Asymptotics: - Convergence in probability

$$Pr(|X_N - X| < \varepsilon) \to 1 \text{ as } N \to \infty$$

- Other types: - Convergence in distribution (weak convergence)

- $\{X_n\}_{n=1}^{\infty}$ converge in distribution to a random variable Z distributed as $F_Z(\ )$

$X_n \xrightarrow{d} Z$ or $X_n \xrightarrow{d} F_Z(\ )$

if $\quad E f(X_n) \to E f(Z)$ (as $N \to \infty$)

for all $f \in \mathcal{F}$

$\mathcal{F}$ can be one of several equivalent classes of functions

- Examples

1 - $\mathcal{F} = C^3$: continuous functions with 3 bounded continuous derivatives

2 - $\mathcal{F} = \{\mathbb{1}\{x \le t\}\ \forall t \text{ s.t. } P(Z = t) = 0\}$

CDF $\quad F_{X_n}(t) \to F_Z(t) \quad \forall t$ s.t. F_Z is continuous at t

3 - $\mathcal{F} = \{e^{ixt};\ \forall t \in \mathbb{R}\} \quad i = \sqrt{-1}$

$$\underbrace{E e^{iX_n t}}_{\text{Characteristic function}} \to E e^{iZt} \quad \forall t$$

Equivalence of 2 and 3 is called Lévy's continuity theorem

Application: $x_n \xrightarrow{d} Z$ implies $\Pr(x_n \in E) \to \Pr(Z \in E)$ for all E

Allows us to build confidence intervals and tests

Central Limit Theorem

- show convergence in distribution for sums of independent random variables
- Simple version: Let $\{X_i\}_{i=1}^{\infty}$ have $EX_i = M$, $Var(X_i) = \sigma^2$ and be independent and identically distributed

Then $\sqrt{n}\left(\frac{\frac{1}{n}\Sigma X_i - M}{\sigma}\right) \xrightarrow{d} Z \sim N(0, 1)$

pdf of Z is $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

- Triangular Array Central Limit Theorem

Let $\{x_{n,i}\}_{i=1}^{n}$ be a sequence of independent real random variables for each n

$X_{1,1}$

$X_{2,1}, X_{2,2}$

$X_{3,1}, X_{3,2}, X_{3,3}$

- Assume

1. $\Sigma_{i=1}^{n} Ex_{n,i} \to M$ finite
2. $\Sigma_{i=1}^{n} Var(X_{n,i}) \to \sigma^2$ finite
3. $\Sigma_{i=1}^{n} E|X_{n,i}|^3 \to 0$

- Then $\sum_{i=1}^{n} X_{n,i} \xrightarrow{d} X \sim N(\mu, \sigma^2)$

- $N(\mu, \sigma^2)$ has a density $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Implies $P_r\left(\left[\sum_i X_{n,i} - \sigma c_{1-\alpha/2}, \sum_i X_{n,i} - \sigma c_{\alpha/2} \right) \text{ contains } \mu \right) \to 1-\alpha$

$P_r\left(\sum X_{n,i} < \mu + \sigma c_{1-\alpha/2} \right) \to P_r\left(X < \mu + \sigma c_{1-\alpha/2} \right)$

$P_r\left(\sum X_{n,i} > \mu + \sigma c_{\alpha/2} \right) \to P_r\left(X < \mu + \sigma c_{\alpha/2} \right)$

If c_t is value such that $P_r(X \leq c_t) = t$ for $X \sim N(0,1)$

Proof Sketch

If $f \in C^3$ Taylor expansion gives

$$f(x+y) = f(x) + y f'(x) + \frac{1}{2} y^2 f''(x) + \frac{1}{6} y^3 f'''(x^*)$$

for some $x^* \in [x, x+y]$. By boundedness

$$\left| \frac{1}{6} y^3 f'''(x^*) \right| \leq C |y|^3$$

Let X and Y be independent random variables.
Take expectations

$$\left| E f(X+Y) - E f(X) - E[Y] E[f'(X)] \right.$$

$$-\frac{1}{2}\, E[Y^2]\, E[f''(X)]\,\Big| \le C\, E|Y|^3$$

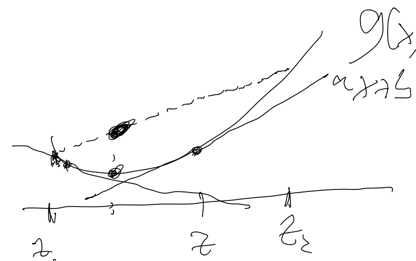
- Let $Z = \mu + \sigma W$ with $W \sim N(0,1)$ independent of X, Y where $\mu = EY$, $\sigma^2 = Var(Y)$

Then $$\big| E f(X+Y) - E f(X+Z) \big| \le C\left(E|Y|^3 + E|Z|^3\right)$$

- Can show $E|Z|^3 \le C_1 E|Y|^3$

Uses Jensen's inequality

Let g be a convex function



At every z $\exists a, b$ $\quad g(x) \ge ax + b$ for all x

and $g(z) = az + b$

Then $g(E[Z]) \le E\, g(Z)$ for any random variable

Pf: Let z_0 be EZ then $\quad g(Z) \ge aZ + b$ for a, b the subgradient

so $$E\, g(Z) \ge E\, aZ + b$$
$$= a\,EZ + b$$
$$= g(z_0) = g(EZ)$$

$$E|\mu + \sigma W|^3 \le E(|\mu| + \sigma|W|)^3 \quad \text{triangle inequality}$$
$$\le E\left(2 \max\{|\mu|, \sigma|W|\}\right)^3$$

$$\leq 8\left(E|M|^3 + \sigma^3|W|^3\right)$$

$$\leq 8\left(E|Y|^3 + (E|Y|^2)^{\frac{3}{2}} E|W|^3\right)$$

$\mathrm{Var}(Y)$
$= EY^2 - (EY)^2$
$\leq E|Y|^2$

$$\leq \Big(8 + 8\underbrace{E|W|^3}_{\text{about } 1.6}\Big) E|Y|^3$$

$$\leq C_1 E|Y|^3$$

Let S_n be $\sum_{i=1}^n X_{n,i}$, $T_n = \sum_{i=1}^n \eta_{n,i}$

with $\eta_{n,i} \overset{\text{independent}}{\sim} N(EX_{n,i}, \mathrm{Var}(X_{n,i}))$

