Econ 6190: Econometrics I

Asymptotic Theory

Chen Qiu

Cornell Economics

2024 Fall

Motivation for asymptotic theory

- We derived the distribution of $\bar{X}_n$ under normal distribution assumption
- This can be quite restrictive
  - What happens when the population is not normal?
  - What is the distribution of nonlinear transformations of $\bar{X}_n$?
- Idea: Allow sample size n to grow to infinity and investigate the behavior of the estimators as this happens
  - Pros: provide useful approximations of the finite-sample case; simpler results
  - Cons: never realistic
- Main tools of asymptotic theory
  - Law of large numbers (LLN)
  - Central limit theorem (CLT)
  - Continuous mapping theorem (CMT)

Contents

Reference

- Hansen Ch. 7 and 8

1. Convergence in Probability

Asymptotic limits

- **Definition**: A sequence of numbers a_n has the **limit** a, or **converges** to a as $n \to \infty$ if for all $\delta > 0$, there exists some n_δ such that for all $n \geq n_\delta$, $|a_n - a| \leq \delta$
- Notations to indicate "a_n converges to a" include:

$$a_n \to a, \text{ as } n \to \infty; \text{ or } \lim_{n\to\infty} a_n = a$$

- Intuitively, a_n gets arbitrarily close to a as $n \to \infty$

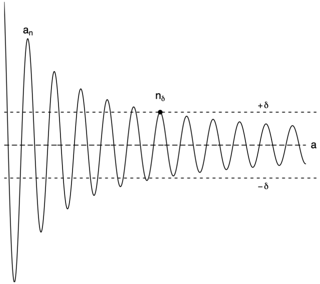

Figure: Limit of a sequence of numbers

Motivation for convergence in probability

- A (non-random) sequence may converge to a limit. What about a sequence of random variables?
- For example, $\bar{X}_n$ is a sequence of random variables indexed by sample size n
- As n changes, the distribution of $\bar{X}_n$ also changes
- In what sense does $\bar{X}_n$ converge when n becomes large?
- Since $\bar{X}_n$ is random, we need to modify definition of convergence and limit
- There are different ways to define convergence of sequence of random variables

Convergence in probability

- Let $\{X_n, n = 1, 2, \ldots\}$ be a sequence of random variables
- Let X be another random variable (X could be a constant)
- **Definition**: We say X_n **converges in probability** to X if for all $\delta > 0$

$$\lim_{n\to\infty} P\{|X_n - X| > \delta\} = 0$$

or equivalently

$$\lim_{n\to\infty} P\{|X_n - X| \leq \delta\} = 1$$

or equivalently, for all $\delta > 0$, $\varepsilon > 0$, there exists some $n_{\delta,\varepsilon}$ such that for all $n \geq n_{\delta,\varepsilon}$

$$P\{|X_n - X| > \delta\} < \varepsilon$$

i.e.

$$P\{|X_n - X| \leq \delta\} \geq 1 - \varepsilon$$

- Notations to indicate convergence in probability include

$$X_n \xrightarrow{p} X, \quad \text{plim} X_n = X, \quad X_n = X + o_p(1)$$

Example

- Consider discrete random variable Z_n such that

$$P\{Z_n = 0\} = 1 - \frac{1}{n}$$
$$P\{Z_n = a_n\} = \frac{1}{n}$$

  where a_n is an arbitrary sequence

- We can show $Z_n \xrightarrow{p} 0$ since for each $\delta > 0$

$$P\{|Z_n - 0| > \delta\} \leq P\{Z_n = a_n\} = \frac{1}{n} \to 0$$

Convergence in probability of vectors

- Let X_n, X be $k \times 1$ random vector with jth element denoted as X_{nj}, $j = 1 \dots k$
- Then $X_n \xrightarrow{p} X$ if and only if $X_{nj} \xrightarrow{p} X_j$ for each $j = 1 \dots k$
- Convergence in probability of a vector is defined as convergence in probability of all elements in the vector
- Same would apply for matrices

Consistency

- **Definition**: An estimator $\hat{\theta}_n$ based on a sample of size n for parameter θ is **(weakly) consistent** if $\hat{\theta}_n - \theta \xrightarrow{p} 0$, i.e., $\hat{\theta}_n \xrightarrow{p} \theta$
- Consistency is
  - an asymptotic property of an estimator
  - typically a minimum requirement for any estimator
  - a different notion compared to finite sample property such as unbiasedness
- In fact, many estimators are biased or asymptotically biased

