ECMT 676: Econometrics II

Homework I - Due date 02/06/2025

Sebastian Sanin-Restrepo*

Exercise 1

Prove the Law of Large Numbers stated above. (Hint: Use Markov's inequality).

Theorem *(Law of Large Numbers)*: Suppose that $(X_1, ..., X_n)$ are independent and identically distributed copies of an underlying random variable X satisfying $Var(X) < \infty$. Then $\bar{X}_n \to_p \mathbb{E}[X]$.

Markov's inequality states:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

The LLN implies that:

$$\bar{X}_n \to_p \mathbb{E}[X] \implies \lim_{n\to\infty} \mathbb{P}(||\bar{X} - \mathbb{E}[X]|| \geq \epsilon) = 0$$

Since the sample mean $\bar{X}_n$ is an unbiased estimator $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X]$ and $Var(\bar{X}_n) = \frac{\sigma^2}{n}$, where $\sigma^2 = Var(X_i)$, we can use the Chebyshev's inequality, as a special case of the Markov's inequality. Chebyshev's inequality states:

$$\mathbb{P}(|Y - \mathbb{E}[Y]| \geq \epsilon) \leq \frac{Var(Y)}{\epsilon^2}$$

In this case,

$$\mathbb{P}(|X_n - \mathbb{E}[X]| \geq \epsilon) \leq \frac{Var(X_n)}{\epsilon^2}$$

Substituting $Var(\bar{X}_n) = \frac{\sigma^2}{n}$,

$$\mathbb{P}(|X_n - \mathbb{E}[X]| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

Now, take the limit as $n \to \infty$. Notice that the right side becomes $\lim_{n\to\infty} \frac{}{\sigma^2} n\epsilon^2 = 0$. Therefore,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - \mathbb{E}[X]| \geq \epsilon) \leq 0$$

*Economics, Ph.D. Student. E-mail: sr.sebastian@tamu.edu

since a probability cannot be less than zero, we have:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - \mathbb{E}[X]| \geq \epsilon) = 0$$

Thus, it must be the case that $\bar{X}_n \to_p \mathbb{E}[X]$.

Exercise 2

Suppose that $X_n \to_p 0$. Prove that there exists a sequence $\varepsilon_n \searrow 0$ such that $Pr(X_n > \varepsilon_n) < \varepsilon_n$.

Suppose that $X_n \to_p 0$, meaning that for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n| \geq \epsilon) = 0.$$

We wish to prove that there exists a sequence $\varepsilon_n \searrow 0$ (i.e., a decreasing sequence converging to zero) such that for every n,

$$\mathbb{P}(X_n > \varepsilon_n) < \varepsilon_n.$$

Note that simply taking $\varepsilon_n = 1/n$ is not sufficient since the inequality may fail for some n. The key idea is to tune the level ε_n so that it adapts to the tail probability $\mathbb{P}(X_n > \varepsilon)$ as n varies.

For any fix n, and any fixed $\epsilon > 0$, since $X_n \to_p 0$ there exists N_ε (depending on ϵ) such that for all $n \geq N_\varepsilon$ we have $\mathbb{P}(X_n > \epsilon) < \epsilon$. Thus, for each n, the set

$$A_n := \{\varepsilon > 0 : \mathbb{P}(X_n > \varepsilon) < \varepsilon\}$$

is nonempty (at least for large n). So, we can define ε_n^* as the infimum of this set:

$$\varepsilon_n^* := \inf A_n,$$

Notice that it could still be the case that the sequence $\{\varepsilon_n^*\}$ may not be monotone. So, to ensure that this sequence is convergent to zero and is decreasing ($\varepsilon_n \searrow 0$), we modify ε_n^* by considering *all* indices larger than n. Define

$$\delta_n := \inf\left\{\varepsilon > 0 : \sup_{k\geq n} \mathbb{P}(X_k > \varepsilon) < \varepsilon\right\}.$$

So, notice that for each fixed $\varepsilon > 0$, the convergence $X_k \to_p 0$ guarantees that for all sufficiently large k, we have $\mathbb{P}(X_k > \varepsilon) < \varepsilon$. Therefore, for n large enough, the set

$$\left\{\varepsilon > 0 : \sup_{k\geq n} \mathbb{P}(X_k > \varepsilon) < \varepsilon\right\}$$

is nonempty, making δ_n well-defined.Now, I'd like to point out that this sequence is monotonic decreasing. If $m \geq n$, then

$$\sup_{k \geq m} \mathbb{P}(X_k > \varepsilon) \leq \sup_{k \geq n} \mathbb{P}(X_k > \varepsilon)$$

for any $\varepsilon > 0$. Thus, the admissible set for δ_m is a subset of that for δ_n. This implies $\delta_m \leq \delta_n$ for $m \geq n$, so the sequence δ_n is decreasing.

