ECMA31000: Introduction to Empirical Analysis

Asymptotics II

Joe Hardwick

University of Chicago

Autumn 2021

Outline

- Last week:
  - Studied 4 modes of convergence, $\xrightarrow{p}$and $\xrightarrow{a.s.}$, $\xrightarrow{d}$, $\xrightarrow{\text{r-th}}$.
- This week:
  - Continuous mapping theorem.
  - Slutsky's Theorem.
  - Law of large numbers and central limit theorem. $\rightarrow$ These will provide (with CMT) Necessary $\rightarrow^p$, $\rightarrow^d$ results.
  - Stochastic Order notation.
  - Delta Method.

$$\sqrt{n}(\bar{X}_n - \mu) \rightarrow^d N(0, V)$$

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \rightarrow^d ?$$ (Delta Method).

Definitions

- A sequence of $(K \times 1)$ random vectors $X_n \xrightarrow{p} X$ if $\forall \epsilon > 0$:

$$\lim_{n \to \infty} \mathrm{P}\left(\|X_n - X\| > \epsilon\right) = 0.$$

- $X_n \xrightarrow{a.s} X$ if:

$$\mathrm{P}\left(\{\omega : \|X_n(\omega) - X(\omega)\| \to 0\}\right) = 1.$$

- $X_n \xrightarrow{\text{r-th}} X$ if

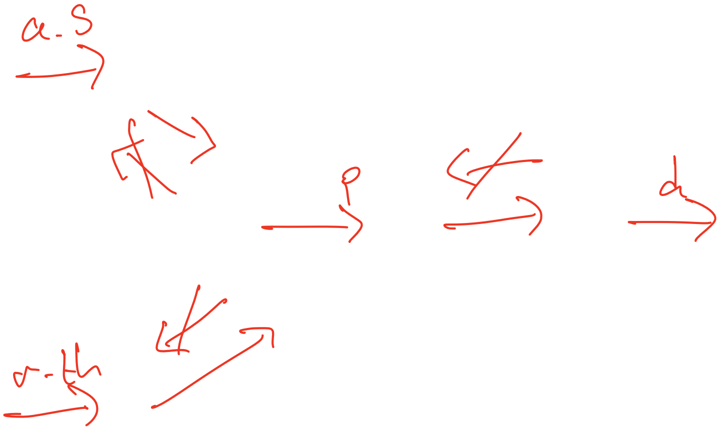
$$\lim_{n \to \infty} \mathrm{E}\left(\|X_n - X\|^r\right) = 0.$$

- $X_n \xrightarrow{d} X$ if

$$F_{X_n}(x) \to F_X(x)$$

for all x such that F_X is continuous at x.

Summary of implications



Continuous Mapping Theorem

- Let $g : \mathbb{R}^k \to \mathbb{R}^n$ be a function that is continuous on a set $S \subset \mathbb{R}^k$ with $\mathrm{P}(X \in S) = 1$. Then the following hold:

$$X_n \xrightarrow{a.s.} X \implies g(X_n) \xrightarrow{a.s.} g(X);$$
$$X_n \xrightarrow{p} X \implies g(X_n) \xrightarrow{p} g(X);$$
$$X_n \xrightarrow{d} X \implies g(X_n) \xrightarrow{d} g(X).$$

- The theorem doesn't hold for $\xrightarrow{\text{r-th}}$: Take

$$X_n \xrightarrow{1st} 0.$$

$$X_n = \begin{cases} n & \text{with probability } \frac{1}{n^2}; \\ 0 & \text{with probability } 1 - \frac{1}{n^2}. \end{cases}$$

Letting $g(x) = x^2$, we see $\mathrm{E}(|X_n - 0|) = \frac{1}{n} \to 0$ but, for all n:

$$\underbrace{\mathrm{E}(|g(X_n) - g(0)|)}_{\Rightarrow \ \downarrow 0} = \mathrm{E}\left(\left|X_n^2 - 0\right|\right) = 1.$$

$$n^2 \cdot \frac{1}{n^2} + 0 \cdot \left(1 - \frac{1}{n^2}\right)$$

$\xrightarrow{1\text{-st}}$

Continuous Mapping theorem

$= 1.$

- It is important that $\mathrm{P}(X \in S) = 1$. To see this, suppose

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ c \end{pmatrix}.$$

- Consider the continuous function $g : \mathbb{R}^2 \to \mathbb{R}$ given by $g(x, y) = \frac{x}{y}$. g is continuous except at points $(x, 0) \in \mathbb{R}^2$. Set

$$S = \mathbb{R}^2 \setminus \{(x, 0) : x \in \mathbb{R}\}.$$

If $c \neq 0$, then $\mathrm{P}((X, c) \in S) = 1$, and the CMT gives

$$g(X_n, Y_n) = \frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c} = g(X, c).$$

If $c = 0$, then $\mathrm{P}((X, c) \in S) = 0$. This would lead to the nonsensical result

$$g(x_n, y_n) = \frac{x_n}{y_n} \xrightarrow{d} \frac{X}{0} \quad \times .$$

Example: Slutsky's Theorem

Theorem

Suppose $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{p} c$ *for some constant* c*. Then:*

$g(x,y) = x+y.$

$g(x,y) = xy.$

$g(x,y) = \frac{x}{y}.$

$$X_n + Y_n \xrightarrow{d} X + c;$$

$$X_n Y_n \xrightarrow{d} Xc;$$

$$X_n / Y_n \xrightarrow{d} X/c \text{ provided } c \neq 0.$$

$X_n \to^d X \quad Y_n \to^p c$

$\Rightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \to^d \begin{pmatrix} X \\ c \end{pmatrix}$

Proof.

We stated that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ implies

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ c \end{pmatrix}.$$

The result follows by noting that $x + y$, xy and x/y are all continuous functions of (x, y), provided $y \neq 0$ in the last case. □

Example: Slutsky's Theorem

$X_n \xrightarrow{d} X \quad Y_n \xrightarrow{p} C$.

$X_n \sim N(0,1) \quad Y_n \sim N(0,1)$

- It is important that Y_n converges in probability to a constant! If instead, for all n

n odd: $Y_n = -X_n$

$X_n \xrightarrow{d} N(0,1)$

$Y_n \xrightarrow{d} N(0,1)$

$$Y_n = (-1)^n X_n \sim \mathcal{N}(0,1),$$

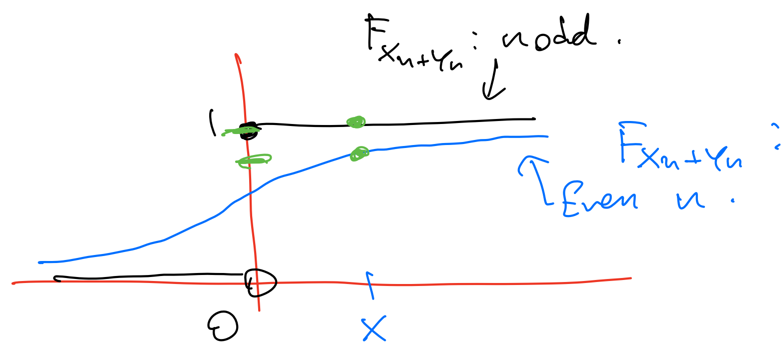
n even: $Y_n = X_n$.

then $X_n + Y_n = 0$ for odd n, and $X_n + Y_n \sim \mathcal{N}(0,4)$ for even n. Thus,

$= 2X_n \sim N(0,4)$.

$$F_{X_n+Y_n}(x)$$

does not converge for any $x \in \mathbb{R}$. This means $X_n + Y_n$ cannot converge in distribution.



