

Lecture 8: Asymptotic theory

Economics 527: Econometric Methods

Vadim Marmer, UBC

Convergence in probability

OLS as a function of sample averages

- Model and OLS estimator:

$$Y_i = X_i^\top \beta + U_i, \quad \hat{\beta} = \left(\sum_{i=1}^n X_i X_i^\top \right)^{-1} \sum_{i=1}^n X_i Y_i$$

- Normal errors, exact distribution given the regressors:

$$U_i \mid X_i \sim N(0, \sigma^2) \\ \Rightarrow \hat{\beta} \mid X_1, \dots, X_n \sim N\left(\beta, \sigma^2 \left(\sum_{i=1}^n X_i X_i^\top \right)^{-1}\right)$$

- Without normality, write OLS in terms of sample averages:

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^n X_i X_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n X_i Y_i$$

- Use approximations; all limits below are as $n \rightarrow \infty$.

Limit of a non-random sequence

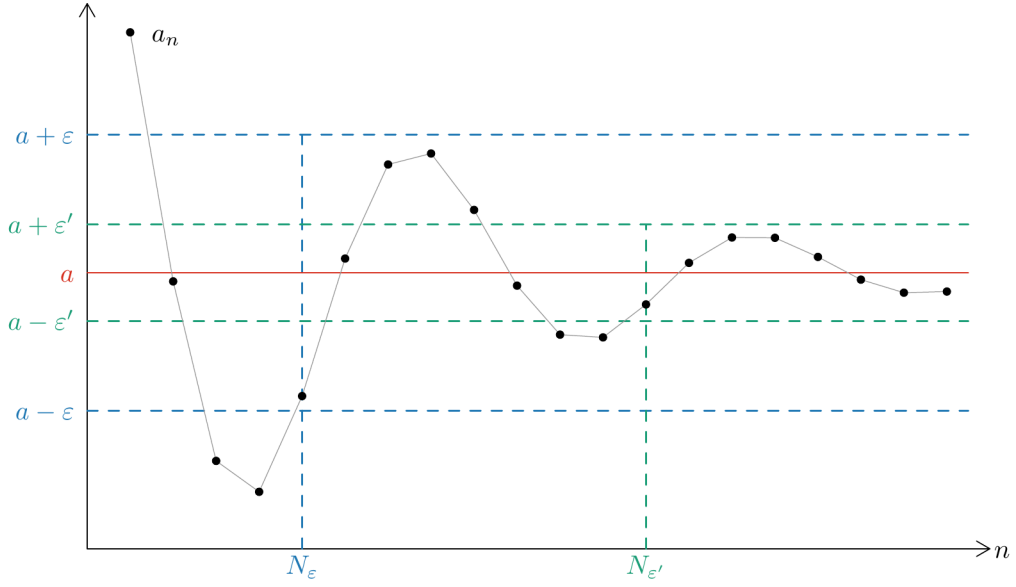
- Non-random sequence:

$$\{a_n \in \mathbb{R} : n \geq 1\}$$

- Definition.** a_n converges to a , $a_n \rightarrow a$, $a = \lim a_n$, if for all $\varepsilon > 0$ there is $N_\varepsilon \in \mathbb{N}$:

$$|a_n - a| < \varepsilon \quad \text{for all } n \geq N_\varepsilon$$

- a : good approximation of a_n for all $n \geq N_\varepsilon$.
- Sequence a_n with bands $a \pm \varepsilon$ and $a \pm \varepsilon'$, $\varepsilon' < \varepsilon$:



Estimator as a random sequence

- $\hat{\theta}_n$: estimator of θ from n observations; random sequence:

$$\{\hat{\theta}_n : n \geq 1\}, \quad \hat{\theta}_n = \hat{\theta}_n(\omega), \quad \omega \in \Omega$$

- OLS from n observations: $\hat{\beta}_n = \hat{\beta}_n(\omega)$, random.
- Non-random sequences from $\hat{\theta}_n$; apply the definition of a limit to them:

$$\text{Moments, } r > 0 : \quad a_n = \mathbb{E} [|\hat{\theta}_n - \theta|^r]$$

$$\text{Probabilities, } \delta > 0 : \quad a_n = \mathbb{P}(|\hat{\theta}_n - \theta| > \delta)$$

Convergence in r -th mean

- **Definition.** $\hat{\theta}_n$ converges to θ in r -th mean, $r > 0$, if:

$$\mathbb{E} [|\hat{\theta}_n - \theta|^r] \rightarrow 0$$

- $a_n = \mathbb{E} [|\hat{\theta}_n - \theta|^r] \rightarrow 0$: for all $\varepsilon > 0$ there is $N_\varepsilon \in \mathbb{N}$:

$$\mathbb{E} [|\hat{\theta}_n - \theta|^r] < \varepsilon \quad \text{for all } n \geq N_\varepsilon$$

- $r = 2$, mean squared error (MSE):

$$\mathbb{E} [(\hat{\theta}_n - \theta)^2]$$

Definition of convergence in probability

- Non-random sequence, $\delta > 0$ fixed:

$$a_n = \mathbb{P}(|\hat{\theta}_n - \theta| > \delta), \quad n \geq 1$$

- **Definition.** $\hat{\theta}_n$ converges in probability to θ , $\hat{\theta}_n \rightarrow_p \theta$, if for all $\delta > 0$:

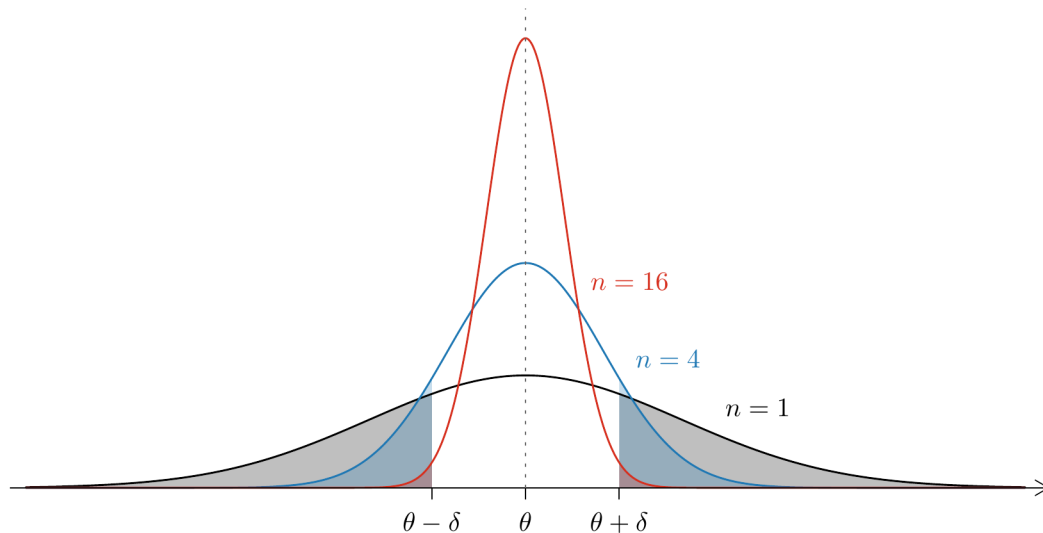
$$\mathbb{P}(|\hat{\theta}_n - \theta| > \delta) \rightarrow 0$$

- Equivalently, for all $\delta > 0$:

$$\begin{aligned} \mathbb{P}(|\hat{\theta}_n - \theta| \leq \delta) &= 1 - \underbrace{\mathbb{P}(|\hat{\theta}_n - \theta| > \delta)}_{\rightarrow 0} \\ &\rightarrow 1 \end{aligned}$$

Densities of $\hat{\theta}_n$ as n grows

- Example: densities for $\hat{\theta}_n \sim N(\theta, 1/n)$.
- Shaded: $\mathbb{P}(|\hat{\theta}_n - \theta| > \delta)$.



- $\hat{\theta}_n \rightarrow_p \theta$: distribution of $\hat{\theta}_n$ more concentrated around θ .

Markov's inequality

- Markov's inequality connects tail probabilities to expectations; the core device for establishing convergence in probability.
- **Theorem.** X : random variable, $M > 0$:

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

- Expectation of an indicator, X with density f_X , split at $-M$ and M :

$$\begin{aligned}
\mathbb{E}[\mathbf{1}(|X| \geq M)] &= \int_{-\infty}^{\infty} \mathbf{1}(|x| \geq M) f_X(x) dx \\
&= \int_{-\infty}^{-M} \underbrace{\mathbf{1}(|x| \geq M)}_{=1} f_X(x) dx \\
&\quad + \overbrace{\int_{-M}^M \mathbf{1}(|x| \geq M) f_X(x) dx}^{=0} \\
&\quad + \int_M^{\infty} \underbrace{\mathbf{1}(|x| \geq M)}_{=1} f_X(x) dx \\
&= \int_{-\infty}^{-M} f_X(x) dx + \int_M^{\infty} f_X(x) dx \\
&= \mathbb{P}(X \leq -M) + \mathbb{P}(X \geq M) \\
&= \mathbb{P}(|X| \geq M)
\end{aligned}$$

- Split 1 into two indicators:

$$1 = \mathbf{1}(|X| \geq M) + \mathbf{1}(|X| < M)$$

- Expectation of $|X|$:

$$\begin{aligned}
\mathbb{E}[|X|] &= \mathbb{E}[|X| \cdot 1] \\
&= \mathbb{E}[|X| (\mathbf{1}(|X| \geq M) + \mathbf{1}(|X| < M))] \\
&= \mathbb{E}[|X| \mathbf{1}(|X| \geq M)] + \underbrace{\mathbb{E}[|X| \mathbf{1}(|X| < M)]}_{\geq 0} \\
&\geq \mathbb{E}[|X| \mathbf{1}(|X| \geq M)] \\
&\geq \mathbb{E}[M \mathbf{1}(|X| \geq M)] \quad \text{where } \mathbf{1}(|X| \geq M) = 1 \text{ when } |X| \geq M \\
&= M \mathbb{E}[\mathbf{1}(|X| \geq M)] \\
&= M \mathbb{P}(|X| \geq M)
\end{aligned}$$

- Divide by M :

$$\boxed{\mathbb{P}(|X| \geq M) \leq \frac{\mathbb{E}[|X|]}{M}}$$

Convergence in r -th mean and in probability

- **Claim.**

$$\mathbb{E}[|\hat{\theta}_n - \theta|^r] \rightarrow 0 \text{ for some } r > 0 \implies \hat{\theta}_n \rightarrow_p \theta$$