Finally, notice that it converges to zero. Given any $\epsilon > 0$, by the convergence in probability, there exists N such that for all $k \geq N_\varepsilon$, $\mathbb{P}(X_k > \epsilon) < \epsilon$. Consequently, for all $n \geq N_\varepsilon$,

$$\sup_{k \geq n} \mathbb{P}(X_k > \epsilon) \leq \epsilon.$$

This shows that ϵ is an admissible value for the infimum defining δ_n when n is large, and hence $\delta_n \leq \epsilon$. Since this holds for every $\epsilon > 0$, it follows that $\delta_n \to 0$ as $n \to \infty$.

Finally, we verify if this approach is what we were looking for. By the very definition of δ_n, we have that for each n,

$$\sup_{k \geq n} \mathbb{P}(X_k > \delta_n) < \delta_n.$$

In particular, when $k = n$, this gives

$$\mathbb{P}(X_n > \delta_n) < \delta_n.$$

Thus, setting

$$\varepsilon_n := \delta_n = \inf \left\{ \varepsilon > 0 : \sup_{k \geq n} \mathbb{P}(X_k > \varepsilon) < \varepsilon \right\},$$

we obtain a sequence that is decreasing, converges to 0, and satisfies

$$\mathbb{P}(X_n > \varepsilon_n) < \varepsilon_n \quad \text{for every } n.$$

Thus, we have successfully constructed a sequence $\varepsilon_n \searrow 0$ satisfying $\mathbb{P}(X_n > \varepsilon_n) < \varepsilon_n$ for every n by defining

$$\varepsilon_n = \inf \left\{ \varepsilon > 0 : \sup_{k \geq n} \mathbb{P}(X_k > \varepsilon) < \varepsilon \right\}.$$

This sequence works because the convergence $X_n \to_p 0$ ensures that for any fixed ϵ the tail probabilities eventually fall below ϵ. By taking the supremum condition over all indices $k \geq n$, we can be sure that ε_n is such that $\varepsilon_n \searrow 0$.

Exercise 3

Show that:

1. $O_p(1) + O_p(1) = O_p(1)$

   I'd like first to state the definitions applicable to all exercises onwards. We say that a random variable X_n is $O_p(1)$ if for all $\epsilon > 0$ there is a constant M such that $\sup_n \mathbb{P}(|X_n| > M) < \epsilon$. On the other hand, we say a random variable X_n is $o(1)$ if $X_n \to_p 0$, that is if X_n converges in probability to zero. Recall that convergence in probability $X_n \to_p X$ means for any $\epsilon > 0$ there exists N_ϵ such that for all $n \geq N_\epsilon$, $\mathbb{P}(|X_n - X| \geq \epsilon) < \epsilon$

   Let $X_n = O_p(1)$ and $Y = O_p(1)$. By definition, for every $\epsilon > 0$ there exist constants $M_1 > 0$ and $M_2 > 0$, and corresponding integers N_1 and N_2 such that

   $$\mathbb{P}\left(|X_n| > M_1\right) < \frac{\epsilon}{2} \quad \text{for all } n \geq N_1,$$

   $$\mathbb{P}\left(|Y_n| > M_2\right) < \frac{\epsilon}{2} \quad \text{for all } n \geq N_2.$$

   We want to show that $Z_n = X_n + Y_n$ is also $O_p(1)$. Let $N = \max\{N_1, N_2\}$ and define $M = M_1 + M_2$. We want to show that

   $$\mathbb{P}\left(|X_n + Y_n| > M\right) < \epsilon \quad \text{for all } n \geq N.$$

   We know by the triangle inequality that:

   $$|Z_n| = |X_n + Y_n| \leq |X_n| + |Y_n|$$

   Thus, if $|X_n + Y_n| > M$, then it must be the case that $|X_n| + |Y_n| > M_1 + M_2$. This implies that at least one of the following must occur:

   $$|X_n| > M_1 \quad \text{or} \quad |Y_n| > M_2.$$

   Thus,

   $$\mathbb{P}(|Z_n| > M) = \mathbb{P}\left(|X_n + Y_n| > M\right) \leq \mathbb{P}\left(|X_n| > M_1\right) + \mathbb{P}\left(|Y_n| > M_2\right)$$