Example: Sample Correlation

$\frac{1}{n}\sum X_i \to^p E(X_i)$

X_i iid $E(X_i)=\mu$ $<\infty$.

- Suppose $\{(X_i, Y_i)\}_{i\geq 1}$ is a sequence of (2×1) iid random vectors with $\mathrm{E}\left(X_i^2\right) < \infty$, $\mathrm{E}\left(Y_i^2\right) < \infty$.
- The sample correlation between X, Y is given by

$P(|\hat{\rho}-\rho|>\varepsilon) \to 0$ as $n\to\infty$.

$$\hat{\rho}_{XY} = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2}}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^n X_i Y_i - \bar{X}\,\bar{Y}}{\sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2}\sqrt{\frac{1}{n}\sum_{i=1}^n Y_i^2 - \bar{Y}^2}},$$

where $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$ and $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ are the sample means.

$g(x)=x^2$

$\bar{X} \to^p E(X_i)$ $\quad$ $\bar{X}^2 \to^p g(E(X_i)) = [E(X_i)]^2.$

Example: Sample Correlation

$E|X_iY_i| \le E(X_i^2)^{1/2} \cdot E(Y_i^2)^{1/2}$

- We have already seen that $\bar{X} \xrightarrow{p} \mathrm{E}(X)$ and $\bar{Y} \xrightarrow{p} \mathrm{E}(Y)$.
- Next, note that since the vectors (X_i, Y_i) are iid, the product $X_i Y_i$ is an iid sequence of random variables.

$E(X_iY_i)^2 \le E(X_i^2) \cdot E(Y_i)^2.$

- By the weak law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{p} \mathrm{E}(XY);$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} \mathrm{E}(X^2);$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \xrightarrow{p} \mathrm{E}(Y^2).$$

Example: Sample Correlation

- It follows that

$$\left(\overset{x}{\bar{X}}, \overset{y}{\bar{Y}}, \overset{s}{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i}, \overset{t}{\frac{1}{n}\sum_{i=1}^{n} X_i^2}, \overset{w}{\frac{1}{n}\sum_{i=1}^{n} Y_i^2}\right)$$

$$\xrightarrow{p} \left(\underset{x}{\mathrm{E}(X)}, \mathrm{E}(Y), \mathrm{E}(XY), \underset{t}{\mathrm{E}\left(X^2\right)}, \mathrm{E}\left(Y^2\right)\right).$$

- Now let

$E(X^2) - [E(X)]^2 = 0.$

$$g(x, y, s, t, w) = \frac{s - xy}{\sqrt{t - x^2}\sqrt{w - y^2}}.$$

- g is continuous at all points except where $t = x^2$ and $w = y^2$. Provided neither X nor Y are constant random variables (if they were the sample variances would also be 0!) we get

$$\mathrm{E}\left(X^2\right) > \mathrm{E}(X)^2; \qquad \mathrm{E}\left(Y^2\right) > \mathrm{E}(Y)^2,$$

so g is continuous at $\left(\mathrm{E}(X), \mathrm{E}(Y), \mathrm{E}(XY), \mathrm{E}\left(X^2\right), \mathrm{E}\left(Y^2\right)\right)$.

Example: Sample Correlation

$$g\begin{pmatrix} \bar{X} \\ \bar{Y} \\ \frac{1}{n}\sum X_i Y_i \\ \frac{1}{n}\sum X_i^2 \\ \frac{1}{n}\sum Y_i^2 \end{pmatrix} \xrightarrow{p} g\begin{pmatrix} E(X) \\ E(Y) \\ E(XY) \\ E(X^2) \\ E(Y^2) \end{pmatrix}$$

- It follows by the continuous mapping theorem that:

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2}} \xrightarrow{p} \frac{\mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)}{\sqrt{\mathrm{E}(X^2) - \mathrm{E}(X)^2}\sqrt{\mathrm{E}(Y^2) - \mathrm{E}(Y)^2}} = \frac{Cov(X, Y)}{\sqrt{Var(X)\,Var(Y)}},$$

which is the population correlation coefficient.

Example

$d(A,B) = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2}$ where i indexes the entries of the matrix

$A_n \to^p A$.

Convergence in p of a matrix is convergence of the vector composed by stacking columns vertically to form a single column vector.

- A form of the following result will appear several times when we analyse OLS/IV estimators.
- Let $A_n \in \mathbb{R}^{P \times K}$ be a sequence of matrices converging in probability to a constant matrix A.
  - This is just the same as vector convergence: Stack the columns on top of each other!
- Let B_n be a sequence of $(K \times 1)$ random vectors such that

$$B_n \xrightarrow{d} \mathcal{N}(\mu, \Sigma).$$

$K \times K$

$K \times 1$

Then:

$$A_n B_n \xrightarrow{d} A\mathcal{N}(\mu, \Sigma) \sim \mathcal{N}(A\mu, A\Sigma A').$$

B

Example

$X_n \to^d X$ $\Rightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \to^d \begin{pmatrix} X \\ c \end{pmatrix}$

$Y_n \to^p c$

Proof.

Since the vector consisting of the columns of A_n, denoted $vec\,(A_n)$ converges in probability to $vec\,(A)$, a constant vector, we obtain

$$\begin{pmatrix} B_n \\ vec\,(A_n) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathcal{N}\,(\mu, \Sigma) \\ vec\,(A) \end{pmatrix}.$$

By the continuous mapping theorem,

$$A_n B_n \xrightarrow{d} A\mathcal{N}\,(\mu, \Sigma)\,.$$

$E(A\,N(\mu, \Sigma))$
$= A\,E(N(\mu, \Sigma))$
$= A\mu.$

Since $\mathrm{E}\,(AX) = A\mathrm{E}\,(X)$, and we showed in problem set 1 that $Var\,(AX) = AVar\,(X)\,A'$, we conclude that

$$A\mathcal{N}\,(\mu, \Sigma) \sim \mathcal{N}\,(A\mu, A\Sigma A')$$

because linear transformations of multivariate normals are also (multivariate) normal. □

Questions?

Strong Law of Large Numbers (SLLN)

- In Lecture 3, we proved that if $\{X_i\}_{i\geq 1}$ is an iid sequence with $\mathrm{E}\left(X_i^2\right) < \infty$, then

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mathrm{E}\left(X_1\right).$$

- Called this a "weak" law of large numbers because it establishes $\xrightarrow{p}$, not $\xrightarrow{a.s.}$.
- The SLLN delivers a stronger result under a weaker condition:

Theorem

If $\{X_i\}_{i\geq 1}$ is an iid sequence with $\mathrm{E}\left(X_i\right) = \mu$, then

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mathrm{E}\left(X_1\right).$$

X_i iid (have same mean).

Strong Law of Large Numbers (SLLN)

$$\frac{1}{n}\sum \begin{pmatrix} X_{i1} \\ \vdots \\ X_{iK} \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum X_{i1} \\ \vdots \\ \frac{1}{n}\sum X_{iK} \end{pmatrix}$$

- Result also holds for a sequence of $(K \times 1)$ iid random vectors $\{X_i\}_{i\geq 1}$ such that $\mathrm{E}(X_i) = \mu$, since

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_{iK} \end{pmatrix},$$

and $\{X_{ij}\}_{i\geq 1}$ are iid sequences of random variabes with $\mathrm{E}(X_{ij}) = \mu_j$, for each $j = 1, \ldots, K$.