Asymptotic unbiasedness

- **Definition**: An estimator $\hat{\theta}_n$ based on a sample of size n for parameter θ is **asymptotically unbiased** (AU) if

$$\lim_{n\to\infty}\left\{\mathbb{E}[\hat{\theta}_n]-\theta\right\}=\left\{\lim_{n\to\infty}\mathbb{E}[\hat{\theta}_n]\right\}-\theta=0$$

- **Theorem**: Consistency and asymptotic unbiasedness do not imply each other

- **Proof**: (by counterexamples)
- (1): show AU$\not\Rightarrow$Consistency
  - Suppose population is $X \sim \mathrm{N}(\mu, \sigma^2)$. Parameter of interest is μ. Given a sample $\{X_1, X_2 \dots X_n\}$ drawn from X, let

$$\hat{\mu} = X_1$$

  - Since $\mathbb{E}[\hat{\mu}] = \mathbb{E}[X_1] = \mu$, $\hat{\mu}$ is unbiased and thus AU
  - But $P\{|\hat{\mu} - \mu| > \delta\} = P\{|X - \mu| > \delta\} \nrightarrow 0$ as $n \to \infty$. Thus not consistent

- (2): show Consistency$\nRightarrow$AU
  - Consider the following artificial example
  - Suppose true parameter is θ, and $\hat{\theta}_n$ is binary

$$P\{\hat{\theta}_n = \theta\} = 1 - \frac{1}{n}, \quad P\{\hat{\theta}_n = n\} = \frac{1}{n}$$

  - $\hat{\theta}_n$ is consistent since for all $\delta > 0$

$$P\{|\hat{\theta}_n - \theta| > \delta\} \leq P\{\hat{\theta}_n = n\} = \frac{1}{n} \to 0, \text{ as } n \to \infty$$

  - However $\hat{\theta}_n$ is not AU since

$$\mathbb{E}[\hat{\theta}_n] = \theta \left(1 - \frac{1}{n}\right) + \frac{1}{n} n = \theta - \frac{\theta}{n} + \frac{n}{n}$$
$$\to \theta + 1, \text{ as } n \to \infty$$

Continuous mapping theorem

- **Theorem**: Let X_n, X be $k \times 1$ random vectors. If $X_n \xrightarrow{p} X$ and g is a real valued continuous function, then
$$g(X_n) \xrightarrow{p} g(X)$$
- **Corollary 1** [Slutsky's theorem]: Let g be continuous at c. Then
$$X_n \xrightarrow{p} c \Rightarrow g(X_n) \xrightarrow{p} g(c)$$
- **Corollary 2**: $X_n \xrightarrow{p} X \Rightarrow \|X_n - X\| \xrightarrow{p} 0$, where $\|\cdot\|$ is the Euclidean norm

2. Proving Convergence in Probability

Markov inequality

- **Definition**: Let X be a random variable and A be an event. An indicator function is

$$\mathbf{1}\{X \in A\} = \begin{cases} 1 & \text{if } X \in A \\ 0 & \text{if } X \notin A \end{cases}$$

- Note $\mathbb{E}[\mathbf{1}\{X \in A\}] = P\{X \in A\}$

- **Theorem [Markov Inequality]**: For each $r > 0$

$$P\{|X| > \delta\} \le \frac{\mathbb{E}[|X|^r]}{\delta^r}, \text{ for all } \delta > 0$$

provided $\mathbb{E}[|X|^r] < \infty$

- **Proof**

$$\begin{aligned} P\{|X| > \delta\} &= \mathbb{E}[\mathbf{1}\{|X| > \delta\}] \\ &\le \mathbb{E}\left[\mathbf{1}\{|X| > \delta\}\frac{|X|^r}{\delta^r}\right] \\ &= \frac{1}{\delta^r}\mathbb{E}\left[\mathbf{1}\{|X| > \delta\}|X|^r\right] \\ &\le \frac{\mathbb{E}\left[|X|^r\right]}{\delta^r} \end{aligned}$$