   Then, by our assumptions we have that $\forall n \geq N$

   $$\mathbb{P}(|Z_n| > M) \leq \mathbb{P}(|X_n| > M_1) + \mathbb{P}(|Y_n| > M_2) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

   Since $\epsilon > 0$ was arbitrary, this proves that $Z_n = X_n + Y_n$ is bounded in probability, so

$Z_n = O_p(1)$. For any $\epsilon > 0$, we can find an $M > 0$ such that $\sup_n \mathbb{P}(|Z_n| > M) < \epsilon$.

2. $O_p(1) \cdot O_p(1) = O_p(1)$

   Let $X_n = O_p(1)$ and $Y = O_p(1)$. We want to show that $Z_n = X_n \cdot Y_n$ is also $O_p(1)$. Since $X_n = O_p(1)$ for any $\epsilon > 0$ there exists $M_1 > 0$ such that:

$$\mathbb{P}(|X_n| > M_1) < \frac{\epsilon}{2} \text{ for all } n \geq N_1,$$

   In the same spirit, since $Y_n = O_p(1)$ for any $\epsilon > 0$ there exists $M_2 > 0$ such that:

$$\mathbb{P}(|Y_n| > M_2) < \frac{\epsilon}{2} \text{ for all } n \geq N_2,$$

   Let $N = \max\{N_1, N_2\}$ and set $M = \max\{M_1, M_2\}$. Notice that if $|X_n Y_n| > M^2$, then it must be that

$$|X_n| > M \quad \text{or} \quad |Y_n| > M.$$

   Thus, we have:

$$\mathbb{P}\big(|Z_n| > M^2\big) = \mathbb{P}\big(|X_n Y_n| > M^2\big) \leq \mathbb{P}(|X_n| > M) + \mathbb{P}(|Y_n| > M)$$

   Since $M \geq M_1$ and $M \geq M_2$, we have

$$\mathbb{P}(|X_n| > M) \leq \mathbb{P}(|X_n| > M_1) < \frac{\epsilon}{2},$$

   and symmetrically for Y_n

$$\mathbb{P}(|Y_n| > M) < \frac{\epsilon}{2}$$

   Therefore, for all $n \geq N$,

$$\mathbb{P}\big(|Z_n| > M^2\big) = \mathbb{P}\big(|X_n Y_n| > M^2\big) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

   This shows that $Z_n = X_n Y_n$ is bounded in probability. In other words, $Z_n = O_p(1)$

3. $O_p(1) \cdot o_p(1) = o_p(1)$

   Let $X_n = O_p(1)$ and $Y = o_p(1)$. We want to show that $Z_n = X_n \cdot Y_n$ is also $o_p(1)$. Since $X_n = O_p(1)$ for any $\epsilon > 0$ there exists $M_1 > 0$ such that:

$$\mathbb{P}(|X_n| > M_1) < \frac{\epsilon}{2} \text{ for sufficiently large } n$$

   On the other hand, since $Y_n = o_p(1)$ for any $\delta > 0$, there exists N such that for

$n \geq N$:

$$P(|Y_n| > \delta) \to 0 \quad \text{as } n \to \infty.$$

We want to show that $Z_n = X_n Y_n = o_p(1)$ meaning $\forall \eta$,

$$P(|X_n Y_n| > \eta) \to 0 \quad \text{as } n \to \infty.$$

To do so, fix an arbitrary $\eta > 0$. By the stochastic boundedness of X_n, there exists $M > 0$ such that

$$\limsup_{n\to\infty} P(|X_n| > M) < \frac{\eta}{2}$$

Then, we can bound the probability as follows:

$$P(|X_n Y_n| > \eta) \leq P(|X_n| > M) + P\Big(|Y_n| > \frac{\eta}{M}\Big).$$

Since $P(|X_n| > M) < \frac{\eta}{2}$ for sufficiently large n, and $P\Big(|Y_n| > \frac{\eta}{M}\Big) \to 0$ as $n \to \infty$ (because $Y_n = o_p(1)$), we can see that

$$\limsup_{n\to\infty} P(|X_n Y_n| > \eta) \leq \frac{\eta}{2} + 0 = \frac{\eta}{2}$$

