International School of Economics at TSU
Econometrics 2
Lasha Chochua

Problem Set 5

Instructions: You are encouraged to solve the problems before the recitation. Additionally, you are encouraged to work in groups. It is **not mandatory** to submit solutions unless stated otherwise. However, if you would like to share your solution, I would be happy to review it.

Problem 1: Show that:

1. If $e \sim \mathcal{N}(0, I_n \sigma^2)$ and $H'H = I_n$, then $u = H'e \sim \mathcal{N}(0, I_n \sigma^2)$.
2. If $e \sim \mathcal{N}(0, AA')$, then $u = A^{-1}e \sim \mathcal{N}(0, I_n)$.

Solution:

Part 1: we have:

$$u = H'e \sim \mathcal{N}(H'0, H' I_n \sigma^2 H).$$

Since $H'0 = 0$, this simplifies to:

$$u \sim \mathcal{N}(0, H' I_n \sigma^2 H).$$

Using the condition that $H'H = I_n$, we get:

$$u \sim \mathcal{N}(0, I_n \sigma^2).$$

Part 2:

$$u = A^{-1}e \sim \mathcal{N}(A^{-1}0, A^{-1}AA'(A^{-1})').$$

Since $A^{-1}0 = 0$, this simplifies to:

$$u \sim \mathcal{N}(0, A^{-1}AA'(A^{-1})').$$

Using the property that $A^{-1}AA'(A^{-1})' = I_n$, we obtain:

$$u \sim \mathcal{N}(0, I_n).$$

Problem 2:

Let $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ be $N_3(\mu, \Sigma)$ with

$$\mu^T = (2, -3, 1)$$

and

$$\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

(a) Find the distribution of $3X_1 - 2X_2 + X_3$.

Solution:

(a)

Let $\mathbf{a} = (3, -2, 1)^T$, then:

$$\mathbf{a}^T\mathbf{X} = 3X_1 - 2X_2 + X_3.$$

Therefore,

$$\mathbf{a}^T\mathbf{X} \sim N(\mathbf{a}^T\mu, \mathbf{a}^T\Sigma\mathbf{a}),$$

where

$$\mathbf{a}^T\mu = (3 \quad -2 \quad 1) \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} = 13$$

and

$$\mathbf{a}^T\Sigma\mathbf{a} = (3 \quad -2 \quad 1)\begin{bmatrix}1 & 1 & 1\\ 1 & 3 & 2\\ 1 & 2 & 2\end{bmatrix}\begin{bmatrix}3\\ -2\\ 1\end{bmatrix} = 9.$$

The distribution of $3X_1 - 2X_2 + X_3$ is $N(13, 9)$.

Problem 3: Let $\mathbf{X}$ be distributed as $N_3(\mu, \Sigma)$, where

$$\mu^T = (1, -1, 2)$$

and

$$\Sigma = \begin{bmatrix}4 & 0 & -1\\ 0 & 5 & 0\\ -1 & 0 & 2\end{bmatrix}.$$

Which of the following random variables are independent? Explain.

1. X_1 and X_2
2. X_1 and X_3
3. X_2 and X_3
4. X_1 and $X_1 + 3X_2 - 2X_3$

Solution:

(1) Independence of X_1 and X_2

Since $\sigma_{12} = \sigma_{21} = 0$, we conclude that X_1 and X_2 are independent.

(2) Independence of X_1 and X_3

Since $\sigma_{13} = \sigma_{31} = -1 \neq 0$, X_1 and X_3 are **not** independent.

(3) Independence of X_2 and X_3

Since $\sigma_{23} = \sigma_{32} = 0$, we conclude that X_2 and X_3 are independent.

(4) Independence of X_1 and $X_1 + 3X_2 - 2X_3$

Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & -2 \end{bmatrix}.$$

Then,

$$A\mathbf{X} = \begin{bmatrix} X_1 \\ X_1 + 3X_2 - 2X_3 \end{bmatrix} \sim N(A\mu, A\Sigma A^T),$$

where

$$A\Sigma A^T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 3 \\ 0 & -2 \end{bmatrix}.$$

Expanding,

$$A\Sigma A^T = \begin{bmatrix} 4 & 6 \\ 6 & 61 \end{bmatrix}.$$

Since the **off-diagonal** term is nonzero, X_1 and $X_1 + 3X_2 - 2X_3$ are **not** independent.

Problem 4: The model is given by:

$$y_i = x_i\beta + e_i, \quad \mathbb{E}(e_i|x_i) = 0$$

where x_i, β, and e_i are scalar. We consider the estimator:

$$\tilde{\beta} = \frac{\bar{y}}{\bar{x}} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}.$$

We assume that x_i and e_i have finite fourth moments and that $\{y_i, x_i\}$ are a random sample.

a. Find $\mathbb{E}(\tilde{\beta}|X)$.
b. Find $\text{Var}(\tilde{\beta}|X)$.
c. Show that $\tilde{\beta} \to_p \beta$ as $n \to \infty$. Does this require any additional assumptions?
d. Find the asymptotic distribution of $\sqrt{n}(\tilde{\beta} - \beta)$ as $n \to \infty$.

