MATH 531 HOMEWORK 4
A Central Limit Theorem for OLS estimates
February 15, 2025

This homework tracks the lectures from class but for completeness we restate some of the key features below.

Consider a linear model for data as

$$\mathbf{y}_n = X_n \beta + \mathbf{e}_n$$

where $\mathbf{y}$ is an vector of length n, β is an vector of length k, and X_n is a full rank, $n \times p$ matrix. Assume that $\{\mathbf{e}_i\}$ are independent mean zero, with variance $\sigma^2 I_n$ and $E(|be_i|^3) \leq M < \infty$.

For further notation we let $\mathbf{x}_i$ be the i^{th} of X_n , aka the covariates for the i^{th} observation. So X_n is an array that grows as a function of sample size by just adding more rows.

(Another approach is to let X_n be based on a completely new set of n covariates. This is referred to as a triangular array because every sample size has a different X_n. We are avoiding this generalization to make things easier to follow.)

The final set of assumptions on this model is a "story" about how the $\{\mathbf{x}_i\}$ vary with n. We will assume

Assumption A

$$(1/n) X_n^T X_n \to \Gamma$$

uniformly as $n \to \infty$ and where Γ has full rank.

Assumption B

$$\|(n)(X_n^T X_n)^{-1} \mathbf{x}_i\| \leq \alpha_2 < \infty$$

uniformly for all n and i.

For the problems below you should not reproduce parts of the proofs that are the same as what we covered in lecture. Just indicate the new features needed to work the problems.

Problem 1

The first part of these lectures on the CLT considered the simple case when $\mathbf{x}_i$ are scalars and so X_n is just a single column – simple linear regression where the intercept is set equal to zero. We showed the OLS estimate has a limiting normal distribution based on the Lyapunov Central Limit Theorem (LCLT). Specifically we choose $\delta = 1$ in the Lyapunov condition. Show that the Lyapunov condition will hold if we relax the moment condition to be

$$E(|\mathbf{e}_i|^{2+\delta}) \leq M < \infty$$

for any $\delta > 0$.

(See page 4 of Lecture 8 for the details.)

Proof. Suppose that $\mathbb{E}[|e_i|^{2+\delta}] \leq M < \infty$ for any $\delta > 0$. Recall that $Z_i = \frac{x_i e_i}{\sum x_j^2}$. The Lyapunov condition is satisfied if

$$\frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|Z_i|^{2+\delta}] \to 0$$

as $n \to \infty$ for some $\delta > 0$, where $S_n^2 = \sum_{i=1}^{n} \text{Var}[\hat{\beta}]$. From the assumptions, we know that the errors e_i are iid with mean 0 and variance $\sigma^2 I_n$. We know that $\mathbb{E}[|e_i|^{2+\delta}] \leq M$ for some finite constant M. Using the notion that $\mathbb{E}[|e_i|^{2+\delta} \leq M$, we find that

$$\mathbb{E}[|Z_i|^{2+\delta}] = \mathbb{E}\left[\frac{|x_i|^{2+\delta}}{(\sum_j^n x_j^2)^{2+\delta}} |e_i|^{2+\delta}\right] \leq \mathbb{E}\left[\frac{|x_i|^{2+\delta}}{(\sum_j^n x_j^2)^{2+\delta}} M\right] = \frac{|x_i|^{2+\delta}}{(\sum_j^n x_j^2)^{2+\delta}} M$$

We can plug this back into the Lyapunov Condition to find

$$\frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|Z_i|^{2+\delta}] \leq \frac{1}{S_n^{2+\delta}} \frac{\sum_{i=1}^{n} |x_i|^{2+\delta}}{(\sum_{j=1}^{n} x_j^2)^{2+\delta}} M$$

By the assumption from Lecture 08 page 2, we know that $0 < a \leq x_i \leq b < \infty$, therefore we can say that

$$\sum_{i=1}^{n} |x_i|^{2+\delta} = nb^{2+\delta}$$

Hence,

$$\frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|Z_i|^{2+\delta}] \leq \frac{1}{S_n^{2+\delta}} \frac{nb^{2+\delta}}{(\sum_j^n x_j^2)^{2+\delta}} M$$