Since $\eta > 0$ was arbitrary, we conclude that $P(|X_n Y_n| > \eta) \to 0$ as $n \to \infty$. This means that $Z_n = X_n Y_n = o_p(1)$

4. $o_p(1) \cdot o_p(1) = o_p(1)$

   Let $X_n = o_p(1)$ and $Y = o_p(1)$. We want to show that $Z_n = X_n \cdot Y_n$ is also $o_p(1)$. Since $X_n = o_p(1)$ and $Y_n = o_p(1)$ for any $\delta > 0$,:

   $$P(|X_n| > \delta) \to 0 \quad \text{and} \quad P(|Y_n| > \delta) \to 0 \quad \text{as } n \to \infty$$

   We want to show that $Z_n = o_p(1)$ so for any $\delta > 0$:

   $$P(|Z_n| > \delta) \to 0 \quad \text{as } n \to \infty$$

   To prove this, fix an arbitrary $\varepsilon > 0$ and choose any $\delta > 0$. Notice that if $|X_n Y_n| > \varepsilon$, then either

   $$|X_n| > \delta \quad \text{or} \quad |Y_n| > \frac{\varepsilon}{\delta}.$$

   Hence,

   $$P(|X_n Y_n| > \varepsilon) \leq P(|X_n| > \delta) + P\left(|Y_n| > \frac{\varepsilon}{\delta}\right).$$

   Since $X_n = o_p(1)$ and $Y_n = o_p(1)$, both probabilities on the right-hand side tend to zero as $n \to \infty$. Therefore,

$$P(|X_n Y_n| > \varepsilon) \to 0,$$

Therefore, we have shown that $Z_n = o_p(1)$.

Exercise 4

Suppose that there exists a constant $M > 0$ such that $Var(X_n) < M$ for all n. Show that $X_n = O_p(1)$

We have that $Var(X_n) < M$, we want to prove that $X_n = O_p(1)$, which means that for any $\epsilon > 0$, there exists a constant $C > 0$ such that $\mathbb{P}(|X_n| > C) < \epsilon$ for a sufficiently large n. From the definition of variance, we have:

$$Var(X_n) = \mathbb{E}[(X_n - \mathbb{E}[X_n])^2]$$

Using Chebyshev's inequality, we obtain that for any $\delta > 0$:

$$\mathbb{P}(|X_n - \mathbb{E}[X_n]| \geq \delta) \leq \frac{Var(X_n)}{\delta^2}$$

Since $Var(X_n) < M$ we have:

$$\mathbb{P}(|X_n - \mathbb{E}[X_n]| \geq \delta) < \frac{M}{\delta^2}$$

For any constant $C > 0$,

$$\mathbb{P}(|X_n| > C) \leq \mathbb{P}(|X_n - \mathbb{E}[X_n]| > C/2) + \mathbb{P}(|\mathbb{E}[X_n]| > M/2)$$

Using again Chebyshev's inequality we find for the first term:

$$\mathbb{P}(|X_n - \mathbb{E}[X_n]| > C/2) \leq \frac{4M}{C^2}$$

For the second term, since $|E[X_n]|$ is finite for all n,

$$\mathbb{P}(|\mathbb{E}[X_n]| > M/2) = 0 \quad \text{for sufficiently large } M$$

Thus,

$$\mathbb{P}(|X_n| > C) \leq \frac{4M}{C^2}$$

For any $\epsilon > 0$, choose C sufficiently large such that $\frac{4M}{C^2} < \epsilon$, obtaining:

$$\mathbb{P}(|X_n| > C) \leq \epsilon$$

which guarantees that $X_n = O_p(1)$.

Exercise 5

Suppose that $X \in \mathbb{R}$ and that $\tilde{X} = X + \eta$. What is the solution to the minimization problem $\tilde{\beta} := arg\ min_{b \in \mathbb{R}} \mathbb{E}[(Y - b\tilde{X})^2]$ and how does it relate to the true parameter $\beta := arg\ min_{b \in \mathbb{R}} \mathbb{E}[(Y - bX)^2]$.