- Markov's inequality, for all $\delta > 0$:

$$\begin{aligned}
\mathbb{P}(|\hat{\theta}_n - \theta| > \delta) &= \mathbb{P}(|\hat{\theta}_n - \theta|^r > \delta^r) \\
&\leq \mathbb{P}(\underbrace{|\hat{\theta}_n - \theta|^r}_X \geq \underbrace{\delta^r}_M) \\
&\leq \frac{1}{\delta^r} \underbrace{\mathbb{E}[|\hat{\theta}_n - \theta|^r]}_{\rightarrow 0} \\
&\rightarrow 0
\end{aligned}$$

- Therefore:

$$\boxed{\hat{\theta}_n \rightarrow_p \theta}$$

Chebyshev's inequality

- As on the previous slide, $r = 2$, $Y - E[Y]$ in place of $\hat{\theta}_n - \theta$, $\delta > 0$:

$$\begin{aligned} P(|Y - E[Y]| > \delta) &\leq \frac{1}{\delta^2} E[|Y - E[Y]|^2] \\ &= \frac{1}{\delta^2} \text{Var}(Y) \end{aligned}$$

Bias and variance tending to zero

- **Claim.** For an estimator $\hat{\theta}_n$ of θ :

$$\begin{aligned} 1. & E[\hat{\theta}_n] \rightarrow \theta \\ 2. & \text{Var}(\hat{\theta}_n) \rightarrow 0 \\ \implies & \hat{\theta}_n \rightarrow_p \theta \end{aligned}$$

- Recall:

$$\underbrace{E[(\hat{\theta}_n - \theta)^2]}_{\text{MSE}} = \text{Var}(\hat{\theta}_n) + \underbrace{(E[\hat{\theta}_n] - \theta)^2}_{\text{Bias}^2}$$

- Markov with $r = 2$, for all $\delta > 0$:

$$\begin{aligned} P(|\hat{\theta}_n - \theta| > \delta) &\leq \frac{1}{\delta^2} E[(\hat{\theta}_n - \theta)^2] \\ &= \frac{1}{\delta^2} \left(\underbrace{\text{Var}(\hat{\theta}_n)}_{\rightarrow 0 \text{ by 2.}} + \underbrace{(E[\hat{\theta}_n] - \theta)^2}_{\rightarrow 0 \text{ by 1.}} \right) \\ &\rightarrow 0 \end{aligned}$$

- Therefore:

$$\boxed{\hat{\theta}_n \rightarrow_p \theta}$$

Law of large numbers (LLN)

LLN for uncorrelated observations

- **Theorem.** $X_1, \dots, X_n$: random variables; for all i and all $j \neq i$:

$$\begin{aligned} E[X_i] &= \mu, \quad \text{Var}(X_i) = \sigma^2 < \infty, \quad \text{Cov}(X_i, X_j) = 0 \\ \implies \bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_i \rightarrow_p \mu \end{aligned}$$

Proof of the LLN, step 1: mean and variance

- $E[\bar{X}_n] = \mu$, so bias is zero:

$$\begin{aligned} E[\bar{X}_n] &= \frac{1}{n} \sum_{i=1}^n E[X_i] \\ &= \mu \end{aligned}$$

- Variance of $\bar{X}_n$:

$$\begin{aligned}
\text{Var}(\bar{X}_n) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\
&= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) \\
&= \frac{1}{n^2} \left(\sum_{i=1}^n \text{Var}(X_i) + \underbrace{\sum_{i=1}^n \sum_{j \neq i} \text{Cov}(X_i, X_j)}_{=0} \right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \underbrace{\text{Var}(X_i)}_{=\sigma^2} \\
&= \frac{1}{n^2} (\sigma^2 \cdot n) \\
&= \frac{\sigma^2}{n} \\
&\rightarrow 0
\end{aligned}$$

Proof of the LLN, step 2: Chebyshev's inequality

- Chebyshev with $E[\bar{X}_n] = \mu$, for all $\delta > 0$:

$$\begin{aligned}
P(|\bar{X}_n - \mu| > \delta) &\leq \frac{1}{\delta^2} \text{Var}(\bar{X}_n - \mu) \\
&= \frac{1}{\delta^2} \text{Var}(\bar{X}_n) \\
&= \frac{1}{\delta^2} \frac{\sigma^2}{n} \\
&\rightarrow 0
\end{aligned}$$

- Therefore:

$$\boxed{\bar{X}_n \rightarrow_p \mu = E[X_i]}$$

Extension to different means

- $E[X_i] = \mu_i$, not constant; $\text{Var}(X_i) = \sigma^2 < \infty$; $\text{Cov}(X_i, X_j) = 0$ for $i \neq j$.
- Centre each X_i at its mean:

$$\begin{aligned}
Y_i &= X_i - E[X_i] \\
\Rightarrow E[Y_i] &= 0, \quad \text{Var}(Y_i) = \sigma^2, \quad \text{Cov}(Y_i, Y_j) = 0 \text{ for } i \neq j
\end{aligned}$$

- LLN for $Y_1, \dots, Y_n$:

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E[X_i] &= \frac{1}{n} \sum_{i=1}^n (X_i - E[X_i]) \\
&= \frac{1}{n} \sum_{i=1}^n Y_i \\
&\rightarrow_p 0
\end{aligned}$$

Extension to different variances

- $Y_1, \dots, Y_n$: random variables; for all i and all $j \neq i$:

$$\begin{aligned} E[Y_i] &= 0, & \text{Var}(Y_i) &= E[Y_i^2] = \sigma_i^2 \text{ not constant} \\ \text{Cov}(Y_i, Y_j) &= 0 \end{aligned}$$