Strong Law of Large Numbers (SLLN)

- By the SLLN for random variables,

$$\frac{1}{n}\sum_{i=1}^{n} X_{ij} \xrightarrow{a.s.} \mathrm{E}\left(X_{1j}\right)$$

for each $j = 1, \ldots, K$. It follows that

Marginal $\xrightarrow{a.s}$

$\Rightarrow$ Joint $\xrightarrow{a.s}$

$$\frac{1}{n}\sum_{i=1}^{n} X_{i} \xrightarrow{a.s.} \mathrm{E}\left(X_{1}\right).$$

Central Limit Theorem

- Let $\{X_i\}_{i\geq 1}$ be an iid sequence with $\mathrm{E}\left(X_i\right) = \mu$ and $Var\left(X_i\right) = \sigma^2$ and consider the distribution of the sample mean $\bar{X}_n$.
- We know that

$$\mathrm{E}\left(\bar{X}_n\right) = \mu; \quad Var\left(\bar{X}_n\right) = \frac{\sigma^2}{n}.$$

Central Limit Theorem

$E(\bar{X}_n - \mu) = 0.$ $\quad E(\sqrt{n}(\bar{X}_n - \mu)) = 0.$

$Var(\bar{X}_n - \mu) = \frac{\sigma^2}{n}$ $\quad Var(\sqrt{n}(\bar{X}_n - \mu)) = \sigma^2.$

- It follows that

$$\mathrm{E}\left(\sqrt{n}\left(\bar{X}_n - \mu\right)\right) = 0; \quad Var\left(\sqrt{n}\left(\bar{X}_n - \mu\right)\right) = \sigma^2.$$

- Note that if $X_i \sim \mathcal{N}\left(\overset{\mu}{\cancel{0}}, \sigma^2\right)$, then for all n:

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \sim \mathcal{N}\left(0, \sigma^2\right).$$

- The Central Limit Theorem provides an approximation to the distribution of $\sqrt{n}\left(\bar{X}_n - \mu\right)$ when X_i are non-normal.
- In "large" samples, $\sqrt{n}\left(\bar{X}_n - \mu\right)$ is approximately normally distributed, sometimes written as

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \overset{a}{\sim} \mathcal{N}\left(0, \sigma^2\right).$$

$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$

Central Limit Theorem

$\mathbb{E}(X_n) \to \mu$

$Var(X_n) \to 0$

$X_n \xrightarrow{p} \mu.$

- Notice that if instead of $\sqrt{n}$, we choose n^r, for some $r > 0$:

$$Var\left(n^r\left(\bar{X}_n - \mu\right)\right) = \frac{\sigma^2 n^{2r}}{n} \to \begin{cases} 0 & \text{if } r < \frac{1}{2} \\ \sigma^2 & \text{if } r = \frac{1}{2} \\ +\infty & \text{if } r > \frac{1}{2} \end{cases}.$$

- By Problem Set 2 Q3, for $r < \frac{1}{2}$, $\mathrm{E}\left(n^r\left(\bar{X}_n - \mu\right)\right) = 0$ and $Var\left(n^r\left(\bar{X}_n - \mu\right)\right) \to 0$ implies $n^r\left(\bar{X}_n - \mu\right) \xrightarrow{p} 0$.
- We say the limiting distribution of $n^r\left(\bar{X}_n - \mu\right)$ is degenerate, because it has the same distribution as a constant random variable (in this case, 0).
- We now consider the cases $r = \frac{1}{2}, r > \frac{1}{2}$.

Central Limit Theorem

- We say a sequence of random variables $\{Y_n\}_{\geq 1}$ converges in distribution to a non-degenerate limit if:
  - Y_n converges in distribution to some random variable Y,
  - Y is not (almost surely) constant.
- If $r = \frac{1}{2}$, The distribution of $\sqrt{n}\left(\bar{X}_n - \mu\right)$ converges to the distribution of $\mathcal{N}\left(0, \sigma^2\right)$, no matter what the initial distribution of the X_i! This is the classical Central Limit Theorem (CLT):

Theorem

(Lindeberg-Lévy CLT) Let $\{X_i\}_{i\geq 1}$ *be an iid sequence of random variables with* $\mathrm{E}\left(X_i\right) = \mu$ *and* $Var\left(X_i\right) = \sigma^2$ *. Then*

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right).$$

Central Limit Theorem

$n^r(\bar{X}_n - \mu) = n^{r-\frac{1}{2}} \cdot \sqrt{n}(\bar{X}_n - \mu)$.

- Next, denote $Y_n^r := n^r\left(\bar{X}_n - \mu\right)$ and let $r > \frac{1}{2}$.
- For any $M > 0$, and $n \geq M^{2/(r-1/2)}$:

$$\mathrm{P}\left(Y_n^r \leq M\right) = \mathrm{P}\left(\sqrt{n}\left(\bar{X}_n - \mu\right) \leq \frac{M}{n^{r-1/2}}\right)$$

Eventually $\leq \frac{M}{M^2}$.

$$\left(\text{since } n^{r-1/2} \geq M^2\right) \leq \mathrm{P}\left(\sqrt{n}\left(\bar{X}_n - \mu\right) \leq \frac{M}{M^2}\right)$$

$$\xrightarrow{CLT} \Phi\left(\frac{1}{M}\right)$$

$\frac{1}{M}$.

as $n \to \infty$.

Φ is the CDF of a standard normal

(Assumed $\sigma^2 = 1$ WLOG, so $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0,1)$).

Central Limit Theorem

$F_{Y_n}(M) \to F_Y(M)$ if M is a continuity pt. of F_Y.

- If it held that $Y_n^r \xrightarrow{d} Y$ for some Y, the definition of convergence in distribution and the above would imply:

$$F_Y(M) = \lim_{n\to\infty} F_{Y_n^r}(M) \le \Phi\left(\frac{1}{M}\right),$$

for all continuity points M of F_Y.

- Since F_Y is a distribution function, we obtain

$\Phi(0)$.

$$1 = \lim_{M\to\infty} F_Y(M) \le \lim_{M\to\infty} \Phi\left(\frac{1}{M}\right) = \frac{1}{2},$$

a contradiction. Therefore, Y_n^r cannot converge in distribution for $r > \frac{1}{2}$.

Questions?

Example: Gamma Distribution

$\sqrt{n}(\bar{X}_n - \mu) \overset{a}{\sim} N(0, \sigma^2)$

- We now demonstrate that there is no good rule of thumb for how large n should be to invoke the CLT.
- Let X_i be iid with $\mathrm{E}(X_i) = \mu$, $Var(X_i) = \sigma^2$ and skewness $\kappa = \mathrm{E}\left(\left(\frac{X_i - \mu}{\sigma}\right)^3\right)$.
- For symmetric distributions (e.g. normal), $\kappa = 0$.
- Can show that the skewness of the sample mean is

$$\mathrm{E}\left(\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right)^3\right) = \frac{\kappa}{\sqrt{n}},$$

which decreases with n, as we may expect.