Notice that the solution of:

$$\tilde{\beta} := arg\ min_{b \in \mathbb{R}} \mathbb{E}[(Y - b\tilde{X})^2]$$

is given by:

$$\tilde{\beta} = \mathbb{E}[\tilde{X}'\tilde{X}]^{-1}\mathbb{E}[\tilde{X}'Y]$$

while the solution of

$$\beta := arg\ min_{b \in \mathbb{R}} \mathbb{E}[(Y - bX)^2]$$

is given by:

$$\beta = \mathbb{E}[X'X]^{-1}\mathbb{E}[X'Y]$$

Since $\tilde{X} = X + \eta$, we have:

$$\begin{aligned}\tilde{\beta} =& \mathbb{E}[(X+\eta)'(X+\eta)]^{-1}\mathbb{E}[(X+\eta)'Y]\\ =& \mathbb{E}[X'X + X'\eta + \eta'X + \eta'\eta]^{-1}\mathbb{E}[X'Y + \eta'Y]\end{aligned}$$

Notice in the numerator:

$$\mathbb{E}[X'Y + \eta'Y] = \mathbb{E}[X'Y] + \mathbb{E}[\eta'Y]$$

We can assume that $\eta \perp Y$, so $\mathbb{E}[\eta'Y] = \mathbb{E}[\eta']\mathbb{E}[Y]$. If $\mathbb{E}[\eta] = 0$ we have $\mathbb{E}[\eta'Y] = 0$, so the numerator would be $\mathbb{E}[X'Y + \eta'Y] = \mathbb{E}[X'Y]$.

Now the denominator is given by:

$$\mathbb{E}[X'X + X'\eta + \eta'X + \eta'\eta] = \mathbb{E}[X'X + 2X'\eta + \eta'\eta] = \mathbb{E}[X'X] + \mathbb{E}[2X'\eta] + \mathbb{E}[\eta'\eta]$$

Since $X \perp \eta$, $\mathbb{E}[X'\eta] = 0$, so the denominator would be $\mathbb{E}[X'X] + \mathbb{E}[\eta'\eta]$.

Taking both conclusions together, we have:

$$\tilde{\beta} = [\mathbb{E}[X'X] + \mathbb{E}[\eta'\eta]]^{-1}[\mathbb{E}[X'Y]]$$

Using the equations for β and $\tilde{\beta}$ we have,

$$\beta[\mathbb{E}[X'X]] = [\mathbb{E}[X'Y]] \quad \text{and} \quad \tilde{\beta}[\mathbb{E}[X'X] + \mathbb{E}[\eta'\eta]] = [\mathbb{E}[X'Y]]$$

Substituting one in another,

$$\beta[\mathbb{E}[X'X]] = \tilde{\beta}[\mathbb{E}[X'X] + \mathbb{E}[\eta'\eta]]$$

Solving for $\tilde{\beta}$,

$$\tilde{\beta} = \beta[\mathbb{E}[X'X] + \mathbb{E}[\eta'\eta]]^{-1}[\mathbb{E}[X'X]]$$

If they are vectors, we would have

$$\tilde{\beta} = \beta\frac{Var(X)}{Var(X) + Var(\eta)}$$

Exercise 6

Let A be positive definite. Show that $A^{-1/2} = (A^{1/2})^{-1}$ is a symmetric square root of A^{-1}.

Since A is positive definite, this ensures that $A^{1/2}$ exists, is symmetric, and satisfies $A^{1/2}A^{1/2} = A$, and $A^{1/2}$ is positive definite. Since the inverse of a positive definite matrix is also positive definite, A^{-1} is positive definite so $A^{-1/2}$ exists, is symmetric, and satisfies $A^{-1/2}A^{-1/2} = A^{-1}$.

To prove that $A^{-1/2} = (A^{1/2})^{-1}$, we have that by the property of matrix inverse, if $A^{1/2}$ satisfies $A^{1/2}A^{1/2} = A$, then $(A^{1/2})^{-1}(A^{1/2})^{-1} = (A^{1/2}A^{1/2})^{-1} = A^{-1}$. Thus, $(A^{1/2})^{-1}$ satisfies the defining property of $A^{-1/2}$ meaning that $A^{-1/2} = (A^{1/2})^{-1}$.

Since $A^{1/2}$ is symmetric so its inverse $A^{-1/2}$ is also symmetric.

Exercise 7

Corollary 1.1: Suppose that Assumptions 1.1 and 1.2 hold and that we have a consistent estimator $\hat{\Sigma}$ such that $\hat{\Sigma} \to_p \Sigma$. Then, $\sqrt{n}\hat{\Sigma}^{-1/2}(\hat{\beta} - \beta) \to_d N(0, 1)$.

Prove Corollary 1.1. (Hint: you can take Theorem 1.1 as given).