Solution

It is worth first noting that:

$$\tilde{\beta} = \frac{\sum_{i=1}^n (x_i\beta + e_i)}{\sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i}\beta + \frac{\sum_{i=1}^n e_i}{\sum_{i=1}^n x_i} = \beta + \frac{\sum_{i=1}^n e_i}{\sum_{i=1}^n x_i}.$$

a.

$$\mathbb{E}\left(\tilde{\beta} - \beta \mid X\right) = \mathbb{E}\left(\frac{\sum_{i=1}^n e_i}{\sum_{i=1}^n x_i}\middle| X\right) = \frac{\sum_{i=1}^n \mathbb{E}\left(e_i \mid x_i\right)}{\sum_{i=1}^n x_i} = 0,$$

$$\mathbb{E}(\tilde{\beta} \mid X) = \beta,$$

so $\tilde{\beta}$ is unbiased for β.

b.

Since $\mathbb{E}(\tilde{\beta} \mid X) = \beta$, we calculate:

$$\mathrm{Var}(\tilde{\beta} \mid X) = \mathbb{E}\left((\tilde{\beta} - \beta)^2 \mid X\right).$$

Expanding:

$$\mathrm{Var}(\tilde{\beta} \mid X) = \mathbb{E}\left(\left(\frac{\sum_{i=1}^n e_i}{\sum_{i=1}^n x_i}\right)^2 \middle| X\right).$$

Since e_i are independent,

$$\mathrm{Var}(\tilde{\beta} \mid X) = \sum_{i=1}^n \frac{\mathbb{E}(e_i^2 \mid X)}{\left(\sum_{i=1}^n x_i\right)^2}.$$

where $\sigma_i^2 = \mathbb{E}(e_i^2 \mid x_i)$. Note: Under the stated assumptions, σ_i^2 may be random, not a constant.

c.

As $n \to \infty$, by the Weak Law of Large Numbers (WLLN) (since the data are iid):

$$\frac{1}{n}\sum_{i=1}^{n} e_i \quad \overset{p}{\to} \quad \mathbb{E}(e_i) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i \quad \overset{p}{\to} \quad \mathbb{E}(x_i) = \mu,$$

say, so if $\mu \neq 0$, then

$$\tilde{\beta} - \beta = \frac{\frac{1}{n}\sum_{i=1}^{n} e_i}{\frac{1}{n}\sum_{i=1}^{n} x_i} \quad \overset{p}{\to} \quad \frac{0}{\mu} = 0.$$

This requires the *assumption* that $\mu \neq 0$.

d.

As $n \to \infty$, by the Central Limit Theorem (CLT) (as e_i is iid)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i \quad \overset{p}{\to} \quad N(0, \sigma^2)$$

where $\sigma^2 = \mathbb{E}(e_i^2)$. Thus (if again $\mu \neq 0$),

$$\sqrt{n}(\tilde{\beta} - \beta) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i}{\frac{1}{n}\sum_{i=1}^{n} x_i} \quad \overset{d}{\to} \quad \frac{N(0, \sigma^2)}{\mu} = N\left(0, \frac{\sigma^2}{\mu^2}\right).$$

Problem 5: Take the linear model

$$y_i = x_i\beta + e_i$$

$$E(e_i \mid x_i) = 0$$

with n observations and x_i is scalar (real-valued). Consider the estimator

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i^3 y_i}{\sum_{i=1}^n x_i^4}$$

Find the asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta)$ as $n \to \infty$.

Solution

Substituting $y_i = x_i\beta + e_i$,

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i^3 y_i}{\sum_{i=1}^n x_i^4}$$

$$= \frac{\sum_{i=1}^n x_i^3 (x_i\beta + e_i)}{\sum_{i=1}^n x_i^4}$$

$$= \beta + \frac{\sum_{i=1}^n x_i^3 e_i}{\sum_{i=1}^n x_i^4}$$

Thus

$$\sqrt{n}(\hat{\beta} - \beta) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i^3 e_i}{\frac{1}{n} \sum_{i=1}^n x_i^4}$$

By the WLLN, if $Ex_i^4 < \infty$, then as $n \to \infty$

$$\frac{1}{n} \sum_{i=1}^n x_i^4 \to_p Ex_i^4.$$

By the LIE and $E(e_i \mid x_i) = 0$, then

$$E(x_i^3 e_i) = E(E(x_i^3 e_i \mid x_i)) = E(x_i^3 E(e_i \mid x_i)) = 0.$$

Then by the CLT, if $E(x_i^6 e_i^2) < \infty$, as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i^3 e_i \to_d N(0, E(x_i^6 e_i^2)).$$

Together,

$$\sqrt{n}(\hat{\beta} - \beta) \to_d N\left(0, \frac{E(x_i^6 e_i^2)}{(Ex_i^4)^2}\right).$$

Problem 6: Use the provided data **attend.dta** to answer this question.

a. To determine the effects of attending lecture on final exam performance, estimate a model relating *stndfnl* (the standardized final exam score) to *atndrte* (the percent of lectures attended). Include the binary variables *frosh* and *soph* as explanatory variables. Interpret the coefficient on *atndrte*, and discuss its significance.

b. How confident are you that the OLS estimates from part a are estimating the causal effect of attendance? Explain.

c. As proxy variables for student ability, add to the regression *priGPA* (prior cumulative GPA) and *ACT* (achievement test score). Now what is the effect of *atndrte*? Discuss how the effect differs from that in part a.

d. What happens to the significance of the dummy variables in part c as compared with part a? Explain.

e. Add the squares of *priGPA* and *ACT* to the equation. What happens to the coefficient on *atndrte*? Are the quadratics jointly significant?

f. To test for a nonlinear effect of *atndrte*, add its square to the equation from part e. What do you conclude?