- Bounded variances, $\bar{\sigma}^2$ not depending on n :

$$\sigma_i^2 \leq \max_{j=1, \dots, n} \sigma_j^2 \leq \bar{\sigma}^2$$

- Variance of the average $\rightarrow 0 \implies \frac{1}{n} \sum_{i=1}^n Y_i \rightarrow_p 0$, by Chebyshev:

$$\begin{aligned} \text{Var}\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) &= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n Y_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(Y_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \\ &\leq \frac{1}{n^2} n \bar{\sigma}^2 \\ &= \frac{\bar{\sigma}^2}{n} \\ &\rightarrow 0 \end{aligned}$$

Extension to dependent observations

- Variance of the average, non-zero covariances:

$$\begin{aligned} \text{Var}\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(Y_i, Y_j) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(Y_i) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} \text{Cov}(Y_i, Y_j) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(Y_i) + 2 \frac{1}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}(Y_i, Y_j) \end{aligned}$$

- Number of distinct covariances, order n^2 :

$$\frac{n^2 - n}{2} = \frac{n(n-1)}{2}$$

- Triangle inequality, $|a + b| \leq |a| + |b|$:

$$\begin{aligned} \left| \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} \text{Cov}(Y_i, Y_j) \right| &\leq \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} \underbrace{|\text{Cov}(Y_i, Y_j)|}_{\text{if of smaller order than } n^2} \\ &\rightarrow 0 \end{aligned}$$

- Double sum smaller order than n^2 : for example, each observation correlated with only a fixed number of others, with bounded covariances.

Example: $Y_i = \varepsilon_i + \varepsilon_{i+1}$

- $\varepsilon_1, \varepsilon_2, \dots$: iid (independent and identically distributed), $0 < \text{Var}(\varepsilon_i) < \infty$.
- Neighbours share one ε :

$$\begin{aligned} Y_i &= \varepsilon_i + \varepsilon_{i+1} \\ Y_{i+1} &= \varepsilon_{i+1} + \varepsilon_{i+2} \\ Y_{i+2} &= \varepsilon_{i+2} + \varepsilon_{i+3} \end{aligned}$$

- Covariances of Y_i and Y_{i+h} :

$$\text{Cov}(Y_i, Y_{i+1}) \neq 0, \quad \text{Cov}(Y_i, Y_{i+h}) = 0 \quad \text{for } h \geq 2$$

Weak LLN for iid observations

- Independence instead of uncorrelatedness: finite variance not required.
- **Theorem.** $X_1, \dots, X_n$: iid random variables:

$$\begin{aligned} & \mathbb{E}[|X_1|] < \infty \\ \Rightarrow \quad \bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_i \rightarrow_p \mathbb{E}[X_1] \\ \mathbb{E}[X_1] &= \int_{-\infty}^{\infty} x f_X(x) dx \quad \text{with density } f_X \end{aligned}$$

- Stated without proof; this is the LLN we will use.

Vectors and Slutsky's theorem

Convergence in probability of a random vector

- Estimator $\hat{\theta}_n$ and parameter θ , $k \times 1$:

$$\hat{\theta}_n = \begin{pmatrix} \hat{\theta}_{n,1} \\ \vdots \\ \hat{\theta}_{n,k} \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$$

- **Definition.** $\hat{\theta}_n$ converges in probability to θ , $\hat{\theta}_n \rightarrow_p \theta$, two versions:

1. Element by element: $\hat{\theta}_{n,j} \rightarrow_p \theta_j$ for all $j = 1, \dots, k$
2. In norm: $\|\hat{\theta}_n - \theta\| \rightarrow_p 0$, $\|x\| = \sqrt{x_1^2 + \dots + x_k^2}$

- **Claim.** $1 \Leftrightarrow 2$.

$$\|\hat{\theta}_n - \theta\|^2 = (\hat{\theta}_{n,1} - \theta_1)^2 + \dots + (\hat{\theta}_{n,k} - \theta_k)^2$$

- $1 \Rightarrow 2$: each element converges to zero, so the finite sum converges to zero.
- $2 \Rightarrow 1$: non-negative summands; the sum converges to zero only if each summand does.

Sum of two sequences

- **Claim.** For 1×1 sequences $\hat{\theta}_{n,1}$ and $\hat{\theta}_{n,2}$:

$$\begin{aligned} \hat{\theta}_{n,1} &\rightarrow_p \theta_1, \quad \hat{\theta}_{n,2} \rightarrow_p \theta_2 \\ \implies \hat{\theta}_{n,1} + \hat{\theta}_{n,2} &\rightarrow_p \theta_1 + \theta_2 \end{aligned}$$

- Triangle inequality, then a union bound, for all $\delta > 0$:

$$\begin{aligned} &P(|\hat{\theta}_{n,1} + \hat{\theta}_{n,2} - (\theta_1 + \theta_2)| > \delta) \\ &\leq P(|\hat{\theta}_{n,1} - \theta_1| + |\hat{\theta}_{n,2} - \theta_2| > \delta) \\ &\leq P\left(|\hat{\theta}_{n,1} - \theta_1| > \frac{\delta}{2} \text{ or } |\hat{\theta}_{n,2} - \theta_2| > \frac{\delta}{2}\right) \\ &\leq P\left(|\hat{\theta}_{n,1} - \theta_1| > \frac{\delta}{2}\right) + P\left(|\hat{\theta}_{n,2} - \theta_2| > \frac{\delta}{2}\right) \\ &\rightarrow 0 \end{aligned}$$