Application: convergence in $r-$th mean implies convergence in probability

- **Definition**: Assuming $\mathbb{E}[|X|^r] < \infty$. Then X_n converges in $r-th$ mean, written as $X_n \to_r X$, if

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0$$

- **Theorem**: For any $r > 0$

$$X_n \to_r X \text{ implies } X_n \xrightarrow{p} X$$

- **Proof**: by Markov inequality

$$P\{|X_n - X| > \delta\} \leq \frac{\mathbb{E}\left[|X_n - X|^r\right]}{\delta^r} \to 0, \text{ as } n \to \infty$$

Application: consistency by mean square convergence

- "Mean square convergence" is convergence in $r-$th mean for $r = 2$
- We can also show estimator $\hat{\theta}_n \xrightarrow{p} \theta$ if

$$\underbrace{\mathbb{E}[\hat{\theta}_n - \theta]^2}_{\text{mean square error}} \to 0, \text{ as } n \to \infty$$

- Since

$$\underbrace{\mathbb{E}[\hat{\theta}_n - \theta]^2}_{\text{mean square error}} = \Big[\text{bias}(\hat{\theta}_n)\Big]^2 + \text{var}(\hat{\theta}_n)$$

- We can show estimator $\hat{\theta}_n \xrightarrow{p} \theta$ if

$$\text{bias}(\hat{\theta}_n) \to 0, \text{ and } \text{var}(\hat{\theta}_n) \to 0, \text{ as } n \to \infty$$

Convergence in $r-$th mean implies AU

- **Theorem**: $\hat{\theta}_n \to_r \theta$ for some $r \geq 1$ implies $\lim_{n\to\infty} \mathbb{E}[\hat{\theta}_n] = \theta$
- **Proof**: Note

$$\begin{aligned}
\mathbb{E}[\hat{\theta}_n] - \theta &\leq |\mathbb{E}[\hat{\theta}_n - \theta]| \\
&\leq \mathbb{E}[|\hat{\theta}_n - \theta|] && \text{(Jensen's Inequality)} \\
&\leq \left\{\mathbb{E}|\hat{\theta}_n - \theta|^r\right\}^{1/r} && \text{(Jensen's Inequality again)} \\
&\to 0, \text{ as } n \to \infty
\end{aligned}$$

- **Remark**: $\hat{\theta}_n \to_r \theta$, g continuous $\Rightarrow g(\hat{\theta}_n) \xrightarrow{p} g(\theta)$
However, it is NOT true that $g(\hat{\theta}_n) \to_r g(\theta)$. $\mathbb{E}|g(\hat{\theta}_n)|^r$ might not even exist

Chebyshev's inequality

- By applying Markov inequality with $r = 2$ and replacing X with demeaned version $X - \mathbb{E}X$

  we have **Chebyshev's Inequality**

$$P\{|X - \mathbb{E}X| > \delta\} \leq \frac{\mathbb{E}[|X - \mathbb{E}X|^2]}{\delta^2} = \frac{\text{var}(X)}{\delta^2}, \text{ for all } \delta > 0$$

- **Implication**
  - An estimator $\hat{\theta}_n \xrightarrow{p} \mathbb{E}\left[\hat{\theta}_n\right]$ if $\text{var}[\hat{\theta}_n]$ is vanishing to zero

Application: Chebyshev's weak law of large numbers

- **Theorem**: If $\{X_i, i = 1, \dots n\}$ are i.i.d with mean μ and finite variance σ^2, then
$$\bar{X}_n \xrightarrow{p} \mu$$
- **Proof**: Recall we've shown under i.i.d assumption,
$$\mathbb{E}\bar{X}_n = \mu, \quad \mathrm{var}(\bar{X}_n) = \frac{\sigma^2}{n}$$
Applying Chebyshev's Inequality yields
$$P\{|\bar{X}_n - \mu| > \delta\} = P\{|\bar{X}_n - \mathbb{E}\bar{X}_n| > \delta\} \le \frac{\mathrm{var}(\bar{X}_n)}{\delta^2} = \frac{\sigma^2}{n\delta^2} \to 0, \text{ for all } \delta > 0$$