Example: Gamma Distribution

- Suppose $X_i \sim \textit{Gamma}(k, \theta)$, where k, θ are shape and scale parameters respectively.
- Conveniently $\sum_{i=1}^{n} X_i \sim \textit{Gamma}(kn, \theta)$ and $\bar{X}_n \sim \textit{Gamma}\left(kn, \frac{\theta}{n}\right)$.
- We have $\mathrm{E}(X_i) = k\theta$ and $\textit{Var}(X_i) = k\theta^2$.
- The skewness of the gamma distribution depends only on the shape parameter:

$$\kappa = \mathrm{E}\left(\left(\frac{X_i - k\theta}{\sqrt{k\theta^2}}\right)^3\right) = \frac{2}{\sqrt{k}}.$$

Example: Gamma Distribution

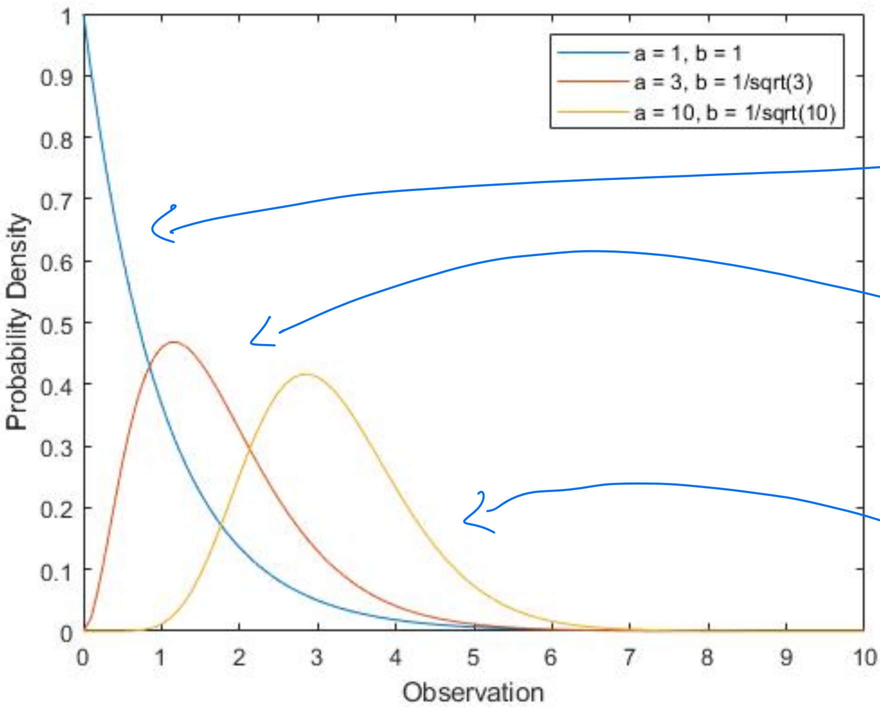
- For the sample mean:

$$\mathrm{E}\left(\left(\frac{\bar{X}_n - k\theta}{\sqrt{\left(k\theta^2\right)/n}}\right)^3\right) = \frac{\kappa}{\sqrt{n}} = \frac{2}{\sqrt{nk}}.$$

  Therefore, the skewness is only close to 0 if $\sqrt{nk}$ is large.
- So if we have $n = 1000$, but X_i happens to be drawn from $Gamma\,(0.001, \theta)$, distribution of $\bar{X}_n$ still positively skewed:

Example: Gamma Distribution

- Can show $\frac{\bar{X}_n}{\sqrt{(k\theta^2)/n}} \sim Gamma\left(nk, \frac{1}{\sqrt{nk}}\right)$:



$n = 1000$

$k = \frac{1}{1000}$

$n = 3000$

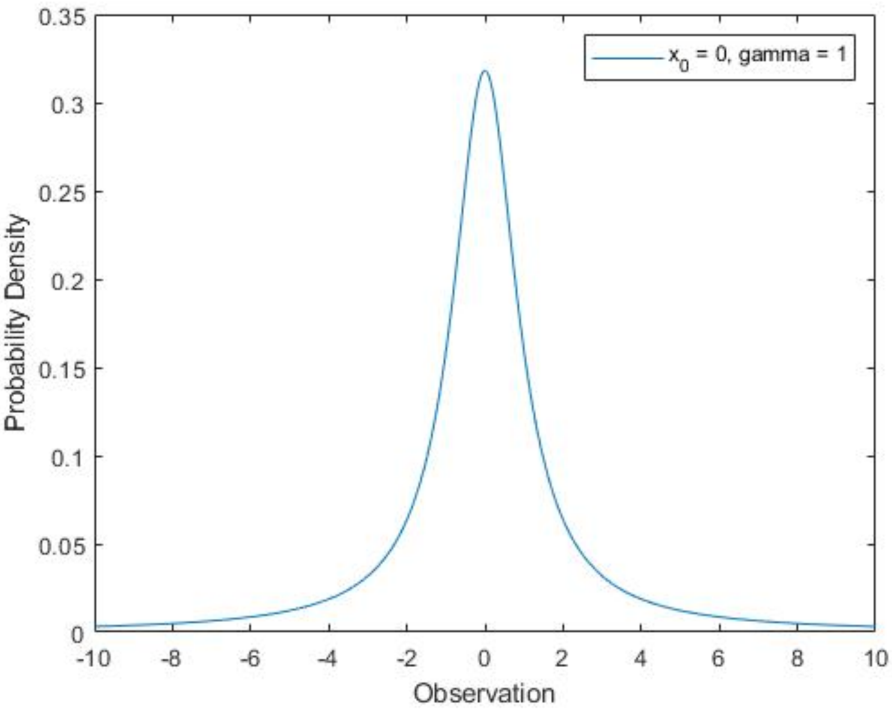
$k = \frac{1}{1000}$

$n = 10{,}000$

$k = \frac{1}{1000}$.

Example: Cauchy Distribution

- We consider X_i iid with $f_X(t) = \frac{1}{\pi t^2 + 1}$.
- Can show that $\bar{X}_n$ has pdf f_X for all n! This occurs because the Cauchy distribution does not have mean or variance.



Multivariate CLT

$\{X_i\}_{i\geq 1}$ iid sequence $E(X_i) = \mu$

$Var(X_i) = \sigma^2$

$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$.

Theorem

Let $\{X_i\}_{i\geq 1}$ be an iid sequence of $(K \times 1)$ random vectors with mean μ and finite variance matrix Σ. Then

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma\right).$$

Multivariate CLT

$\sqrt{n}(\bar{X}_n-\mu) = \frac{1}{\sqrt{n}} \sum (X_i-\mu)$.

$= \frac{1}{\sqrt{n}} \sum t'(X_i-\mu)$.

Proof.

By the Cramér-Wold Theorem, it suffices to show that for all $t \in \mathbb{R}^K$,

$$t'\left(\sqrt{n}\left(\bar{X}_n-\mu\right)\right) \xrightarrow{d} t'\mathcal{N}\left(0,\Sigma\right).$$

$Var(Ax) = A\,Var(x)A'$

Note that

$$t'\left(\sqrt{n}\left(\bar{X}_n-\mu\right)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t'\left(X_i-\mu\right),$$

$Var(t'(X_i-\mu)) = t'\Sigma t$

where $\{t'(X_i-\mu)\}_{i\geq 1}$ is an iid sequence with mean 0 and variance $t'\Sigma t$, so by the univariate CLT:

$$t'\left(\sqrt{n}\left(\bar{X}_n-\mu\right)\right) \xrightarrow{d} \mathcal{N}\left(0, t'\Sigma t\right),$$

$\overset{d}{=} t'N(0,\Sigma)$

which has the same distribution as $t'\mathcal{N}\left(0,\Sigma\right)$. □

Cramér Wold $\Rightarrow \sqrt{n}(\bar{X}_n-\mu) \to^d N(0,\Sigma)$.