From Theorem 1.1. we know that:

$$\sqrt{n}(\hat{\beta} - \beta) \to_d N(0, \Sigma)$$

where $\Sigma := (E_n[Z_iX_i'])^{-1}E[\varepsilon^2 ZZ'](E_n[Z_iX_i'])^{-1}$. We want to show that, given $\hat{\Sigma} \to_p \Sigma$, then $\sqrt{n}\hat{\Sigma}^{-1}(\hat{\beta} - \beta) \to_d N(0, 1)$. Notice that this is an application of Slutsky's Theorem, where $Y_n \to_p c_1$ could be replaced by $\hat{\Sigma} \to_p \Sigma$.

Normalize $\sqrt{n}(\hat{\beta}-\beta)$. Define $W_n := \sqrt{n}\hat{\Sigma}^{-1/2}(\hat{\beta}-\beta)$. We aim to show that $W_n \to_d N(0,1)$. Using Theorem 1.1 we know $\sqrt{n}(\hat{\beta}-\beta) \to_d N(0,\Sigma)$. Let's call $Z_n = \sqrt{n}(\hat{\beta}-\beta)$, so $Z_n \to_d N(0,\Sigma)$. Substituting into our definition of W_n, we have $W_n = \hat{\Sigma}^{-1/2}Z_n$. Since we have that $\hat{\Sigma} \to_p \Sigma$ and using the Continuous Mapping Theorem, we infer that $\hat{\Sigma}^{-1/2} \to_p \Sigma^{-1/2}$. Therefore, we can write $W_n = \Sigma^{-1/2}Z_n$. Using the properties of the normal distribution, we have:

$$W_n = \Sigma^{-1/2}Z_n \to_d N(0, \Sigma^{-1/2}\Sigma\Sigma^{-1/2}) = N(0,1)$$

Exercise 8

Suppose that $X_n \in \mathbb{R}$ converges in distribution to $Z \in \mathbb{R}$ conditional on Y_n. Show that $X_n \to_d Z$ unconditionally as well. (Hint: by the law of iterated expectations $Pr(X_n \leq x) = \mathbb{E}[Pr(X_n < x|Y_n)]$)

Conditional convergence in distribution from $X_n \to_d Z$ conditional on Y_n means that for every $x \in \mathbb{R}$:

$$\mathbb{P}(X_n \leq x|Y_n) \to_p \mathbb{P}(Z \leq n)$$

Let x be any point at which the CDF $F_Z(x)$ of Z is continuous. By the law of iterated expectations, we have

$$P(X_n \leq x) = \mathbb{E}\Big[P(X_n \leq x \mid Y_n)\Big]$$

Since X_n converges in distribution to Z conditional on Y_n, it follows that

$$P(X_n \leq x \mid Y_n) \to F_Z(x)$$

Note that for all n and x,

$$0 \leq P(X_n \leq x \mid Y_n) \leq 1$$

Because the conditional probabilities are bounded, we can apply the bounded convergence theorem to interchange the limit and the expectation. Thus, we obtain

$$\lim_{n\to\infty} P(X_n \leq x) = \lim_{n\to\infty} \mathbb{E}\Big[P(X_n \leq x \mid Y_n)\Big] = \mathbb{E}\Big[\lim_{n\to\infty} P(X_n \leq x \mid Y_n)\Big] = \mathbb{E}\Big[F_Z(x)\Big]$$

Since $F_Z(x)$ is constant with respect to Y_n, $\mathbb{E}\Big[F_Z(x)\Big] = F_Z(x)$. Therefore, we have shown that for every continuity point x of F_Z,

$$\lim_{n\to\infty} P(X_n \leq x) = F_Z(x)$$

This is exactly the definition of convergence in the distribution of X_n to Z. Thus $X_n \to_d Z$

unconditionally.

Exercise 9

Assumption 3.1 *(Sampling): $\{Y_i, X_i, Z_i\}_{i=1}^n$ are collected i.i.d. underlying random variables (Y, X, Z) satisfying (1.1) for some $\beta \in \mathbb{R}^d$.*

Assumption 3.2 *(Weighting matrix): (i) The estimated weighting matrix $\hat{\Omega}$ satisfies $\hat{\Omega} \to_p \Omega$ for some symmetric and positive definite matrix Ω and (ii) The rank of the $d_x \times d_z$ matrix $\mathbb{E}[XZ']$ is equal to d_x.*

Assumption 3.3 *(Moments): (i) The moment $\mathbb{E}[||XZ'||] < \infty$ and (ii) the eigenvalues of $\mathbb{E}[\varepsilon^2 ZZ']$ are bounded from above and away from zero.*