- Therefore:

$$\boxed{\hat{\theta}_{n,1} + \hat{\theta}_{n,2} \rightarrow_p \theta_1 + \theta_2}$$

Slutsky's theorem

- More general than sums: convergence is preserved by continuous functions.
- **Theorem.** $\hat{\theta}_n$: $k \times 1$; $\theta \in \mathbb{R}^k$ non-random; $h: \mathbb{R}^k \rightarrow \mathbb{R}^m$ continuous at θ ; $\rightarrow_p$ in norm:

$$\hat{\theta}_n \rightarrow_p \theta \implies h(\hat{\theta}_n) \rightarrow_p h(\theta)$$

- h continuous at θ : for all $\delta > 0$ there is $\eta_\delta > 0$:

$$\|\hat{\theta}_n - \theta\| \leq \eta_\delta \implies \|h(\hat{\theta}_n) - h(\theta)\| \leq \delta$$

- Inclusion of events, then probabilities, for all $\delta > 0$:

$$\begin{aligned} &\{\|\hat{\theta}_n - \theta\| \leq \eta_\delta\} \subset \{\|h(\hat{\theta}_n) - h(\theta)\| \leq \delta\} \\ \implies &\underbrace{P(\|\hat{\theta}_n - \theta\| \leq \eta_\delta)}_{\rightarrow 1 \text{ (in norm)}} \leq P(\|h(\hat{\theta}_n) - h(\theta)\| \leq \delta) \leq 1 \\ \implies &P(\|h(\hat{\theta}_n) - h(\theta)\| \leq \delta) \rightarrow 1 \\ \implies &\boxed{h(\hat{\theta}_n) \rightarrow_p h(\theta)} \end{aligned}$$

Examples: product, ratio

- $\hat{\theta}_{n,1} \rightarrow_p \theta_1$, $\hat{\theta}_{n,2} \rightarrow_p \theta_2$; element by element $\implies$ in norm; Slutsky for product and ratio:

$$\begin{aligned} \hat{\theta}_{n,1} \hat{\theta}_{n,2} &\rightarrow_p \theta_1 \theta_2 \\ \frac{\hat{\theta}_{n,1}}{\hat{\theta}_{n,2}} &\rightarrow_p \frac{\theta_1}{\theta_2} \text{ if } \theta_2 \neq 0 \end{aligned}$$

- $h(\theta_1, \theta_2) = \theta_1/\theta_2$: not continuous at $\theta_2 = 0$.
- $X_1, \dots, X_n$ iid, $E[|X_1|] < \infty$; weak LLN, then Slutsky for $h(u) = u^2$ and $h(u) = 1/u$:

$$\begin{aligned} (\bar{X}_n)^2 &\rightarrow_p (E[X_1])^2 \\ \frac{1}{\bar{X}_n} &\rightarrow_p \frac{1}{E[X_1]} \text{ if } E[X_1] \neq 0 \end{aligned}$$

Example: inverse of a matrix

- X_i : $k \times 1$, iid, $E[X_i X_i^\top]$ finite; weak LLN, element by element:

$$\frac{1}{n} \sum_{i=1}^n X_i X_i^\top \rightarrow_p E[X_i X_i^\top]$$

- $E[X_i X_i^\top]$: symmetric, p.s.d.
- Assume $E[X_i X_i^\top]$ p.d.; $h(\cdot) = (\cdot)^{-1}$: continuous at a p.d. matrix.
- Element by element $\Rightarrow$ in norm, then Slutsky on the k^2 elements:

$$\left(\frac{1}{n} \sum_{i=1}^n X_i X_i^\top \right)^{-1} \rightarrow_p (E[X_i X_i^\top])^{-1}$$

Convergence in distribution

Normal distributions and Φ

- Standard normal CDF:

$$\Phi(t) = P(Z \leq t), \quad Z \sim N(0, 1)$$

- For $X \sim N(\mu, \sigma^2)$, $\sigma > 0$:

$$\begin{aligned} P(X \leq t) &= P(X - \mu \leq t - \mu) \\ &= P\left(\underbrace{\frac{X - \mu}{\sigma}}_{= Z \sim N(0,1)} \leq \frac{t - \mu}{\sigma} \right) \\ &= \Phi\left(\frac{t - \mu}{\sigma}\right) \end{aligned}$$

- For any X , $\mu \in \mathbb{R}$ and $\sigma > 0$:

$$P(X \leq t) = \Phi\left(\frac{t - \mu}{\sigma}\right) \text{ for all } t \in \mathbb{R} \iff X \sim N(\mu, \sigma^2)$$

Asymptotic normality

- $\hat{\theta}_n$, θ : 1×1 .
- **Definition.** $\hat{\theta}_n$ asymptotically normal: for some $\omega > 0$:

$$\begin{aligned} P(\sqrt{n}(\hat{\theta}_n - \theta) \leq t) &= P\left(\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\omega} \leq \frac{t}{\omega}\right) \\ &\rightarrow \Phi\left(\frac{t}{\omega}\right) \text{ for all } t \in \mathbb{R} \\ \iff \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\omega} &\rightarrow_d N(0, 1) \\ \iff \sqrt{n}(\hat{\theta}_n - \theta) &\rightarrow_d N(0, \omega^2) \end{aligned}$$

