Quantitative Methods in Economics
Asymptotics of least squares

Maximilian Kasy

Harvard University

Roadmap, Part II

1. **Asymptotics of least squares**
2. Inference: Testing and confidence sets

Takeaways for these slides

- Convergence in probability, convergence in distribution
- Law of large numbers: sample means go to population expectations in probability
- Central limit theorem: rescaled sample means go to a standard normal in distribution
- Slutsky theorem: combining convergence of parts of some expression
- Application: Least squares is consistent for the best linear predictor, and asymptotically normal

- Random sampling:

$$(Y_i, X_i) \text{ i.i.d.}, \quad X_i' = \begin{pmatrix} X_{i1} & \dots & X_{iK} \end{pmatrix}.$$

where i.i.d. means “independently identically distributed”

- Recall: Linear Predictor

$$E^*(Y_i \,|\, X_i) = X_i'\beta, \quad \beta = [E(X_i X_i')]^{-1} E(X_i Y_i).$$

- Recall: Least-Squares Estimator:

$$b = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i Y_i.$$

- Our goal: Understanding how b relates to β.

Convergence in probability

- Definition: The sequence of random variables Q_n *converges in probability* to a constant α if

$$\lim_{n \to \infty} P(|Q_n - \alpha| > \varepsilon) = 0$$

for all $\varepsilon > 0$. Notation: $Q_n \xrightarrow{p} \alpha$.

- Definition: The estimator b is consistent for β if it converges in probability to β,

$$b \xrightarrow{p} \beta.$$

Questions for you

- Try to describe convergence in probability in words.
- How does it relate to convergence of sequences of numbers?

Convergence in distribution

- Definition: The sequence of random variables Q_n *converges in distribution* to a random variable Q if and only if for all continuity points of F_Q
$$F_{Q_n}(q) \to F_Q(q).$$
- Convergence in probability implies convergence in distribution.
- The reverse is not true,
- except when X is non-random.

Three important theorems
Law of Large Numbers

- Let $W_1, W_2, \ldots$ be a sequence of iid random variables with $E[W_i] = \mu$,
- Let $\overline{W}_n = n^{-1} \sum_{i=1}^{n} W_i$.
- Then

$$\overline{W}_n \xrightarrow{p} E(W_1).$$

Questions for you

- Suppose additionally $\text{Var}(W_i) = \sigma^2 < \infty$.
- What's $E(\overline{W}_n)$?
- What's $\text{Var}(\overline{W}_n)$?

Central limit theorem

- Let $W_1, W_2, \dots$ be a sequence of iid random variables with
  1. $E[W_i] = \mu$,
  2. $Var(W_i) = \sigma^2$,
  3. and $0 < \sigma^2 < \infty$.
- Let $\overline{W}_n = n^{-1} \sum_{i=1}^{n} W_i$.
- Then

$$\frac{\sqrt{n}}{\sigma}(\overline{W}_n - \mu) \to^d N(0,1).$$

Slutsky's theorem

- Let c be a constant,
- suppose $W_n \to^d W$ and $Q_n \to^p c$
- then
    1. $W_n + Q_n \to^d W + c$
    2. $W_n Q_n \to^d Wc$
    3. $W_n / Q_n \to^d W/c$, provided $c \neq 0$.
- In particular, if $W_n \to^d W$ and $Q_n \to^p 0$, then $W_n Q_n \to^p 0$.

OLS and best linear predictor

- Recall again

$$b = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$$
$$\beta = [E(X_i X_i')]^{-1} E(X_i Y_i)$$

- Thus

$$\begin{aligned} b - \beta &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i X_i \beta\right) \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} X_i U_i\right) \end{aligned}$$

where

$$U_i = Y_i - X_i \beta.$$

Applying these theorems to least squares

Questions for you

Use our theorems to characterize the large sample behavior of

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \tag{1}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i \tag{2}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i U_i \tag{3}$$

Solution:

1. Law of large numbers:

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \xrightarrow{p} E(X_1 X_1')$$

2. Law of large numbers:

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{p} E(X_1 Y_1)$$

3. Central limit theorem and $E[X_i U_i] = 0$ (orthogonality condition):

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i U_i \xrightarrow{d} N(0, \mathrm{Var}(X_1 U_1))$$

Questions for you

Use these results and Slutsky's theorem to characterize the large sample behavior of

1. b
2. $\sqrt{n}(b-\beta)$

Solution:

1. Consistency of least squares.
2. Asymptotic normality of least squares.

$$b \xrightarrow{p} [E(X_1 X_1')]^{-1} E(X_1 Y_1) = \beta. \tag{4}$$

$$\sqrt{n}(b - \beta) \xrightarrow{p} N(0, V) \tag{5}$$

where

$$V = [E(X_1 X_1')]^{-1} \operatorname{Var}(X_1 U_1) [E(X_1 X_1')]^{-1}.$$

Questions for you

- Interpret these results.
- How do they relate to each other?
- Make sure you understand where the formula for V is coming from!