Theorem 3.1: Suppose that Assumptions 3.1, 3.2, and 3.3(i) hold. Then $\hat{\beta} \to_p \beta$. If additionally Assumptions 3.3(ii) holds then

$$\sqrt{n}(\hat{\beta} - \beta) \to_d N(0, \Sigma(\Omega))$$

where $\Sigma(\Omega) = (\mathbb{E}[XZ']\Omega\mathbb{E}[ZX'])^{-1}\mathbb{E}[XZ']\Omega\mathbb{E}[\varepsilon^2 ZZ']\Omega\mathbb{E}[ZX'](\mathbb{E}[XZ']\Omega\mathbb{E}[ZX'])^{-1}$.

Describe where each assumption is used to obtain the result in Theorem 3.1.

These assumptions are very important for the derivation of consistency and asymptotic normality. In particular, we can assign the importance of each assumption to each part of the Theorem:

- *Consistency* ($\hat{\beta} \to_p \beta$) is achieved using Assumption 3.1 (i.i.d. sampling) to invoke the LLN and Assumption 3.3(i) (finite moments), along with the identification provided by Assumption 3.2(ii).
- *Asymptotic Normality* ($\sqrt{n}(\hat{\beta} - \beta) \to_d N(0, \Sigma(\Omega))$) is derived using the CLT (enabled by Assumption 3.1), the convergence of the weighting matrix in Assumption 3.2(i), and the bounded, non-degenerate variance condition in Assumption 3.3(ii).

Let's describe one-by-one to describe better how these assumptions are required in the Theorem.

- **Assumption 3.1 (Sampling)**

  The i.i.d. sampling assumption is essential for the following reasons:

  - *Law of Large Numbers (LLN):* It ensures that sample averages converge to their population counterparts. For example, the sample moment conditions converge to the corresponding expectations, which is used to establish the consistency $\hat{\beta} \to_p \beta$.

- *Central Limit Theorem (CLT):* It justifies the asymptotic normality of the estimator, i.e., the result that $\sqrt{n}(\hat{\beta} - \beta)$ converges in distribution to a normal distribution.

- **Assumption 3.2 (Weighting Matrix)**

  This assumption serves two key purposes:

  - *Convergence of the Weighting Matrix:* Part (i) ensures that the estimated weighting matrix $\hat{\Omega}$ converges in probability to a fixed matrix Ω. This stability is necessary for the sample criterion function to have a well-behaved limit.
  - *Identification:* Part (ii) requires that the matrix $\mathbb{E}[XZ']$ has full column rank (rank equal to d_x). This condition is crucial to guarantee that the moment condition uniquely identifies the parameter β. It also ensures that the matrix $\mathbb{E}[XZ']\Omega\mathbb{E}[ZX']$ is invertible, which is needed in the expression for the asymptotic variance $\Sigma(\Omega)$.

- **Assumption 3.3 (Moments)**

  The moment conditions play a critical role in both consistency and asymptotic normality:

  - *Finiteness of Moments (Part (i)):* The condition $\mathbb{E}[||XZ'||] < \infty$ is required to apply the LLN, ensuring that the sample moments converge to their expectations and thus yielding $\hat{\beta} \to_p \beta$.
  - *Non-Degeneracy of the Variance (Part (ii)):* By bounding the eigenvalues of $\mathbb{E}[\varepsilon^2 ZZ']$ away from zero and above, this condition prevents the asymptotic variance from degenerating or exploding. It is essential for obtaining a valid asymptotic normal distribution for $\sqrt{n}(\hat{\beta} - \beta)$ and for the expression of $\Sigma(\Omega)$.

Exercise 10

Definition 3.1 (PSD Ordering): We say a matrix $\Sigma_1 \in \mathbb{R}^{d\times d}$ is weakly smaller than a matrix $\Sigma_2 \in \mathbb{R}^{d\times d}$ denoted $\Sigma_1 \leq \Sigma_2$, in the positive semidefinite ordering if the matrix $\Sigma_2 - \Sigma_1$ is positive semidefinite. That is, for any $v \in \mathbb{R}^d$, $v'(\Sigma_2 - \Sigma_1)v \geq 0$. If the matrix $\Sigma_2 - \Sigma_1$ is strictly positive definite, then we say Σ_1 is strictly smaller than Σ_2 in the positive semidefinite ordering and write $\Sigma_1 < \Sigma_2$.