Example

$$\frac{1}{S_n} \xrightarrow{a.s} \frac{1}{\sigma}$$

- Let $\{X_i\}_{i\geq 1}$ be an iid sequence of random variables with $\mathrm{E}(X_i) = \mu$ and $Var(X_i) = \sigma^2 > 0$.
- By the CLT
$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right).$$
- Suppose we wish to test $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ at significance level α.
  - Since σ^2 is typically unknown, how do we pick a test stat.?
  - Since the distribution of $\sqrt{n}\left(\bar{X}_n - \mu\right)$ is unknown, how do we pick a critical value?

Example

$\cancel{\frac{1}{n}\sum(X_i-\mu)^2}$

- To solve first issue: estimate σ^2 by

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \bar{X}_n^2\right).$$

- The denominator $1/(n-1)$ is used because

$$\mathrm{E}\left(S_n^2\right) = \sigma^2.$$

$\min_a \sum (X_i - a)^2 \leq \sum (X_i - \mu)^2$

- From the SLLN:

$a^* = \bar{X}_n$

$$\frac{1}{n}\sum_{i=1}^{n}X_i^2 \overset{a.s.}{\rightarrow} \mathrm{E}\left(X_i^2\right); \qquad \bar{X}_n \overset{a.s.}{\rightarrow} \mathrm{E}\left(X_i\right)$$

and by ordinary analysis, $\frac{n}{n-1} \rightarrow 1$. It follows that:

$$\left(\frac{n}{n-1}, \frac{1}{n}\sum_{i=1}^{n}X_i^2, \bar{X}_n\right) \overset{a.s.}{\rightarrow} \left(1, \mathrm{E}\left(X_i^2\right), \mathrm{E}\left(X_i\right)\right).$$

Example

$x \quad y \quad z \qquad x \quad y \quad z$

$S_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum X_i^2 - \bar{X}_n^2\right) \qquad E[X_i^2] - [E(X_i)]^2 > 0$

provided $Var(X_i) > 0$.

- By the continuous mapping theorem, with

$$g(x, y, z) = \frac{1}{\sqrt{x(y - z^2)}}$$

we obtain (since $\sigma^2 > 0$):

$$\frac{1}{S_n} = \frac{1}{\sqrt{S_n^2}} \xrightarrow{a.s.} \frac{1}{\sqrt{\sigma^2}} = \frac{1}{\sigma}.$$

$$\Rightarrow \frac{1}{S_n} \xrightarrow{P} \frac{1}{\sigma}.$$

Example

$X_n \to^d X$

$Y_n \to^d c$ (constant)

$\frac{X_n}{Y_n} \to^d \frac{X}{c}$ provided $c \neq 0$

- By Slutsky's theorem, since $\frac{1}{S_n} \xrightarrow{p} \frac{1}{\sigma}$, and $\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$, it follows that

$$\frac{1}{S_n} \times \sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \frac{1}{\sigma}\mathcal{N}\left(0, \sigma^2\right), \stackrel{d}{=} N(0,1).$$

  which has the standard normal distribution.

- Therefore, under H_0:

$$Y_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu_0\right)}{S_n} \xrightarrow{d} \mathcal{N}(0, 1).$$

$F_{Y_n}(x) \to \Phi(x)$ $\forall x \in \mathbb{R}$, Φ CDF of standard normal.

Example

$1 - F_{Y_n}(x) \to 1 - \Phi(x)$.

- By definition of $\xrightarrow{d}$:

Y_n

$$\mathrm{P}\left(\frac{\sqrt{n}\left(\bar{X}_n - \mu_0\right)}{S_n} > z_{1-\alpha}\right) \to \alpha,$$

$1 - \Phi(z_{1-\alpha})$

$1 - (1-\alpha) = \alpha$.

  where $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution.

- Therefore, comparing our test statistic with the $1-\alpha$ quantile of $\mathcal{N}(0,1)$ produces <u>asymptotically</u> correct null rejection probability, although the finite sample significance level of this test will usually <u>not</u> be α.

Questions?

Berry-Esseen Theorem

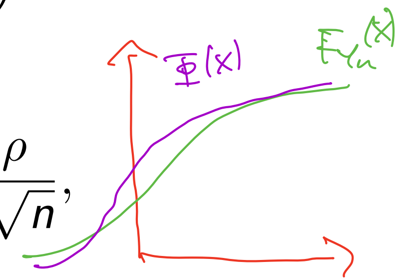
- We know that no fixed n is large enough to guarantee a good normal approximation. But.. the example used really highly skewed distributions. Berry-Esseen gives a finite sample bound on how far away from normality the distribution actually is:

CLT : $P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq t\right) \to \Phi(t)$

Theorem

(Berry-Esseen) Let $\{X_i\}_{i\geq 1}$ *be an iid sequence of random variables with* $\mathrm{E}(X_i) = \mu$, $0 < Var(X_i) < \infty$ *and* $\mathrm{E}\left(|X_i - \mu|^3\right) = \rho < \infty$. *Then*

$$\sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\overbrace{\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}}^{Y_n} \leq x \right) - \Phi(x) \right| \leq \frac{C\rho}{\sigma^3\sqrt{n}},$$



where Φ *is the standard normal CDF and* C *is a constant that does not depend on* n *or the distribution of the* X_i.

Example: Bernoulli Distribution

$\lesssim \frac{C\rho}{\sigma^3} \cdot \frac{1}{\sqrt{n}}$.

$C \leq 0.47...$

$C_E \geq 0.409...$

- If $X_i \sim Bernoulli(p)$, then $\mathrm{E}(X_i) = p$, $Var(X_i) = p(1-p)$, and

$$\frac{\rho}{\sigma^3} = \mathrm{E}\left(\left|\frac{X_i - \mathrm{E}(X_i)}{\sqrt{Var(X_i)}}\right|^3\right) = \frac{p^2 + (1-p)^2}{\sqrt{p(1-p)}}.$$

$p \approx 1$

$\approx \frac{1+0}{\sqrt{1(0)}}$

- Berry-Esseen theorem tells us that

$$\sup_{x \in \mathbb{R}} \left| \mathrm{P}\left(\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \leq x\right) - \Phi(x)\right| \leq C\frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}.$$

Tightness

$\frac{1}{\sqrt{n}} \sum \text{Terms}_i = \frac{1}{\sqrt{n}} \sum \text{iid Terms}_i + \text{Stuff}.$

Hope $\text{Stuff} \xrightarrow{p} 0$.

$X_n \xrightarrow{d} X \quad Y_n \xrightarrow{p} 0 \quad X_n + Y_n \xrightarrow{d} X.$

- There are cases in which a sequence of random vectors does not have a limiting distribution, but it does remain "bounded" in a probabilistic sense, so that none of the probability mass escapes to $+\infty$ or $-\infty$.