Application: Khinchine's Weak Law of Large Numbers

- **Theorem**: If $\{X_i, i = 1, \dots n\}$ are i.i.d with $\mathbb{E}|X_i| < \infty$, then

$$\bar{X}_n \xrightarrow{p} \mathbb{E}[X_i] = \mu$$

- Notice Khinchine's WLLN does not require finiteness of variance and thus is a stronger result than Chebyshev's LLN
- Khinchine's WLLN is often referred to as "the WLLN"
- Proof is technical and done by showing

$$\mathbb{E}[|\bar{X}_n - \mu|] \to 0,$$

or convergence in $r-$th mean when $r = 1$

Khinchine's WLLN for vector case

- We now extend Khinchine's WLLN to vector case
- **Theorem**: Suppose $X_i \in \mathbb{R}^m, i = 1 \dots n$ are iid distributed and $\mathbb{E}\,\|X_i\| = \mathbb{E}\,\|X\| < \infty$, then

$$\bar{X}_n \xrightarrow{p} \mathbb{E}X$$

  as $n \to \infty$

- Note $\mathbb{E}\,\|X\| < \infty$ if and only if $\mathbb{E}|X_j| < \infty$ for all $j = 1, \dots, m$

3. Almost Sure Convergence

Almost sure convergence

- Convergence in probability is sometimes called **weak convergence**
- A stronger concept is **almost sure convergence**, also known as **strong convergence**, or **convergence with probability one**
- **Definition**: We say X_n **converges almost surely** to X, denoted $X_n \stackrel{a.s.}{\to} X$, if

$$P\left\{\lim_{n\to\infty} X_n = X\right\} = 1$$

or equivalently, for all $\delta > 0$ and $\varepsilon > 0$

$$P\{|X_m - X| \leq \delta \text{ for all } m \geq n_{\delta,\varepsilon}\} > 1 - \varepsilon$$

- **Theorem**: $X_n \stackrel{a.s.}{\to} X$ implies $X_n \stackrel{p}{\to} X$

Proof

- **Proposition**: If $(C \Rightarrow D)$, then $P\{C\} \leq P\{D\}$
- Recall $X_n \xrightarrow{p} X$ if for all $\delta > 0$, $\varepsilon > 0$
  there exists some $n_{\delta,\varepsilon}$ such that for all $m \geq n_{\delta,\varepsilon}$
  $$P\{|X_m - X| \leq \delta\} > 1 - \varepsilon$$
- $X_n \xrightarrow{a.s} X$ if for all $\delta > 0$, $\varepsilon > 0$
  there exists some $n_{\delta,\varepsilon}$ such that for all $m \geq n_{\delta,\varepsilon}$
  $$P\{|X_m - X| \leq \delta \text{ for all } m \geq n_{\delta,\varepsilon}\} > 1 - \varepsilon$$
  $$\iff P\left\{\cap_{m=n_{\delta,\varepsilon}}^{\infty}\{|X_m - X| \leq \delta\}\right\} > 1 - \varepsilon$$
- Take
  $$D = |X_m - X| \leq \delta \text{ for any } m \geq n_{\delta,\varepsilon}$$
  $$C = \cap_{m=n_{\delta,\varepsilon}}^{\infty}\{|X_m - X| \leq \delta\}$$
- Clearly $C \Rightarrow D$. Hence for any $m \geq n_{\delta,\varepsilon}$
  $$\begin{aligned} P\{|X_m - X| \leq \delta\} &= P\{D\} \\ &\geq P\{C\} = P\left\{\cap_{m=n_{\delta,\varepsilon}}^{\infty}\{|X_m - X| \leq \delta\}\right\} \\ &> 1 - \varepsilon \end{aligned}$$

Strong law of large numbers (SLLN)

- **Theorem**: if X_i, $i = 1 \dots n$ are i.i.d with finite mean $\mathbb{E}|X_i| = \mathbb{E}|X| < \infty$, then

$$\bar{X}_n \stackrel{a.s.}{\to} \mathbb{E}X$$

- SLLN is a stronger asymptotic result
- Proof uses more advanced tools
- For most practical purposes weak laws of large numbers are sufficient

4. Stochastic Orders of Magnitude

Introduction

- It is convenient to have simple symbols for random variables and vectors which converge in probability to zero or are stochastically bounded
- **Definition**: [Nonstochastic orders]