Now, we can substitute in for S_n^2. We know that S_n^2 is equivalent to the variance of $\widehat{\beta}$, which can be represented as

$$\mathrm{Var}(\widehat{\beta}) = \frac{\sigma^2}{\sum_i x_i^2}$$

Then,

$$S_n^{2+\delta} = (S_n^2)^{2+\delta/2} = \left(\frac{\sigma^2}{\sum_j x_j^2}\right)^{2+\delta/2} = \frac{\sigma^{2+\delta}}{(\sum_j x_j^2)^{2+\delta/2}} \implies \frac{1}{S_n^{2+\delta}} = \frac{(\sum_j x_j^2)^{2+\delta/2}}{\sigma^{2+\delta}}$$

Now, we can substitute this expression back into the Lyapunov Condition expression.

$$\frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|Z_i|^{2+\delta}] \leq \frac{(\sum_j x_j^2)^{2+\delta/2}}{\sigma^{2+\delta}} \cdot \frac{nb^{2+\delta}}{(\sum_j x_j^2)^{2+\delta}} M = \frac{nb^{2+\delta}}{\sigma^{2+\delta}} M \cdot \frac{1}{(\sum_j x_j^2)^{2+\delta/2}}$$

Again, we know that $a \leq x_i \implies \frac{1}{a} \geq \frac{1}{x_i}$, therefore we can say that

$$\frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|Z_i|^{2+\delta}] \leq \frac{Mb^{2+\delta}}{(a\sigma)^{2+\delta}} \cdot \frac{n}{n^{2+\delta/2}}$$

This expression simplifies to

$$\frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|Z_i|^{2+\delta}] \leq \frac{Mb^{2+\delta}}{(a\sigma)^{2+\delta}} \cdot \frac{1}{n^{\delta/2}}$$

Clearly, this expression tends to 0 as $n \to \infty$, therefore

$$\lim_{n\to\infty} \frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|Z_i|^{2+\delta}] \leq 0 \implies \frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|Z_i|^{2+\delta}] \to 0$$

as $n \to \infty$. Hence, the Lyapunov Condition holds for any $\delta > 0$ under the relaxed moment condition. $\square$

Problem 2

In class we showed (after much technical detail !!) that under the assumptions described above

$$\sqrt{n}(\hat{\beta} - \beta) \to^{d} MN(0, \sigma^2 \Gamma^{-1})$$

a more useful result is tied to the actual data where we don't have to include a limiting Γ.

$$(1/\sigma) \left[X_n^T X_n\right]^{1/2} (\hat{\beta} - \beta) \to^{d} MN(0, I) \tag{1}$$

Here $\left[X_n^T X_n\right]^{1/2}$ is the symmetric matrix square root for a positive definite matrix.

This problem covers some of the parts of establishing equation (1). Fix $\mathbf{a} \in \mathbf{R}^k$ and focus on the random variable (a scalar)

$$U_n = \mathbf{a}^T (1/\sigma) \left[X_n^T X_n\right]^{1/2} (\hat{\beta} - \beta).$$

- We want to apply the LCLT to U_n. Show that

  $$U_n = \sum_i Z_i$$

  where $Z_i = w_i \mathbf{e}_i$, $\mathbf{e}_i$ are the random errors from the linear model, and $\mathbf{w}$ a vector. Identify the form for w_i and explain why $E(U_n) = 0$.