$-1, 1, -1, 1 \ldots\ldots$

- For example, $X_n = (-1)^n$ is a sequence of constant random variables that does not converge in distribution, but for all n:

$$P(|X_n| \leq 1) = 1.$$

$0 \leq a_n \leq \frac{K}{n}$

Tightness

- We call a sequence of random vectors $\{X_n\}_{n\geq 1}$ tight if for all $\epsilon > 0$ there exists $B_\epsilon > 0$ such that for all n,

$$\mathrm{P}\left(\|X_n\| \leq B_\epsilon\right) \geq 1 - \epsilon.$$

- Tightness requires that for any ϵ, the entire sequence can be contained in a ball of radius B_ϵ with probability at least $1-\epsilon$, so a tight sequence is sometimes called "bounded in probability".

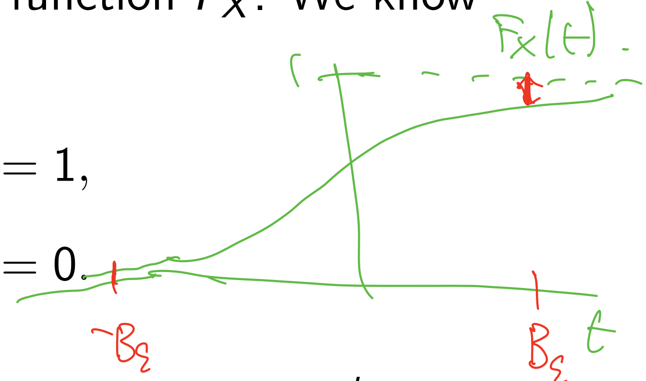
Tightness

Sum of two red distances is less than ε.

- Suppose $X_n \xrightarrow{d} X$, with distribution function F_X. We know that

$$\lim_{t\to+\infty} F_X(t) = 1,$$

$$\lim_{t\to-\infty} F_X(t) = 0.$$



It follows that if $X_n = X$ for all n, then trivially $X_n \xrightarrow{d} X$, but also $\exists B_\epsilon$ such that for any $\epsilon > 0$:

$$\mathrm{P}(|X_n| \leq B_\epsilon) = \mathrm{P}(|X| \leq B_\epsilon) \geq 1 - \epsilon,$$

since B_ϵ can be chosen large enough that

$$F_X(B_\epsilon) - F_X(-B_\epsilon) \geq 1 - \epsilon.$$

- In fact, convergence in distribution always implies tightness.

Tightness

Theorem

If $\{X_n\}_{n\geq 1}$ is a sequence of random vectors such that $X_n \xrightarrow{d} X$, then $\{X_n\}_{n\geq 1}$ is tight.

Proof.

(For random variables). Suppose $X_n \xrightarrow{d} X$, with distribution function F_X. We know that $\exists B_\epsilon > 0$ such that

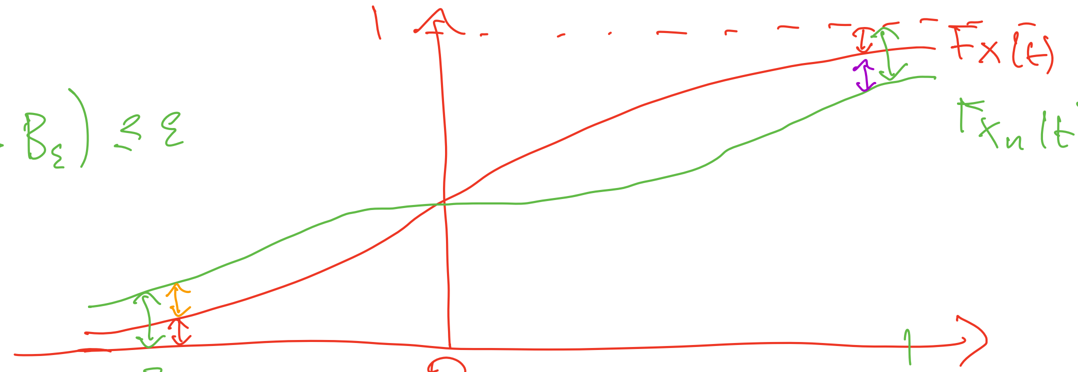
$$\mathrm{P}\left(|X| > B_\epsilon\right) \leq \epsilon/3,$$

where $B_\epsilon, -B_\epsilon$ are continuity points of F_X. We have $F_{X_n}\left(B_\epsilon\right) \to F_X\left(B_\epsilon\right)$, so choose n large enough that

$$\begin{aligned} \left|F_{X_n}\left(B_\epsilon\right) - F_X\left(B_\epsilon\right)\right| &< \epsilon/3, \\ \left|F_{X_n}\left(-B_\epsilon\right) - F_X\left(-B_\epsilon\right)\right| &< \epsilon/3. \end{aligned}$$

Tightness

NTS: $P(\|X_n\| > B_\varepsilon) \leq \varepsilon$



Proof.

Now note that

$$
\begin{aligned}
\mathrm{P}\left(|X_n| > B_\epsilon\right) &= \mathrm{P}\left(X_n > B_\epsilon\right) + \mathrm{P}\left(X_n < -B_\epsilon\right) \\
&= \mathrm{P}\left(X_n > B_\epsilon\right) - \mathrm{P}\left(X > B_\epsilon\right) \\
&+ \mathrm{P}\left(X > B_\epsilon\right) + \mathrm{P}\left(X < -B_\epsilon\right) \\
&+ \mathrm{P}\left(X_n < -B_\epsilon\right) - \mathrm{P}\left(X < -B_\epsilon\right) \\
&\leq \left|1 - F_{X_n}(B_\epsilon) - \left[1 - F_X(B_\epsilon)\right]\right| \\
&+ \mathrm{P}\left(|X| > B_\epsilon\right) \\
&+ \left|F_{X_n}(-B_\epsilon) - F_X(-B_\epsilon)\right| \\
&< \epsilon.
\end{aligned}
$$

Need to make sure sum of green distances is less than ε.

(1) Choose B_ε large enough such that sum of red distances $< \frac{\varepsilon}{3}$.

(2) Choose n large enough such that purple distance $< \frac{\varepsilon}{3}$.

(3) Choose n large enough: orange distance $< \frac{\varepsilon}{3}$.

(1) possible because X is a random variable

(2) + (3) because $X_n \xrightarrow{d} X$.

$\square$

Stochastic order notation

$\frac{1}{\sqrt{n}} \sum \text{terms}_i = \frac{1}{\sqrt{n}} \sum \text{iid terms}_i + o_p(1).$

- Let $\{x_n\}_{n\geq 1}, \{y_n\}_{n\geq 1}$ be sequences of real numbers.
- Write $x_n = o(y_n)$ if

$$\left|\frac{x_n}{y_n}\right| \to 0 \text{ as } n \to \infty.$$

$x_n = n \quad y_n = n^2$

$x_n = o(y_n).$

- Write $x_n = O(y_n)$ if $\exists B < \infty$ such that for all n:

$$\left|\frac{x_n}{y_n}\right| \leq B.$$

$x_n = n \quad y_n = a n$

for any a satisfies

$x_n = O(y_n).$

- Note that $x_n = o(y_n) \implies x_n = O(y_n)$.

$\because$ Convergent sequence is bdd.