- Convergence in distribution concerns distribution functions, not convergence of the random variables themselves.
- Approximate using the normal CDF; valid as $n \rightarrow \infty$:

$$P(\sqrt{n}(\hat{\theta}_n - \theta) \leq t) \approx \Phi\left(\frac{t}{\omega}\right)$$

Why $\sqrt{n}$

- $X_1, \dots, X_n$ iid, mean θ , variance $0 < \sigma^2 < \infty$; $\hat{\theta}_n = \bar{X}_n$:

$$\hat{\theta}_n - \theta \rightarrow_p 0 \quad (\text{LLN})$$

- Variance of $\hat{\theta}_n - \theta$ (proof of the LLN, step 1):

$$\begin{aligned} \text{Var}(\hat{\theta}_n - \theta) &= \text{Var}(\bar{X}_n) \\ &= \frac{\sigma^2}{n} \\ &\rightarrow 0 \end{aligned}$$

- Scaled by $\sqrt{n}$:

$$\begin{aligned} \text{Var}(\sqrt{n}(\hat{\theta}_n - \theta)) &= n \text{Var}(\hat{\theta}_n - \theta) \\ &= n \cdot \frac{\sigma^2}{n} \\ &= \sigma^2 \end{aligned}$$

- Scaled by n^α :

$$\begin{aligned} \text{Var}(n^\alpha(\hat{\theta}_n - \theta)) &= \frac{n^{2\alpha}}{n} \sigma^2 \\ &= n^{2\alpha-1} \sigma^2 \\ &\rightarrow \begin{cases} \infty & \text{if } \alpha > \frac{1}{2} \\ \sigma^2 & \text{if } \alpha = \frac{1}{2} \\ 0 & \text{if } \alpha < \frac{1}{2} \end{cases} \end{aligned}$$

Central limit theorem (CLT)

Univariate CLT

- **Theorem.** $X_1, \dots, X_n$: iid random variables:

$$\begin{aligned} \mathbb{E}[X_i] &= 0, \quad \text{Var}(X_i) = \mathbb{E}[X_i^2] = \sigma^2, \quad 0 < \sigma^2 < \infty \\ \Rightarrow \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i &\rightarrow_d N(0, \sigma^2) \end{aligned}$$

- CDF form, for all $t \in \mathbb{R}$:

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \leq t\right) \rightarrow \Phi\left(\frac{t}{\sigma}\right)$$

- In terms of $\bar{X}_n$:

$$\sqrt{n} \bar{X}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow_d N(0, \sigma^2)$$

Non-zero mean

- $X_1, \dots, X_n$ iid, $\mathbb{E}[X_i] = \mu \neq 0$, $\text{Var}(X_i) = \sigma^2$, $0 < \sigma^2 < \infty$:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow_d N(\mu, \sigma^2)$$

- Add and subtract μ :

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) + \frac{1}{\sqrt{n}} n\mu \\ &= \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^n \underbrace{(X_i - \mu)}_{\substack{\text{E}[X_i - \mu] = 0 \\ \rightarrow_d N(0, \sigma^2)}}}_{\rightarrow_p 0} + \underbrace{\sqrt{n} \mu}_{\rightarrow \pm \infty}\end{aligned}$$

- CLT for $X_i - \mu$, for all $t \in \mathbb{R}$:

$$\begin{aligned}\mathbb{P}(\underbrace{\sqrt{n}(\bar{X}_n - \mu)}_{\rightarrow_p 0} \leq t) &= \mathbb{P}\left(\sqrt{n} \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \leq t\right) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \leq t\right) \\ &\rightarrow \Phi\left(\frac{t}{\sigma}\right)\end{aligned}$$

- Therefore:

$$\boxed{\sqrt{n}(\bar{X}_n - \mu) \rightarrow_d N(0, \sigma^2)}$$

Why $\sigma^2 > 0$

- Zero variance:

$$\begin{aligned}\mathbb{E}[X_i] &= 0, \quad \text{Var}(X_i) = 0 \\ \implies X_i &= 0 \quad \text{with probability 1} \\ \implies \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i &= 0 \not\rightarrow_d N(0, \omega^2) \quad \text{for any } \omega > 0\end{aligned}$$

Proof of the CLT, step 1: moment generating function (MGF)

- MGF of X , assumed finite for t near 0:

$$M(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

- Normal MGF:

$$X \sim N(\mu, \sigma^2) \iff \mathbb{E}[e^{tX}] = e^{\mu t + \sigma^2 t^2 / 2} \quad \text{for all } t \in \mathbb{R}$$

- One can show: MGFs converge for all $t \implies$ convergence in distribution.
- Need to show, for all $t \in \mathbb{R}$:

$$\mathbb{E}\left[e^{t \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i}\right] \rightarrow e^{\sigma^2 t^2 / 2}$$

Proof of the CLT, step 2: product of MGFs

- Factorise, with $M(t) = \mathbb{E}[e^{tX_i}]$:

$$\begin{aligned}
 \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} \sum_{i=1}^n X_i}\right] &= \mathbb{E}\left[e^{\sum_{i=1}^n \frac{t}{\sqrt{n}} X_i}\right] \\
 &= \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} X_1 + \dots + \frac{t}{\sqrt{n}} X_n}\right] \\
 &= \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} X_1} \times \dots \times e^{\frac{t}{\sqrt{n}} X_n}\right] \\
 &= \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} X_1}\right] \times \dots \times \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} X_n}\right] \quad (\text{independence}) \\
 &= \left(\mathbb{E}\left[e^{\frac{t}{\sqrt{n}} X_i}\right]\right)^n \quad (\text{identical distributions}) \\
 &= \left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n
 \end{aligned}$$