  Proof. We have that $U_n = a^T (1/\sigma) \left[X_n^T X_n\right]^{1/2} (\hat{\beta} - \beta)$. We can express the term $(\hat{\beta} - \beta)$ as

$$(\hat{\beta} - \beta) = \frac{\sum x_i e_i}{\sum x_i^2}$$

We substitute into the expression for U_n and we get

$$U_n = a^T (1/\sigma) \left[X_n^T X_n\right]^{1/2} \sum_i \frac{x_i e_i}{\sum x_i^2} = \sum_i \left(a^T (1/\sigma) \left[X_n^T X_n\right]^{1/2} \frac{x_i}{\sum x_i^2} e_i \right)$$

Now, we define a vector w where

$$w_i = a^T (1/\sigma) \left[X_n^T X_n\right]^{1/2} \frac{x_i}{\sum x_i^2}$$

Hence, we have

$$U_n = \sum_i w_i e_i = \sum_i Z_i$$

Next, we check if $\mathbb{E}[U_n] = 0$.

$$\mathbb{E}[U_n] = \mathbb{E}\left[\sum_i Z_i\right] = \mathbb{E}\left[\sum_i w_i e_i\right] = \sum_i w_i \mathbb{E}[e_i] = 0 \quad \square$$

- Explain why $\{Z_i\}$ are independent. (This is easy but important!)

  Recall that $Z_i = w_i e_i$. We know that w_i is a function of X_n and e_i is independent of e_j for all $i \neq j$. X_n is treated as fixed, therefore Z_i is only dependent on e_i, making the set $\{Z_i\}$ mutually independent.

- What is the form of $s_n^2 = Var(U_n)$? (This is easy but important!)

  We have that $U_n = \sum_i w_i e_i$. The errors e_i are independent, and w_i are fixed constants, so we can write the variance as

$$\text{Var}(U_n) = \sum_i \text{Var}(w_i)\text{Var}(e_i) = \sum_i w_i^2 \text{Var}(e_i) = \sigma^2 \sum_i w_i^2$$

  Therefore, $s_n^2 = \sigma^2 \sum_i w_i^2$.

- We want to show U_n/s_n converges in distribution to $N(0,1)$. Verify the relevant Lyapunov condition to prove this result. Use $\delta = 1$ to make this easy and see the lecture notes for hints.

  You may use without proof that Assumptions A and B above imply

  $$\|\sqrt{n}(X_n^T X_n)^{-1/2}\mathbf{x}_i\| \leq \alpha_3 < \infty$$

  uniformly for all n and i.

Proof. We've shown above that the following statements are true

$$\textbf{(1)} \quad U_n = \sum_i^n Z_i = \sum_i^n w_i e_i$$

$$\textbf{(2)} \quad s_n^2 = \text{Var}(U_n) = \sigma^2 \sum_i^n w_i^2$$

We will use these facts to show that $\frac{U_n}{s_n} \to^d N(0,1)$. Using $\delta = 1$, we have the Lyapunov Condition defined as

$$\frac{1}{s_n^3}\sum_i^n \mathbb{E}[|Z_i|^3] \to 0$$

as $n \to \infty$. We can expand the summation term as

$$\sum_i^n \mathbb{E}[|Z_i|^3] = \sum_i^n |w_i|^3 \mathbb{E}|e_i|^3$$

We also know that $\mathbb{E}|e_i|^3 \leq M < \infty$, so it is true that

$$\frac{1}{s_n^3}\sum_i^n \mathbb{E}[|Z_i|^3] \leq \frac{M}{s_n^3}\sum_i^n |w_i|^3$$

Recall that $w_i = \frac{1}{\sigma}a^T(X_n^T X_n)^{-1/2}x_i$. We also have, by assumption, that $\|\sqrt{n}(X_n^T X_n)^{-1/2}x_i\| \leq \alpha_3 < \infty$. This also implies that $\|(X_n^T X_n)^{-1/2}x_i\| \leq \frac{\alpha_3}{\sqrt{n}}$. Now, we consider using the Cauchy-Schwarz inequality in the following way:

$$|w_i| = \frac{1}{\sigma}|a^T(X_n^T X_n)^{-1/2}x_i| \leq \frac{\|a\|}{\sigma}\|(X_n^T X_n)^{-1/2}x_i\| \leq \frac{\|a\|}{\sigma} \cdot \frac{\alpha_3}{\sqrt{n}}$$