Given $\varepsilon > 0$ $\exists N : \forall n \geq N$, $\left|\frac{x_n}{y_n}\right| \leq \varepsilon$

Stochastic order notation

So $\forall n$: $\left|\frac{X_n}{Y_n}\right| \le \max\left\{\varepsilon, \left|\frac{X_1}{Y_1}\right|, \dots, \left|\frac{X_N}{Y_N}\right|\right\}$
$\underbrace{\phantom{\left|\frac{X_1}{Y_1}\right|, \dots}}_{\| \; B}$

- Let $\{X_n\}_{n\geq1}, \{Y_n\}_{n\geq1}$ be sequences of random variables.
- Write $X_n = o_p(Y_n)$ if

$$\left|\frac{X_n}{Y_n}\right| \xrightarrow{p} 0.$$

$\forall \varepsilon > 0$, $\exists B_\varepsilon > 0$ such that $\forall n$:

- Write $X_n = O_p(Y_n)$ if ~~$\exists B_\epsilon < \infty$ such that for all n and $\epsilon > 0$:~~

$\Leftrightarrow$ $\left|\frac{X_n}{Y_n}\right|$ is tight.

$$\mathrm{P}\left(\left|\frac{X_n}{Y_n}\right| \leq B_\epsilon\right) \geq 1 - \epsilon.$$

- This last condition just says that $|X_n/Y_n|$ is tight.
- Note that $X_n = o_p(Y_n) \implies X_n = O_p(Y_n)$.

$X_n = o_p(Y_n) \Leftrightarrow \left|\frac{X_n}{Y_n}\right| \to^p 0 \Rightarrow \left|\frac{X_n}{Y_n}\right| \to^d 0 \Rightarrow \left|\frac{X_n}{Y_n}\right| = O_p(1)$

Example

$\Rightarrow X_n = O_p(Y_n)$

- $\{X_n\}_{n\geq 1}$ is tight is equivalently written as $X_n = O_p(1)$.
- $X_n \xrightarrow{p} 0$ is equivalently written as $X_n = o_p(1)$.
- Some properties:
  - $o_p(1) + o_p(1) = o_p(1)$
  - $o_p(1) + O_p(1) = O_p(1)$
  - $o_p(1)\, O_p(1) = o_p(1)$
  - $(A + o_p(1))^{-1} = O_p(1)$ for an invertible matrix A.
- By Slutsky's Theorem, if $X_n \xrightarrow{d} X$ and $Y_n = o_p(1)$, then $X_n + Y_n \xrightarrow{d} X$.

$y_i = x_i'\beta + u_i$; $E(x_i u_i) = 0$.

$$\hat{\beta}^{OLS} - \beta = \left(\frac{1}{n}\sum x_i x_i'\right)^{-1} \frac{1}{n}\sum x_i u_i$$

$$\left(\frac{1}{n}\sum x_i x_i'\right)^{-1} \xrightarrow{p} \left[E(x_i x_i')\right]^{-1} \qquad \frac{1}{n}\sum x_i u_i \xrightarrow{p} E(x_i u_i) = 0.$$

Can also write

$$\left(\frac{1}{n}\sum x_i x_i'\right)^{-1} \frac{1}{n}\sum x_i u_i$$

$$= \left(E(x_i x_i') + o_p(1)\right)^{-1} \cdot \frac{1}{n}\sum x_i u_i$$

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

Proof: $o_p(1)\, O_p(1) = o_p(1)$

- Let $X_n = o_p(1)$ and $Y_n = O_p(1)$. Need to show

  $\overset{?}{\Rightarrow}$: $\mathrm{P}(|X_n Y_n| > \epsilon) \to 0$ for any $\epsilon > 0$.
- Choose $\gamma > 0$ such that $\mathrm{P}(|Y_n| \geq B_\gamma) \leq \gamma$. $\Leftarrow Y_n = O_p(1)$
- Write

$$
\begin{aligned}
|X_n Y_n| &= |X_n Y_n| \mathbf{1}_{|Y_n| < B_\gamma} + |X_n Y_n| \mathbf{1}_{|Y_n| \geq B_\gamma} \\
&\leq B_\gamma |X_n| + |X_n Y_n| \mathbf{1}_{|Y_n| \geq B_\gamma}.
\end{aligned}
$$

- It follows that $|A_n| + |B_n| > \varepsilon \Rightarrow |A_n| > \frac{\varepsilon}{2}$ or $|B_n| > \frac{\varepsilon}{2}$.

$$
\begin{aligned}
\mathrm{P}(|X_n Y_n| > \epsilon) &\leq \mathrm{P}\left(B_\gamma |X_n| + \left|X_n Y_n \mathbf{1}_{|Y_n| \geq B_\gamma}\right| > \epsilon\right) \leq \mathrm{P}\left(\{B_\gamma |X_n| > \tfrac{\varepsilon}{2}\} \cup \{|X_n Y_n \mathbf{1}| > \tfrac{\varepsilon}{2}\}\right) \\
&\leq \mathrm{P}(B_\gamma |X_n| > \epsilon/2) + \underbrace{\mathrm{P}\left(\left|X_n Y_n \mathbf{1}_{|Y_n| \geq B_\gamma}\right| > \epsilon/2\right)}_{\leq \mathrm{P}(|Y_n| \geq B_\gamma) \leq \gamma.} \\
&\leq o(1) + \gamma.
\end{aligned}
$$

$\downarrow$ 0

Since $\gamma > 0$ was chosen arbitrarily, $\mathrm{P}(|X_n Y_n| > \epsilon) \to 0$.

$X_n Y_n \xrightarrow{p} 0.$

Questions?

Delta Method

- So far we have developed the CLT, SLLN and CMT, now add Delta Method.
- CLT gives:

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right).$$

- CMT gives:

$$g\left(\sqrt{n}\left(\bar{X}_n - \mu\right)\right) \xrightarrow{d} g\left(\mathcal{N}\left(0, \sigma^2\right)\right)$$

- What can we say about

$$\sqrt{n}\left(g\left(X_n\right) - g\left(\mu\right)\right)?$$

$$\sqrt{n}\left(g(\bar{X}_n) - g(\mu)\right) \approx \sqrt{n}\left(g'(\mu)(\bar{X}_n - \mu)\right)$$

$$= g'(\mu)\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N\left(0, \sigma^2 g'(\mu)^2\right).$$

Delta Method

Theorem

Let $\{X_n\}_{n\geq 1}$ *be a sequence of* $(K \times 1)$ *random vectors and suppose that*

$$n^r (X_n - c) \xrightarrow{d} X$$

for some $r > 0$ *and constant vector* c. *Let* $g : \mathbb{R}^K \to \mathbb{R}^d$ *be differentiable at the point* c. *Let* $Dg(c)$ *be the* $d \times k$ *matrix of partial derivatives evaluated at* c. *Then*

$$n^r (g(X_n) - g(c)) \xrightarrow{d} Dg(c) X.$$

In particular, if $X \sim \mathcal{N}(0, \Sigma)$, *then*

$$n^r (g(X_n) - g(c)) \xrightarrow{d} \mathcal{N}\left(0, Dg(c) \Sigma Dg(c)'\right).$$

Delta Method

- $Dg\left(c\right)$ is the following matrix of partial derivatives:

$$Dg\left(c\right)=\begin{bmatrix}\frac{\partial g_1}{\partial x_1}\left(c\right) & \frac{\partial g_1}{\partial x_2}\left(c\right) & \cdots & \frac{\partial g_1}{\partial x_k}\left(c\right)\\ \frac{\partial g_2}{\partial x_1}\left(c\right) & \frac{\partial g_2}{\partial x_2}\left(c\right) & \cdots & \frac{\partial g_2}{\partial x_k}\left(c\right)\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial g_d}{\partial x_1}\left(c\right) & \frac{\partial g_d}{\partial x_2}\left(c\right) & \cdots & \frac{\partial g_d}{\partial x_k}\left(c\right)\end{bmatrix}.$$

Delta Method

Proof.