- $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + ab\,\underbrace{\mathrm{Cov}(X,Y)}_{\underbrace{EXY - EXE(Y)}_{=0 \text{ if } X, Y \text{ are independent}}}$

- $T_n \sim N\left(\sum_{i=1}^n EX_{n,i}, \sum_{i=1}^n \mathrm{Var}(X_{n,i})\right)$

- Let $X_{n,j} = \sum_{i=1}^{j-1} X_{n,i} + \sum_{i=j+1}^{n} \eta_{n,i}$

$Y_{n,j} = X_{n,j} \quad Z_{n,j} = \eta_{n,j}$

$$\left| Ef(X_{n,j} + Y_{n,j}) - Ef(X_{n,j} + Z_{n,j}) \right| \leq \tilde{C} E|X_{n,j}|^3$$

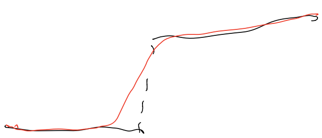
Apply n times from $j=n$ to $j=1$

$$|E f(S_n) - E f(T_n)| \le C\left(\sum_{i=1}^{n} E|X_{n,i}|^3\right) \to 0$$

- Can show $T_n \xrightarrow{d} T$ where $T \sim N(\mu, \sigma^2)$ so $|Ef(T_n) - Ef(T)| \to 0$

So $|Ef(S_n) - Ef(T)| \to 0 \quad \forall f \in C^3$

So $S_n \xrightarrow{d} T$ //



Result: usual version

Let $X_i \sim$ iid $EX_i = \mu$, $Var(X_i) = \sigma^2 < \infty$,

$E|X_i|^3 < \infty$

Apply above with $X_{n,i} = \dfrac{X_i - \mu}{\sqrt{n}}$

Then $\sum_{i=1}^{n} E X_{n,i} = 0$, $\sum_{i=1}^{n} E X_{n,i}^2 = \sigma^2$

$$\sum_{i=1}^{n} E|X_{n,i}|^3 = \frac{1}{n^{3/2}} \sum_{i=1}^{n} E|X_i|^3 = \frac{1}{n^{1/2}} E|X_i - \mu|^3 \to 0$$

So by above $X_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ has

$$\sqrt{n}(X_n - \mu)/\sigma \xrightarrow{d} X \sim N(0, 1)$$

- Can drop bounded 3rd moment condition in iid case
  - Rough argument: apply triangular array version to $\left(\frac{X_i - \mu}{\sigma\sqrt{n}}\right)\mathbb{1}\{\left(\frac{X_i-\mu}{\sigma} \le \sqrt{n}\right)\}$ and bound error from tails.

Stronger results

Can relax assumptions or strengthen to get stronger results

- Multivariate: if $X_n \in \mathbb{R}^k$ iid, $EX_i = \mu$

  $Var(X_i) = E X_i X_i' - EX_i EX_i' = \Sigma < \infty$

  $\Sigma_{jk} = Cov(X_{ij}, X_{ik})$

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

  multivariate normal $a'Z \sim N(0, a'\Sigma a)$ for any $a \in \mathbb{R}^k$

- In independent non identically distributed case with $EX_i^{(2+\delta)} < \infty$ for $\delta > 0$, still holds when standardized by average mean and variance

- Can weaken independence
- Can get stronger modes of convergence with more finite moments ie $EX_i^k < \infty$ for $k > 2$

eg: Berry Esseen theorem for $k=3$

— CLT useful for us because it is what we can show for estimators we use (OLS, IV, GMM)

Extension: Continuous Mapping Theorem:

If $X_n \xrightarrow{d} X$ where $\{X_n\}_{n=1}^{\infty}$ is a sequence of $k \times 1$ random vectors

If $g: \mathbb{R}^k \to \mathbb{R}^J$ is a continuous function on a set C such that $P(X \in C) = 1$,

Then $g(X_n) \xrightarrow{d} g(X)$

Rough argument: $Ef(g(X_n)) \to Ef(g(X))$

by composition of f and g for all $f \in \mathcal{F}$

This lets us apply convergence in distribution to functions of converging statistics.

Next topic: Estimation and Ordinary Least Squares.

Suppose we have data from a cross section of firms

data: $\{y_i, k_i, l_i\}_{i=1}^{N}$

y_i is log output, k_i is log capital

l_i is log labor

Want to know the production process.

Standard model: Cobb Douglas production function

$$\exp(y_i) = \exp(k_i)^{\alpha} \exp(l_i)^{\beta}$$

Want to know α, β.

Different firms with same k_i, l_i may have different y_i.

So need another term, not observable. call it ε_i

$$\exp(y_i) = \exp(\varepsilon_i) \exp(k_i)^{\alpha} \exp(l_i)^{\beta}$$