  For nonstochastic sequences x_n and f_n, $n = 1, \ldots$

  1. (**small oh**) $x_n = o(f_n)$ if $\frac{x_n}{f_n} \to 0$ as $n \to \infty$.
  2. (**big oh**) $x_n = O(f_n)$ if $\frac{x_n}{f_n}$ is bounded for all sufficiently large n, that is

$$\text{there exists some } M < \infty \text{ such that for all } n \geq n_M, \left|\frac{x_n}{f_n}\right| < M$$

Stochastic orders of magnitude

- **Definition**: [Stochastic orders]

  Let X_n and f_n, $n = 1, \ldots$ be a sequence of random variables and constants

  1. (**small oh-p**) $X_n = o_p(f_n)$ if $\frac{X_n}{f_n} \xrightarrow{p} 0$
  2. (**big oh-p**) $X_n = O_p(f_n)$ if $\frac{X_n}{f_n}$ is bounded in probability, that is

     for all $\varepsilon > 0$, **there exists** a constant $M_\varepsilon < \infty$ and $n_{\varepsilon,M} > 0$ such that

     $$P\left\{\left|\frac{X_n}{f_n}\right| > M_\varepsilon\right\} < \varepsilon, \text{ for all } n \geq n_{\varepsilon,M}$$

- $X_n = o_p(1)$ simply means $X_n \xrightarrow{p} 0$

- **Theorem**: If $X_n \xrightarrow{p} c$ for some constant c, then $X_n = O_p(1)$
- Proof: For each $\varepsilon > 0$, we must find **some** constant C_ε such that for each $\varepsilon > 0$

$$P\{|X_n| > C_\varepsilon\} \leq \varepsilon, \text{ for all } n \geq n_{\varepsilon,C}$$

- Since $X_n \xrightarrow{p} c$, we know for each $\varepsilon > 0$, and **each** $\delta > 0$

$$P\{|X_n - c| > \delta\} < \varepsilon, \text{ for all n} \geq n_{\delta,\varepsilon} \tag{1}$$

- By triangle inequality

$$|X_n| \leq |X_n - c| + |c| \tag{2}$$

- Pick $C = |c| + \delta$. Combining (1) and (2) yield

$$\begin{aligned} P\{|X_n| > C\} &= P\{|X_n| > |c| + \delta\} \\ &\leq P\{|X_n - c| + |c| > |c| + \delta\} \\ &= P\{|X_n - c| > \delta\} \\ &< \varepsilon, \text{ for all } n \geq n_{\delta,\varepsilon} \end{aligned}$$

Algebra of stochastic orders

1. If $X_n = O_p(f_n)$, $Y_n = O_p(g_n)$, then
   - $X_n Y_n = O_p(f_n g_n)$
   - $X_n + Y_n = O_p(\max(f_n, g_n))$
2. We can replace O by o everywhere in ❶
3. If $X_n = O_p(f_n)$, $Y_n = o_p(g_n)$, then $X_n Y_n = o_p(f_n g_n)$
4. If $X_n = O_p(f_n)$ and $\frac{f_n}{g_n} \to 0$, then $X_n = o_p(g_n)$

Why stochastic symbols are useful?

- We use stochastic orders because we want a simple characterization of how fast X_n converges to X in probability
- Example: Suppose $\{X_i, i = 1 \dots n\}$ are i.i.d with finite finite variance σ^2. We know from weak law of large numbers

$$\bar{X}_n \xrightarrow{p} \mu$$

- But how fast does $\bar{X}_n$ converge to μ?

- To tackle this, recall by Chebyshev's inequality

$$P\{|\bar{X}_n - \mu| > \delta\} \leq \frac{\sigma^2}{n\delta^2}, \text{ for all } \delta > 0$$

- It also implies that for all δ

$$P\left\{\frac{|\bar{X}_n - \mu|}{\frac{1}{\sqrt{n}}} > \delta\right\} = P\left\{|\bar{X}_n - \mu| > \frac{1}{\sqrt{n}}\delta\right\} \leq \frac{\sigma^2}{\delta^2} \qquad (3)$$

- From (3), for each $\varepsilon > 0$, we can choose $C_\varepsilon = \frac{\sigma}{\sqrt{\varepsilon}}$ such that

$$P\left\{\frac{|\bar{X}_n - \mu|}{\frac{1}{\sqrt{n}}} > C_\varepsilon\right\} \leq \varepsilon$$