Verify that this ordering satisfies transitivity. i.e. if $\Sigma_3 \geq \Sigma_2$ and $\Sigma_2 \geq \Sigma_1$ then $\Sigma_3 \geq \Sigma_1$.

PSD Ordering: $\Sigma_i \leq \Sigma_j$ means $\Sigma_i - \Sigma_j$ is positive semidefinite, meaning for any $v \in \mathbb{R}^d$, $v'(\Sigma_j - \Sigma_i)v \geq 0$.

Assume that $\Sigma_1 \leq \Sigma_2$ and $\Sigma_2 \leq \Sigma_3$, we want to show that $\Sigma_1 \leq \Sigma_3$. From assumptions, we have:

$$\Sigma_1 \leq \Sigma_2 \implies \Sigma_2 - \Sigma_1 \text{ is PSD}$$

$$\Sigma_2 \leq \Sigma_3 \implies \Sigma_3 - \Sigma_2 \text{ is PSD}$$

Therefore, for any $v > 0$ we have:

$$v'(\Sigma_2 - \Sigma_1)v \geq 0$$

$$v'(\Sigma_3 - \Sigma_2)v \geq 0$$

We can add both inequalities and obtain:

$$v'(\Sigma_2 - \Sigma_1)v + v'(\Sigma_3 - \Sigma_2)v \geq 0$$

$$v'(\Sigma_3 + \Sigma_2 - \Sigma_1 - \Sigma_2)v \geq 0$$

$$v'(\Sigma_3 - \Sigma_1)v \geq 0$$

Notice that the last equation implies that $\Sigma_3 - \Sigma_1$ is positive semidefinite, which implies $\Sigma_1 \leq \Sigma_3$, and therefore this ordering **is Transitive**.

Exercise 11

Repeat this analysis to show that $\sigma_1 < \sigma_2$ leads to a locally uniformly more powerful test for the null alternate pair $H_o : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$.

Consider the inference for the null hypothesis:

$$H_0 : \theta \geq \theta_0 \quad vs \quad H_a : \theta < \theta_0$$

The standard level α Wald test based on the estimators $\hat{\theta}^1$ and $\hat{\theta}^2$, is defined as in the lecture notes:

$$T_n^{(j)} = \frac{\hat{\theta}^{(j)} - \theta_0}{\sigma_j/\sqrt{n}}$$

It will reject the null hypothesis whenever $T_n^{(j)} < c_\alpha$. To examine the power properties of this test, we consider the power under $\sqrt{n}$ deviations from the null. So, consider the sequence $\theta_n := \theta - t/\sqrt{n}$ for $t > 0$ again. Then, the local power of the test is computed as the probability of rejecting H_0 given that it is false. So, we compute:

$$\begin{aligned}\mathbb{P}_{\theta_n}(T_n^{(j)} < c_\alpha) =& \mathbb{P}_{\theta_n}\left(\frac{\hat{\theta}^{(j)} - \theta_0}{\sigma_j/\sqrt{n}} < c_\alpha\right) \\ =& \mathbb{P}_{\theta_n}\left(\frac{\hat{\theta}^{(j)} - \theta_n}{\sigma_j/\sqrt{n}} + \frac{\hat{\theta}_n - \theta_0}{\sigma_j/\sqrt{n}} < c_\alpha\right)\end{aligned}$$

Since $\frac{\hat{\theta}^{(j)}-\theta_n}{\sigma_j/\sqrt{n}} \to_d N(0,1)$ and $\frac{\hat{\theta}_n-\theta_0}{\sigma_j/\sqrt{n}} = -\frac{t}{\sigma_j}$, we obtain:

$$\begin{aligned}\mathbb{P}_{\theta_n}(T_n^{(j)} < c_\alpha) =& \mathbb{P}_{\theta_n}\left(Z - \frac{t}{\sigma_j} < c_\alpha\right) = \mathbb{P}_{\theta_n}\left(Z < c_\alpha + \frac{t}{\sigma_j}\right) \\ =& \Phi\left(c_\alpha + \frac{t}{\sigma_j}\right)\end{aligned}$$

Since $\Phi(\cdot)$ is an increasing function, the higher the argument, the higher the power of the test. Notice that it is decreasing in σ_j. Thus, since $\sigma_1 < \sigma_2$, $t/\sigma_1 > t/\sigma_2$. Therefore, the power is higher for the tests based on $\theta^{(1)}$ than $\theta^{(2)}$.