Proof of the CLT, step 3: expansion of $M(t/\sqrt{n})$

- Definition.** Small o :

$$a_n \rightarrow 0 \iff a_n = o(1), \quad o\left(\frac{t^2}{n}\right) = \underbrace{\frac{t^2}{n}}_{\rightarrow 0} o(1)$$

- M and its derivatives at 0:

$$\begin{aligned}
 M(0) &= \mathbb{E}[e^{tX_i}] \Big|_{t=0} = 1 \\
 M^{(1)}(t) &= \frac{d}{dt} \mathbb{E}[e^{tX_i}] \\
 &= \mathbb{E}\left[\frac{d}{dt} e^{tX_i}\right] \\
 &= \mathbb{E}[X_i e^{tX_i}] \\
 M^{(1)}(0) &= \mathbb{E}[X_i] = 0 \\
 M^{(2)}(t) &= \mathbb{E}\left[\frac{d^2}{dt^2} e^{tX_i}\right] \\
 &= \mathbb{E}[X_i^2 e^{tX_i}] \\
 M^{(2)}(0) &= \mathbb{E}[X_i^2] = \sigma^2
 \end{aligned}$$

- Second-order expansion around 0, then substitute the derivatives:

$$\begin{aligned}
 M\left(\frac{t}{\sqrt{n}}\right) &= M(0) + M^{(1)}(0) \frac{t}{\sqrt{n}} + \frac{M^{(2)}(0)}{2} \frac{t^2}{n} + o\left(\frac{t^2}{n}\right) \\
 &= 1 + 0 + \frac{\sigma^2}{2} \frac{t^2}{n} + \frac{t^2}{n} o(1) \\
 &= 1 + \frac{c_n}{n}, \quad c_n = \frac{\sigma^2 t^2}{2} + t^2 o(1) \rightarrow \frac{\sigma^2 t^2}{2}
 \end{aligned}$$

Proof of the CLT, step 4: limit of $(M(t/\sqrt{n}))^n$

- We will use, for $a_n \rightarrow \infty$:

$$\lim \left(1 + \frac{1}{a_n}\right)^{a_n} = e$$

- MGF of $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$, from steps 2 and 3, $t \neq 0$, $a_n = n/c_n \rightarrow \infty$:

$$\begin{aligned}
\mathbb{E} \left[e^{t \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i} \right] &= \left(M \left(\frac{t}{\sqrt{n}} \right) \right)^n \\
&= \left(1 + \frac{c_n}{n} \right)^n \\
&= \left(1 + \frac{1}{n/c_n} \right)^n \\
&= \left[\underbrace{\left(1 + \frac{1}{n/c_n} \right)^{n/c_n}}_{\rightarrow e} \right]^{c_n} \\
&\rightarrow e^{\sigma^2 t^2 / 2} \quad (c_n \rightarrow \sigma^2 t^2 / 2, \text{ step 3})
\end{aligned}$$

- $t = 0$: both sides equal 1.
- $e^{\sigma^2 t^2 / 2}$: MGF of $N(0, \sigma^2)$ (step 1).
- Therefore:

$$\boxed{\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow_d N(0, \sigma^2)}$$

Limit $(1 + 1/n)^n \rightarrow e$

- l'Hôpital's rule in $1/n$:

$$\begin{aligned}
\lim \frac{\ln(1 + 1/n)}{1/n} &= \lim \frac{1/(1 + 1/n)}{1} \\
&= 1
\end{aligned}$$

- Take logarithms:

$$\begin{aligned}
\ln \left[\left(1 + \frac{1}{n} \right)^n \right] &= n \ln \left(1 + \frac{1}{n} \right) \\
&= \frac{\ln(1 + 1/n)}{1/n} \\
&\rightarrow 1 \\
\Rightarrow \left(1 + \frac{1}{n} \right)^n &= e^{\ln[(1+1/n)^n]} \\
&\rightarrow e^1 = e
\end{aligned}$$

- Same steps, $1/a_n \rightarrow 0$ in place of $1/n$: $(1 + 1/a_n)^{a_n} \rightarrow e$.

Multivariate CLT

- **Theorem.** $X_1, \dots, X_n$: iid random $k \times 1$ vectors:

$$\begin{aligned}
\mathbb{E}[X_i] &= 0, \quad \Sigma_{k \times k} = \mathbb{E}[X_i X_i^\top] \text{ finite and p.d.} \\
\Rightarrow \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i &\rightarrow_d N(0, \Sigma)
\end{aligned}$$

- Recall: linear map of a normal vector; $a \in \mathbb{R}^k$, $a \neq 0$:

$$W_{k \times 1} \sim N(0, \Sigma) \quad \Rightarrow \quad a_{1 \times 1}^\top W \sim N(0, a_{1 \times 1}^\top \Sigma a)$$

- Equivalent statements (one can show):