Let $K = \frac{\|a\|\alpha_3}{\sigma}$, since it is a fixed constant. Therefore, we have

$$|w_i| \leq \frac{K}{\sqrt{n}} \implies |w_i|^3 \leq \frac{K^3}{n^{3/2}}$$

By summing over i, we have

$$\sum_i^n |w_i|^3 \leq \sum_i^n \frac{K^3}{n^{3/2}} \implies \sum_i^n |w_i|^3 \leq \frac{K^3}{\sqrt{n}}$$

Putting it back together, we can use this inequality in the Lyapunov Condition argument.

$$\frac{1}{s_n^3} \sum_i^n \mathbb{E}[|Z_i|^3] \leq \frac{M}{s_n^3} \sum_i^n |w_i|^3 \leq \frac{MK^3}{\sqrt{n} s_n^3}$$

Clearly, as $n \to \infty$, the Lyapunov Condition will tend to 0. Therefore, for $\delta = 1$, by the LCLT, we have that

$$\frac{U_n}{s_n} \to^d N(0,1) \quad \Box$$

- What "device" is used to argue that this CLT result for U_n now implies equation (1).

  Cramer-Wold Device!!!

Problem 3

Now for something different. We shift our attention to $\widehat{\sigma^2}$.

$$\widehat{\sigma^2} = (Y - X\hat{\beta})^T (Y - X\hat{\beta})/(n-k) = [(I-M)\mathbf{y}]^T [(I-M)\mathbf{y}]/(n-k)$$

where, of course, M is the projection matrix onto $\mathcal{W}_X$. E.g $M = X(X^T X)^{-1} X^T$.

- By judiciously adding and subtracting $X\beta$ and using the properties of M show that
$$\begin{aligned}\widehat{\sigma^2} &= \mathbf{e}^T (I-M)\mathbf{e}/(n-k) = \mathbf{e}^T\mathbf{e}/(n-k) - \mathbf{e}^T M\mathbf{e}/(n-k) \\ &= (\sum_i \mathbf{e}_i^2)/(n-k) - \mathbf{e}^T M\mathbf{e}/(n-k)\end{aligned}$$
  ($\mathbf{e}$ are the errors in the linear model.)

Proof. By definition, we have that

$$\widehat{\sigma}^2 = \frac{[(I-M)y]^T[(I-M)y]}{n-k}$$

with $M = X(X^TX)^{-1}X^T$. We can left multiply the linear model by $(I-M)$ to find

$$(I-M)Y = (I-M)X\hat{\beta} + (I-M)e$$

We can expand this term by distributing.

$$\begin{aligned}(I-M)Y &= X\hat{\beta} - MX\widehat{\beta} + e - Me \\ (I-M)Y &= X\widehat{\beta} - X(X^TX)^{-1}X^TX(X^TX)^{-1}Xy + e - Me \\ (I-M)Y &= e - Me\end{aligned}$$

Therefore,

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{(e-Me)^T(e-Me)}{n-k} \\ &= \frac{e^Te - 2e^TMe + e^TMe}{n-k} \\ &= \frac{e^Te - e^TMe}{n-k}\end{aligned}$$

Hence,

$$\widehat{\sigma}^2 = \frac{\sum_i^n e_i^2}{n-k} - \frac{e^TMe}{n-k} \quad \square$$

- Use Markov's inequality to show that $\sqrt{n}(\mathbf{e}^TM\mathbf{e})/(n-k) \to^P 0$
  Hint: $E(\mathbf{e}^TM\mathbf{e}) = tr(ME(\mathbf{ee}^T) = \sigma^2 k$. Why?)

Proof. Define $X_n = \sqrt{n}(e^TMe)/(n-k)$. Since M is a projection matrix (and therefore positive semidefinite), then $X_n \geq 0$. Markov's inequality states that for a random variable $X_n \geq 0$, then for any $c > 0$,

$$\Pr[X_n > c] \leq \frac{\mathbb{E}[X_n]}{c}$$

First, we compute $\mathbb{E}[X_n]$.

$$\mathbb{E}[X_n] = \mathbb{E}\left[\frac{\sqrt{n}e^T Me}{n-k}\right] = \frac{\sqrt{n}\mathbb{E}[e^T Me]}{n-k}$$

