrm{E}\,(a)$	$\mathrm{plim}\left(\frac{a_n}{b_n}\right) = \frac{\mathrm{plim}(a_n)}{\mathrm{plim}(b_n)}$ if $\mathrm{plim}\,(b_n) \neq 0$	Cramer
$\mathrm{E}\,[g\,(a_n)] \neq g\,(\mathrm{E}\,(a_n))$	$\mathrm{plim}\,(g\,(a_n)) = g\,(\mathrm{plim}\,(a_n))$	Slutsky

- Similar properties can also be given for random vectors

- Moreover if $a_n \xrightarrow{p} a$, $b_n \xrightarrow{D} b$ where b is a random variable with known distribution, then $a_n b_n \xrightarrow{D} ab$

Ordinary least squares (Large samples)

- Structural equation

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + u$$

- $y_i, x_1, \ldots, x_K$ random variables
- Structural model represents causal relationship
- Necessary condition for consistency of OLS estimators

$$\mathrm{E}(u) = 0, \;\; \mathrm{Cov}(x_j, u) = 0, \; j = 1, \ldots, N \tag{1}$$

- Sufficient condition for (1)

$$\mathrm{E}\left(u \mid x_1, x_2, \ldots, x_K\right) = \mathrm{E}\left(u \mid \boldsymbol{x}\right) = 0 \tag{2}$$

Conditional expected value (population regression function) is

$$\mathrm{E}\left(y \mid x_1, x_2, \ldots, x_K\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K$$

$$\mathrm{E}\left(y \mid \boldsymbol{x}\right) = \boldsymbol{x}\boldsymbol{\beta}$$

- Under assumption (2) no explanatory variable can help to explain u.

Asymptotic properties of OLS

- We assume that the sample is random: $(\boldsymbol{x}_i, y_i)$ can be treated as iid variables

$$y_i = \boldsymbol{x}_i\beta + u_i \text{ for } i = 1, \ldots, N$$

- **Assumption**: population orthogonality condition

$$\mathrm{E}\left(\boldsymbol{x}_i' u_i\right) = 0 \tag{3}$$

which is equivalent to $\mathrm{E}\left(u\right) = 0$, $Cov\left(\boldsymbol{x}_i, u_i\right) = 0$.

- **Assumption**: full rank condition

$$\mathrm{Rank}\,\mathrm{E}\left(\boldsymbol{x}_i' \boldsymbol{x}_i\right) = K \tag{4}$$

expectation of matrix $\boldsymbol{x}_i'\boldsymbol{x}_i$ is positive semidefinite (no multicollinearity)

- Under these 2 conditions parameter $\boldsymbol{\beta}$ is identified. It means that it can be expressed as the function of population moments.

- Premultipling

$$y_i = \boldsymbol{x}_i\,\boldsymbol{\beta} + u_i$$

by $\boldsymbol{x}_i'$ we get

$$\boldsymbol{x}_i' y_i = \boldsymbol{x}_i'\boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{x}_i' u_i$$

taking expectations

$$\mathrm{E}\left(\boldsymbol{x}_i' y_i\right) = \mathrm{E}\left(\boldsymbol{x}_i'\boldsymbol{x}_i\right)\boldsymbol{\beta} + \underbrace{\mathrm{E}\left(\boldsymbol{x}_i' u_i\right)}_{0}$$

premultipling by $\left[\mathrm{E}\left(\boldsymbol{x}_i'\boldsymbol{x}_i\right)\right]^{-1}$ we get

$$\boldsymbol{\beta} = \left[\mathrm{E}\left(\boldsymbol{x}_i'\boldsymbol{x}_i\right)\right]^{-1}\mathrm{E}\left(\boldsymbol{x}_i' y_i\right)$$

- To estimate $\boldsymbol{\beta}$ we can use the *analogy principle:* use the sample means instead of expectations (population means).

- We obtain

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'y_i\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} \\
&= \boldsymbol{\beta} + \left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'u_i\right)
\end{aligned}
$$

- This derivation of Ordinary Least Estimator use principles of methods of moments

- Under assumption (4) and using Slutsky theorem and Law of Large Numbers

$$\operatorname{plim}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i\right)^{-1} = \left[\operatorname{plim}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i\right)\right]^{-1}$$
$$= \left[\mathrm{E}\left(\boldsymbol{x}'\boldsymbol{x}\right)\right]^{-1}$$

$$\operatorname{plim}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'u_i\right) = 0$$

$$\text{plim}\,\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \underbrace{\text{plim}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i\right)^{-1}}_{[\text{E}(\boldsymbol{x}'\boldsymbol{x})]^{-1}}\underbrace{\text{plim}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'u_i\right)}_{0}$$

$$= \boldsymbol{\beta}$$

Theorem 1. *Under assumptions (3) and (4), the OLS estimator $\widehat{\boldsymbol{\beta}}$ is consistent.*

- OLS consistently estimates the parameters of a linear projection under assumptions (3) and (4)
- If condition (3) or (4) fails than $\boldsymbol{\beta}$ is not identified.
- Under conditions (3) or (4) is generally biased in small samples.

Homoscedasticity

Assumption:

Homoscedasticity assumption:

$$\mathrm{E}\left(u_i^2 \boldsymbol{x}_i' \boldsymbol{x}_i\right) = \sigma^2 \,\mathrm{E}\left(\boldsymbol{x}_i' \boldsymbol{x}_i\right) < \infty, \text{ where } \sigma^2 = \mathrm{E}\left(u_i^2\right) = \mathrm{Var}\left(u_i\right)$$

- Sufficient condition for this assumption is the condition that $\mathrm{E}\left(u_i^2 \middle| \boldsymbol{x}_i\right) = \mathrm{Var}\left(u_i \middle| \boldsymbol{x}_i\right) = \sigma^2$

Theorem 2. *(Asymptotic normality of OLS) Under assumptions (3) and (4) plus homoscedasticity assumption*

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{D} N\left(\mathbf{0}, \sigma^2 \boldsymbol{A}^{-1}\right)$$

where $\boldsymbol{A} = \operatorname{plim}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i\right)$ *can be approximated as* $\widehat{\boldsymbol{A}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i$ *and then variance matrix of the* OLS *estimator be estimated with*

$$Avar\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\sigma}^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

- Usually σ^2 is estimated with $\widehat{\sigma}^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{N-K}$. This estimator is consistent under homoscedasticity assumption.

Heteroscedasticity-Robust Inference

- Failure of homoscedasticity assumption has less serious consequences that failure of orthogonality condition
- Homoscedasticity assumption is not needed for consistency of $\widehat{\boldsymbol{\beta}}$
- The default variance estimators and test statistics are derived under homoscedasticity assumption
- However, it is often the case that the homoscedasticity assumption is violated. In this case we may use:
  - weighted Least Squares - each observation divided by estimate of the standard deviation $\sqrt{\mathrm{Var}\left(u\middle|\,\boldsymbol{x}\right)}$

- adjust the estimator of asymptotic variance so that it is consistent even if heteroscedasticity present the model

- The second procedure is less efficient but much easier to implement as we must not build a separate model for $\mathrm{Var}\left(u|\,\boldsymbol{x}\right)$

- In large samples the efficiency considerations are less important but robustness is still important