By Taylor's theorem:

$$g(x) = g(c) + Dg(c)(x - c) + h_1(x)(x - c),$$

for some function $h_1(x)$ with $\lim_{x \to c} h_1(x) = h_1(c) = 0$. It follows that

$$n^r(g(X_n) - g(c)) = Dg(c)\, n^r(X_n - c) + h_1(X_n)\, n^r(X_n - c).$$

Since $n^r(X_n - c) \xrightarrow{d} X$, by Slutsky's Theorem:

$$Dg(c)\, n^r(X_n - c) \xrightarrow{d} Dg(c)\, X.$$

Now show that $h_1(X_n)\, n^r(X_n - c) = o_p(1)$:

Delta Method

Proof.

Since $n^r(X_n - c) \xrightarrow{d} X$, we have $n^r(X_n - c) = O_p(1)$ and $X_n \xrightarrow{p} c$ (See Problem Set 3). Since h_1 is continuous at c by construction,

$$h_1(X_n) \xrightarrow{p} h_1(c) = 0$$

by the CMT. Therefore,

$$h_1(X_n)\, n^r(X_n - c) = o_p(1) \cdot O_p(1) = o_p(1),$$

so

$$\begin{aligned} n^r(g(X_n) - g(c)) &= Dg(c)\, n^r(X_n - c) + o_p(1) \\ &\xrightarrow{d} Dg(c)\, X. \end{aligned}$$

□

Delta Method

- Note that if $Dg(c) = 0$,

$$n^r (g(X_n) - g(c)) \xrightarrow{d} 0,$$

which is a degenerate limiting distribution, (so $\xrightarrow{p} 0$ also).

- If g has higher order derivatives, we can derive an alternate form of the Delta Method when $Dg(c) = 0$.
- Suppose $\{X_n\}_{n\geq 1}$ is a sequence of random variables and $g : \mathbb{R} \to \mathbb{R}$ has 2 derivatives.
- Taylor's theorem implies

$$g(x) = g(c) + g'(c)(x - c) + \frac{g''(c)}{2}(x - c)^2 + h_2(x)(x - c)^2,$$

where $\lim_{x\to c} h_2(x) = h_2(c) = 0$.

Delta Method

- Repeating the argument in the proof of the original Delta Method:

$$n^{2r}\left(g\left(X_n\right)-g\left(c\right)\right)=g'\left(c\right)n^{2r}\left(X_n-c\right)+\frac{g''\left(c\right)}{2}n^{2r}\left(X_n-c\right)^2$$
$$+h_2\left(X_n\right)n^{2r}\left(X_n-c\right)^2.$$

- We have

$$\begin{aligned} g'\left(c\right)n^{2r}\left(X_n-c\right)&=0\\ \frac{g''\left(c\right)}{2}n^{2r}\left(X_n-c\right)^2&\xrightarrow{d}\frac{g''\left(c\right)}{2}X^2\\ h_2\left(X_n\right)n^{2r}\left(X_n-c\right)^2&=o_p\left(1\right)O_p\left(1\right)=o_p\left(1\right).\end{aligned}$$

- In summary, if $g'\left(c\right)=0$,

$$n^{2r}\left(g\left(X_n\right)-g\left(c\right)\right)\xrightarrow{d}\frac{g''\left(c\right)}{2}X^2.$$

Example: Sample variance

- We will tie together the concepts we have learned so far to find the asymptotic distribution of

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

- Let X_i be iid random variables with $\mathrm{E}(X_i) = \mu$, $Var(X_i) = \sigma^2$ and $\mathrm{E}(X_i - \mu)^4 = \kappa$.
- We are looking for an $r > 0$, constant c and random variable X such that

$$n^r\left(S_n^2 - c\right) \xrightarrow{d} X,$$

for some non-degenerate X.

Example: Sample variance

- We have already shown that $S_n^2 \xrightarrow{p} \sigma^2$, so we must take $c = \sigma^2$.
- Unfortunately, S_n^2 is not in a form where we can apply the CLT directly:

$$n^{1/2}\left(S_n^2 - \sigma^2\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left(X_i - \bar{X}_n\right)^2 - \sigma^2\right].$$

- If we could replace $\bar{X}_n$ with μ, would consider

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[(X_i - \mu)^2 - \sigma^2\right].$$

Example: Sample variance

- Note that $\left\{(X_i - \mu)^2\right\}_{i \geq 1}$ is an iid sequence with $\mathrm{E}\,(X_i - \mu)^2 = \sigma^2$ and

$$\begin{aligned} Var\left((X_i - \mu)^2\right) &= \mathrm{E}\left[(X_i - \mu)^4\right] - \left[\mathrm{E}\,(X_i - \mu)^2\right]^2 \\ &= \kappa - \sigma^4. \end{aligned}$$

- Therefore, by the CLT:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[(X_i - \mu)^2 - \sigma^2\right] \xrightarrow{d} \mathcal{N}\left(0, \kappa - \sigma^4\right).$$

- It remains to show that

$$n^{1/2}\left(S_n^2 - \sigma^2\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[(X_i - \mu)^2 - \sigma^2\right] + o_p\,(1)\,.$$

Example: Sample variance

- Note that

$$
\begin{aligned}
\sqrt{n}\left(S_n^2 - \sigma^2\right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left(X_i - \bar{X}_n\right)^2 - \sigma^2\right] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left(X_i - \mu - \left(\bar{X}_n - \mu\right)\right)^2 - \sigma^2\right] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[(X_i - \mu)^2 - \sigma^2\right] + \sqrt{n}\left(\bar{X}_n - \mu\right)^2 \\
&- 2\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[(X_i - \mu)\left(\bar{X}_n - \mu\right)\right] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[(X_i - \mu)^2 - \sigma^2\right] - \sqrt{n}\left(\bar{X}_n - \mu\right)^2 .
\end{aligned}
$$

Example: Sample variance

- Finally, since

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right),$$
$$\left(\bar{X}_n - \mu\right) = o_p(1),$$

we get

$$\sqrt{n}\left(\bar{X}_n - \mu\right)^2 = o_p(1)\, O_p(1) = o_p(1).$$

- In summary:

$$n^{1/2}\left(S_n^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}\left(0, \kappa - \sigma^4\right).$$

Example: Sample variance

- We are not done yet: The limiting distribution is non-degenerate iff
$$\kappa - \sigma^4 > 0.$$
- Jensen's inequality gives
$$\kappa = \mathrm{E}\left[(X_i - \mu)^4\right] \geq \left[\mathrm{E}\,(X_i - \mu)^2\right]^2 = \sigma^4,$$
with equality if and only if the random variable $(X_i - \mu)^2$ is constant almost surely.
- In this case, this does NOT imply X_i is constant, since
$$(X_i - \mu)^2 = \sigma^2$$
will have two solutions: $X_i = \mu \pm \sigma$ when $\sigma > 0$.