It's given that $\mathbb{E}[e^T Me] = \sigma^2 k$. Thus,

$$\mathbb{E}[X_n] = \frac{\sqrt{n}\sigma^2 k}{n-k} = \frac{\sqrt{n}}{n-k} \cdot \frac{\sigma^2}{n-k}$$

Note that this expression behaves like $\frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$. Therefore, we can say that $\mathbb{E}[X_n] \to 0$ as $n \to \infty$. Therefore,

$$\frac{\sqrt{n}e^T Me}{n-k} \to^P 0$$

as $n \to \infty$. $\square$

- Use the weak law of large numbers and Slutsky's theorem (twice) to show

$$\widehat{\sigma^2} \to^P \sigma^2$$

Proof. We have that

$$\widehat{\sigma}^2 = \frac{1}{n-k}\sum_i^n e_i^2 - \frac{e^T Me}{n-k}$$

From the previous problem, we proved that the term $\sqrt{n}\frac{e^T Me}{n-k}$ converges in probability to 0. Thus, it would follow that the term $\frac{e^T Me}{n-k}$ converges in probability to 0. That leaves us to analyze the asymptotic behavior of the first term on the RHS, namely $\frac{1}{n-k}\sum_i^n e_i^2$. In order to apply the weak law of large numbers to this term, we can multiply this term by $\frac{n}{n}$ to obtain

$$\frac{1}{n-k}\sum_i^n e_i^2 = \frac{n}{n-k}\frac{1}{n}\sum_i^n e_i^2$$

By the weak law of large numbers, it is easy to say that $\frac{1}{n}\sum_i^n e_i^2 \to^P \sigma^2$ since $\mathbb{E}(e_i^2) = \sigma^2$ and e_i are iid. Additionally, the term $\frac{n}{n-k} \to 1$ as $n \to \infty$. Now, let $X_n = \frac{n}{n-k}$ and $Y_n = \frac{1}{n}\sum_i^n e_i^2$. Now that we know the asymptotic behaviors of X_n and Y_n, then we can say that

$$X_n \cdot Y_n \to^P 1 \cdot \sigma^2 = \sigma^2$$

Now, we apply Slutsky's Theorem again to the original expression for $\widehat{\sigma}^2$. Let $X_n = \frac{1}{n-k}\sum_i^n e_i^2$ and $Y_n = \frac{e^T M e}{n-k}$. We know now that $X_n \to^P \sigma^2$ and $Y_n \to^P 0$, therefore, by Slutsky's Theorem, we can say

$$X_n - Y_n \to^P \sigma^2 - 0 = \sigma^2$$

Hence, $\widehat{\sigma}^2 \to^P \sigma^2$. $\square$

- Assume that $E(\mathbf{e}_i^4) = \rho < \infty$ and $E(|\mathbf{e}_i|^5) = M < \infty$. Using the LCLT and Slutsky's theorem show that

$$\sqrt{n}(\widehat{\sigma^2} - \sigma^2)/\rho \to N(0,1)$$

Hint Apply the LCLT to just $(\sum_i \mathbf{e}_i^2)/n$ (this is not hard!) then prove this problem's result using Slutsky's theorem to deal with the extra bits.

Proof. Let $Z_i = \widehat{\sigma}^2 - \sigma^2$ for $i = 1, ..., n$. We have the following facts

$$(1)\ \mathbb{E}[Z_i] = 0$$
$$(2)\ \mathrm{Var}(Z_i) = \rho - \sigma^4$$

Given that $\mathbb{E}[|e_i|^5] < \infty$ we can apply the LCLT to $\sum_i Z_i$.

$$\sqrt{n}(\widehat{\sigma}^2 - \sigma^2) = \frac{1}{\sqrt{n}}\sum_i^n Z_i \to^d N(0, \rho^2)$$

with $\rho^2 = \mathbb{E}[e_i^4] - \sigma^4$. If we normalize the expression with ρ and multiply the term by $\frac{n}{n}$, then we have

$$\frac{\sqrt{n}(\widehat{\sigma}^2 - \sigma^2)}{\rho} \to^d N(0,1) \quad \square$$