- Hence $\bar{X}_n - \mu = O_p(\frac{1}{\sqrt{n}})$, or equivalently $\bar{X}_n = \mu + O_p(\frac{1}{\sqrt{n}})$
- $\bar{X}_n$ converges to μ at a rate no slower than $\frac{1}{\sqrt{n}}$

Derive stochastic order from bounded moments

- **Theorem**: $X_n = O_p\left\{[\mathbb{E}|X_n|^r]^{\frac{1}{r}}\right\}$ for $r > 0$
- **Proof**: For each $\varepsilon > 0$, pick $C_\varepsilon = \left(\frac{1}{\varepsilon}\right)^{\frac{1}{r}}$

  It follows by Markov Inequality

$$
\begin{aligned}
P\left\{\left|\frac{X_n}{[\mathbb{E}|X_n|^r]^{\frac{1}{r}}}\right| > C_\varepsilon\right\} &= P\left\{|X_n| > [\mathbb{E}|X_n|^r]^{\frac{1}{r}}\, C_\varepsilon\right\} \\
&\leq \frac{\mathbb{E}|X_n|^r}{\mathbb{E}|X_n|^r C_\varepsilon^r} \\
&= \frac{1}{C_\varepsilon^r} = \varepsilon
\end{aligned}
$$

5. Convergence in Distribution

Motivation

- From previous sections we show sample mean converge to population mean in probability
- And we are also able to characterize is convergence rate by using stochastic symbols
- However, for most economic applications, this is not enough
- In order to do inference, we also need to approximate the sampling distribution of sample mean
  - Sampling distribution is a function of the unknown population distribution F and sample size n
  - Study the sampling distribution by letting $n \to \infty$
  - Hopefully after some standardization, as $n \to \infty$, the sampling distribution becomes much more tractable than the unknown F

Convergence in distribution

- Let $F_X(x) = P\{X \leq x\}$ be the distribution function of random variable X
- Consider a sequence of random variables X_n with distribution function $F_{X_n}(x) = P\{X_n \leq x\}$
- **Definition**: X_n **converges in distribution** to X ($X_n \xrightarrow{d} X$) if

$$F_{X_n}(a) \rightarrow F_X(a) \text{ as } n \rightarrow \infty$$

for all a where $F_X(a)$ **is continuous**

Equivalent conditions for convergence in distribution

- Technically it is often difficult to show $X_n \xrightarrow{d} X$ by working directly with cdf. Following theorem guarantees that instead we can work with characteristic function
- **Theorem**: $X_n \xrightarrow{d} X \Leftrightarrow C_{X_n}(t) \to C_X(t)$, as $n \to \infty$ for all t, where $C_X(t) = \mathbb{E}[\exp(\mathrm{i}tX)]$ is the characteristic function of X

Relationship between $\xrightarrow{p}$, $\xrightarrow{d}$ and $O_p(1)$

- **Theorem**

  1. $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$
  2. $X_n \xrightarrow{p} c \iff X_n \xrightarrow{d} c$ for some constant c
  3. $X_n \xrightarrow{d} X \Rightarrow X_n = O_p(1)$

Proof for statement ❷

- (1): show $X_n \xrightarrow{p} c \Rightarrow X_n \xrightarrow{d} c$
- The cdf of a constant variable X such that $P\{X = c\} = 1$ is degenerate

$$P\{X \leq x\} = \begin{cases} 0 & \text{if } x < c \\ 1 & \text{if } x \geq c \end{cases}$$

- We need to show
  - (a) For each $\delta > 0$, $P\{X_n \leq c - \delta\} \to 0$ as $n \to \infty$
  - (b) For each $\delta > 0$, $P\{X_n \leq c + \delta\} \to 1$ as $n \to \infty$
- To see (a), note

$$P\{X_n \leq c - \delta\} = P\{X_n - c \leq -\delta\} \leq P\{|X_n - c| \geq \delta\} \to 0$$

by definition of $X_n \xrightarrow{p} c$

- To see (b), it suffices to show $P\{X_n > c + \delta\} \to 0$ as $n \to \infty$ and the proof is similar to (a)