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow_d N(0, \Sigma) \\
\iff & a^\top \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right) \rightarrow_d N(0, a^\top \Sigma a) \quad \text{for all } a \neq 0 \\
\iff & P\left(a^\top \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right) \leq t\right) \rightarrow \Phi\left(\frac{t}{\sqrt{a^\top \Sigma a}}\right) \quad \text{for all } a \neq 0, t \in \mathbb{R}
\end{aligned}$$

Proof of the multivariate CLT

- Σ p.d.:

$$a^\top \Sigma a > 0 \quad \text{for all } a \neq 0$$

- Linear combination with a :

$$a^\top \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \underbrace{a^\top X_i}_{1 \times 1}$$

- $a^\top X_1, \dots, a^\top X_n$: iid.
- Recall mean and variance with $\Gamma = a^\top$, $\text{Var}(X_i) = E[X_i X_i^\top] = \Sigma$:

$$\begin{aligned}
E[a^\top X_i] &= a^\top E[X_i] = 0 \\
\text{Var}(a^\top X_i) &= a^\top \text{Var}(X_i) a = a^\top \Sigma a
\end{aligned}$$

- Univariate CLT, $0 < a^\top \Sigma a < \infty$:

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n a^\top X_i \rightarrow_d N(0, a^\top \Sigma a) \quad \text{for all } a \neq 0 \\
\implies & \boxed{\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow_d N(0, \Sigma)} \quad (\text{equivalent statements})
\end{aligned}$$

Summary

Convergence in probability and the LLN

- $a_n \rightarrow a$: for all $\varepsilon > 0$, some N_ε with $|a_n - a| < \varepsilon$ for all $n \geq N_\varepsilon$.
- r -th mean, $r > 0$: $E[|\hat{\theta}_n - \theta|^r] \rightarrow 0$; $r = 2$: MSE.
- $\hat{\theta}_n \rightarrow_p \theta$: $P(|\hat{\theta}_n - \theta| > \delta) \rightarrow 0$ for all $\delta > 0$.
- Markov, $M > 0$: $P(|X| \geq M) \leq E[|X|]/M$; Chebyshev, $\delta > 0$: $P(|Y - E[Y]| > \delta) \leq \text{Var}(Y)/\delta^2$.
- r -th mean $\implies \rightarrow_p$; $E[\hat{\theta}_n] \rightarrow \theta$, $\text{Var}(\hat{\theta}_n) \rightarrow 0 \implies \hat{\theta}_n \rightarrow_p \theta$.
- LLN: uncorrelated, $E[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2 < \infty \implies \text{Var}(\bar{X}_n) = \sigma^2/n$, $\bar{X}_n \rightarrow_p \mu$.
- $Y_i = X_i - E[X_i]$, $\text{Var}(Y_i) \leq \bar{\sigma}^2$ not depending on n , $\sum_i \sum_{j \neq i} |\text{Cov}(Y_i, Y_j)|$ of smaller order than $n^2 \implies \frac{1}{n} \sum Y_i \rightarrow_p 0$.
- Weak LLN: iid, $E[|X_1|] < \infty \implies \bar{X}_n \rightarrow_p E[X_1]$.

Vectors and Slutsky's theorem

- Vectors: $\hat{\theta}_{n,j} \rightarrow_p \theta_j$ for all $j \iff \|\hat{\theta}_n - \theta\| \rightarrow_p 0$.
- Slutsky: $\hat{\theta}_n \rightarrow_p \theta$, h continuous at $\theta \implies h(\hat{\theta}_n) \rightarrow_p h(\theta)$.
- X_i iid, $E[X_i X_i^\top]$ finite and p.d. $\implies \left(\frac{1}{n} \sum X_i X_i^\top\right)^{-1} \rightarrow_p (E[X_i X_i^\top])^{-1}$.

Convergence in distribution and the CLT

- $\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_d N(0, \omega^2)$, $\omega > 0$: $P(\sqrt{n}(\hat{\theta}_n - \theta) \leq t) \rightarrow \Phi(t/\omega)$ for all t .
- X_i iid, mean θ , variance $0 < \sigma^2 < \infty \implies \text{Var}(n^\alpha(\bar{X}_n - \theta)) = n^{2\alpha-1}\sigma^2$; $\alpha = \frac{1}{2}$: variance σ^2 .
- CLT: iid, $E[X_i] = 0$, $\text{Var}(X_i) = \sigma^2$, $0 < \sigma^2 < \infty \implies \frac{1}{\sqrt{n}} \sum X_i \rightarrow_d N(0, \sigma^2)$.
- $E[X_i] = \mu$, other conditions the same $\implies \sqrt{n}(\bar{X}_n - \mu) \rightarrow_d N(0, \sigma^2)$.
- Proof of the CLT, $E[X_i] = 0$, MGF $M(t) = E[e^{tX_i}]$ finite near 0: $(M(t/\sqrt{n}))^n = (1 + c_n/n)^n \rightarrow e^{\sigma^2 t^2/2}$, $c_n \rightarrow \sigma^2 t^2/2$.
- Multivariate CLT: iid $k \times 1$, $E[X_i] = 0$, $\Sigma = E[X_i X_i^\top]$ finite and p.d. $\implies \frac{1}{\sqrt{n}} \sum X_i \rightarrow_d N(0, \Sigma)$.
- Equivalent statements: $\frac{1}{\sqrt{n}} \sum X_i \rightarrow_d N(0, \Sigma) \iff \frac{1}{\sqrt{n}} \sum a^\top X_i \rightarrow_d N(0, a^\top \Sigma a)$ for all $a \neq 0$.