- (2): show $X_n \xrightarrow{d} c \Rightarrow X_n \xrightarrow{p} c$
- Note for each $\delta > 0$,

$$
\begin{aligned}
P\{|X_n - c| > \delta\} &= P\{X_n - c > \delta\} + P\{X_n - c < -\delta\} \\
&\leq 1 - F_{X_n}(\delta + c) + F_{X_n}(c - \delta) \\
&\to 1 - 1 + 0 = 0, \text{ as } n \to \infty
\end{aligned}
$$

Asymptotic distribution of sample mean

- The aim is to approximate the distribution of $\bar{X}_n$ as $n \to \infty$
- By weak law of large numbers $\bar{X}_n \xrightarrow{p} \mu$. Thus $\bar{X}_n \xrightarrow{d} \mu$
  - The asymptotic distribution of $\bar{X}_n$ degenerates to μ
- In order to get more useful results, we need to rescale $\bar{X}_n$ so that it has a stable distribution
- Since $\mathrm{var}(\bar{X}_n) = \frac{\sigma^2}{n}$, consider

$$Z_n = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)$$

- Note $\mathbb{E}[Z_n] = 0$, $\mathrm{var}(Z_n) = 1$. The distribution of Z_n is "stabilized"
- We aim to find the asymptotic distribution of Z_n

Lindeberg-Lévy central limit theorem

- **Theorem**: If $X_i, i = 1, \ldots, n$ are i.i.d and $\mathbb{E}X_i^2 < \infty$ then

$$Z_n \xrightarrow{d} \mathrm{N}(0,1), \text{ or equivalently, } \sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathrm{N}(0,\sigma^2)$$

where $\mathbb{E}[X_i] = \mu$ and $\sigma^2 = \mathrm{var}(X_i)$

Proof of Lindeberg-Lévy CLT

- Wlog, assume $\mu = 0$
- We show $C_{Z_n}(t) \to \exp\left(-\frac{t^2}{2}\right)$ as $n \to \infty$, since $\exp\left(-\frac{t^2}{2}\right)$ is the CF of a standard normal
- Note $Z_n = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) = \sum_{j=1}^{n} x_{jn}$, where $x_{jn} = \frac{(X_j - \mu)}{\sigma\sqrt{n}} = \frac{X_j}{\sigma\sqrt{n}}$.

$$
\begin{aligned}
C_{Z_n}(t) = \mathbb{E}[\exp(\mathrm{i}tZ_n)] &= \mathbb{E}\left[\exp\left(\mathrm{i}t\sum_{j=1}^{n} x_{jn}\right)\right] \\
&= \prod_{j=1}^{n} \mathbb{E}[\exp(itx_{jn})] \text{(by independence)} \\
&= \{\mathbb{E}[\exp(itx_{1n})]\}^n \text{ (by indentical distribution)} \\
&= \left\{C_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}^n
\end{aligned}
$$

where $C_{X_1}(s) = \mathbb{E}[\exp(\mathrm{i}sX_1)]$ is the CF of X_1

- Since $\mathbb{E}X_1^2 < \infty$, by Taylor's Theorem

$$C_{X_1}(s) = \underbrace{C_{X_1}(0)}_{1} + \mathrm{i}s\underbrace{\mathbb{E}X_1}_{0} + \frac{\mathrm{i}^2 s^2}{2}\underbrace{\mathbb{E}X_1^2}_{\sigma^2} + o(s^2), \text{ as } s \to 0$$

- Hence for each fixed t,

$$C_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{\sigma^2 n}\right)$$

- And for each fixed t, as $n \to \infty$

$$C_{Z_n}(t) = \left\{1 - \frac{t^2}{2n} + o\left(\frac{t^2}{\sigma^2 n}\right)\right\}^n \to e^{-\frac{t^2}{2}}$$

since $\left(1 + \frac{a}{n}\right)^n \to e^a$ as $n \to \infty$. Conclusion follows

Multivariate central limit theorem

- **Theorem**: [Cramér-Wold Device]
  For a sequence of random vectors $X_n \in \mathbb{R}^k$,

$$X_n \xrightarrow{d} X \Longleftrightarrow \lambda' X_n \xrightarrow{d} \lambda' X, \text{ for all } \lambda \in \mathbb{R}^k$$

- The above theorem implies that
  to show a random vector X_n is asymptotically multivariate normal, it is necessary and sufficient to show that any linear combination of elements of X_n is asymptotically univariate normal

- **Theorem**: [Multivariate Lindeberg-Lévy CLT]
  If $X_i, i = 1, \ldots, n$ are i.i.d and $\mathbb{E} \|X_i\|^2 < \infty$ then

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathrm{N}(0, \Sigma),$$

  where $\mu = \mathbb{E}[X_i]$ and $\Sigma = \mathbb{E}\left[(X_i - \mu)(X_i - \mu)'\right]$

6. Delta Method

Motivation

- So far we consider $\bar{X}_n$ to estimate $\mathbb{E}[X_i]$
- Same idea applies to transformation of X, say $g(X)$
- We can obtain LLN and CLT like

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{p} \mathbb{E}[g(X)] = \mu$$

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \mathrm{N}(0, \mathrm{var}(g(X)))$$

- Just replace "X" with "$g(X)$" in previous slides

Functions of moments

- How about functions of moments

$$\beta = h(\mu) = h(\mathbb{E}[g(X)])$$

where $h(\cdot)$ is a possibly nonlinear transformation

- Natural estimator is **plug-in estimator**

$$\hat{\beta} = h(\hat{\mu}), \text{where } \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

- How do we derive the asymptotic distribution of $\hat{\beta}$?

Continuous mapping theorem

- **Theorem**: For random vectors $X_n \in \mathbb{R}^k$ and $X \in \mathbb{R}^k$

$$X_n \xrightarrow{d} X,\ g \text{ is continuous} \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

- Convergence in distribution is preserved under continuous transformations

- **Theorem**: If $X_n \xrightarrow{d} X$ and $c_n \xrightarrow{p} c$, then
  - $X_n + c_n \xrightarrow{d} X + c$
  - $X_n c_n \xrightarrow{d} Xc$
  - $\frac{X_n}{c_n} \xrightarrow{d} \frac{X}{c}$ provided $c \neq 0$

- Example 1: $X_n \xrightarrow{d} X \sim \mathrm{N}(0, I_k) \Rightarrow X_n' X_n \xrightarrow{d} X'X \sim \chi_k^2$
- Example 2: [Normal approximation with estimated variance]
  - Suppose $\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{d} \mathrm{N}(0, 1)$ and $\hat{\sigma}$ is a consistent estimator of $\sigma > 0$
  - Then $\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\hat{\sigma}}\right) = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)\left(\frac{\sigma}{\hat{\sigma}}\right) \xrightarrow{d} \mathrm{N}(0, 1)$

Delta method

- Now let us derive asymptotic distribution of $\hat{\beta} = h(\hat{\mu})$
- Note that $\hat{\beta}$ is written as function of $\hat{\mu}$ (not $\sqrt{n}(\hat{\mu} - \mu)$), so CMT is not directly applicable
- Key step is first-order Taylor expansion (by assuming differentiability of $h(\cdot)$)

$$\hat{\beta} = h(\hat{\mu}) = h(\mu) + \frac{\partial h(u)}{\partial u'}|_{u=\mu^*}(\hat{\mu} - \mu)$$

where μ^* is on the line joining $\hat{\mu}$ and μ. Then

$$\sqrt{n}(\hat{\beta} - h(\mu)) = \frac{\partial h(u)}{\partial u'}|_{u=\mu^*}\sqrt{n}(\hat{\mu} - \mu)$$

so we can use asymptotic distribution of $\sqrt{n}(\hat{\mu} - \mu)$ and CMT

- **Theorem**: If $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \xi$ and $h(\cdot)$ is a function continuously differentiable in a neighborhood μ, then

$$\sqrt{n}(h(\hat{\mu}) - h(\mu)) \xrightarrow{d} \mathbf{H}'\xi,$$

where $\mathbf{H}' = \frac{\partial}{\partial u'} h(u) \mid_{u=\mu}$

In particular, if $\xi \sim \mathrm{N}(0, V)$, then

$$\sqrt{n}(h(\hat{\theta}) - h(\theta)) \xrightarrow{d} \mathrm{N}(0, \mathbf{H}' V \mathbf{H}) \tag{4}$$

When μ and h are scalar in (4)

$$\sqrt{n}(h(\hat{\mu}) - h(\mu)) \xrightarrow{d} \mathrm{N}\left(0, \left(\frac{\partial}{\partial u} h(u) \mid_{u=\mu}\right)^2 V\right)$$