Convergence in Distribution and the Central Limit Theorem

Let $Y_1, Y_2, \ldots, Y_n$ be some sequence of statistics. For example, let $X_1, \ldots, X_n$ be iid r.v.'s with $E[X_i] = \mu$ and $var(X_i) = \sigma^2$. Define

$$Y_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \text{"z-statistic"}$$

We say that Y_n <u>converges in distribution</u> to a r.v. W and write $Y_n \xrightarrow{d} W$ if

$$\underbrace{Pr(Y_n \le y)}_{\text{CDF of } Y_n} \longrightarrow \underbrace{Pr(W \le y)}_{\text{CDF of } W} \quad \text{as } n \to \infty$$

<u>Note:</u> In most cases, W is either a normal r.v. or a χ^2 r.v.

<u>Theorem</u> (Lindeberg-Levy CLT)

Let $X_1, \ldots, X_n$ be iid r.v.'s with $E[X_i] = \mu$ and $var(X_i) = \sigma$

Then

$$Y_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim N(0,1) \quad \text{as } n \to \infty$$

(Note: $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \sqrt{n}\left(\frac{\bar{x}-\mu}{\sigma}\right)$)

That is, for all $y \in \mathbb{R}$

$$Pr(Y_n \leq y) \longrightarrow \Phi(y) \quad \text{as } n \to \infty$$

CDF of Standard Normal r.v.

<u>Remark</u> The CLT suggests that we may approximate the distribution of $Y_n = \bar{x}-\mu/\sigma/\sqrt{n}$ by that of $Z \sim N(0,1)$. This, in turn, suggests approximating the distribution of $\bar{x}$ by $W \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. Notationally, we write

$$\sqrt{n}\left(\frac{\bar{x}-\mu}{\sigma}\right) \overset{A}{\sim} N(0,1)$$

and

$$\bar{x} \overset{A}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Not practically useful unless we know σ^2!

where "$\overset{A}{\sim}$" denotes "asymptotically distributed as" and is interpreted as an approximating distribution in finite samples.

Results for the Manipulation of CLT results

1. Suppose

$$Y_n \xrightarrow{d} W \qquad (W \text{ is a r.v.})$$

$$Z_n \xrightarrow{P} c \qquad (c \text{ is a constant})$$

Then

(a) $Z_n \cdot Y_n \xrightarrow{d} c \cdot W$

(b) $\dfrac{Y_n}{Z_n} \xrightarrow{d} \dfrac{W}{c}$ provided $c \neq 0$

(c) $Y_n + Z_n \xrightarrow{d} W + c$

Example. Let $X_1, \ldots, X_n$ be iid with $E[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. In most practical situations we don't know σ^2 and we must estimate it from the data. Let

$$S^2 = \frac{1}{n} \sum_1^n (X_i - \bar{X})^2$$

From the previous example we know that

$$S^2 \xrightarrow{P} \sigma^2 \quad \text{and} \quad S \xrightarrow{P} \sigma$$

In addition, from the Lindeberg-Levy CLT we know that

$$Y_n = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim N(0,1).$$

Now consider

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} = \underbrace{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)}_{\xrightarrow{d} N(0,1)} * \underbrace{\left(\frac{\sigma}{s}\right)}_{\xrightarrow{p} \frac{\sigma}{\sigma} = 1} \xrightarrow{d} Z \cdot 1 = Z \sim N(0,1)$$

~~Using~~ The above used the following results:

(i) $s^2 \xrightarrow{p} \sigma^2 \Rightarrow s \xrightarrow{p} \sigma \Rightarrow \frac{1}{s} \xrightarrow{p} \frac{1}{\sigma}$

via the LLNs and Slutsky's theorem.

(ii) $Z_n = \frac{\sigma}{s} \xrightarrow{p} 1$

(iii) $Y_n \cdot Z_n \xrightarrow{d} Z \cdot 1 = Z \sim N(0,1)$

<u>Remark</u> The above example shows that

$$\sqrt{n}\left(\frac{\bar{X}-\mu}{s}\right) \overset{A}{\sim} N(0,1)$$

$\Rightarrow \cancel{\bar{X} \overset{A}{\sim} N(\mu, s^2/n)}$ $\frac{s^2}{n}$ = consistent estimate of asymptotic variance

$\Rightarrow \quad \sqrt{n}(\bar{X}-\mu) \overset{A}{\sim} N(0, s^2)$

$\Rightarrow \quad S^2$ is a consistent estimate of the asymptotic variance of $\sqrt{n}(\bar{X}-\mu)$;
i.e $\quad S^2 \xrightarrow{P} \sigma^2$

Going further, we have

$$\bar{X} \overset{A}{\sim} N\left(\mu, \frac{S^2}{n}\right) \longleftarrow$$ Practically useful result.

where $\frac{S^2}{n}$ is an estimate of the asymptotic variance of $\bar{X}$.

CLT Relevant for the Linear Regression model

Theorem (Lindeberg - Feller)

Let $X_1, \ldots, X_n$ be independent (not necessarily identically distributed) r.v's with

(a) $E[X_i] = \mu_i < \infty$

(b) $var(X_i) = \sigma_i^2 < \infty$

$$\bar{\mu}_n = \frac{1}{n} \sum_1^n \mu_i \quad = \text{average of means}$$

Define

$$\bar{\sigma}_n^2 = \frac{1}{n} \sum_1^n \sigma_i^2 \quad = \text{average variance}$$

Suppose

$$\text{(i)} \quad \lim_{n \to \infty} \max_i \frac{\sigma_i}{n \bar{\sigma}_n^2} = 0$$

(No individual variance is too big

$$\text{(ii)} \quad \lim_{n \to \infty} \bar{\sigma}_n^2 = \bar{\sigma}^2 < \infty$$

(limit of avg. variance is finite

requires $E[X_i^4] < \infty$

Then

$$\sqrt{n} \left(\frac{\bar{X} - \bar{\mu}_n}{\bar{\sigma}_n} \right) \xrightarrow{d} N(0,1)$$

or

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

Remark: The key condition to check is that the asymptotic variance of $\sqrt{n}\,\bar{X}$,

$$\bar{\sigma}^2,$$

is finite.

Example : Consistency & Asymptotic Normality of β in classical linear regression model

Let $\underset{1\times1}{y_i} = \underset{1\times1}{x_i} \cdot \underset{1\times1}{\beta} + \epsilon_i \qquad i = 1, \dots, n$

x_i is fixed (or analysis is conditional on x_i) and x_i is independent of ϵ_i

$\epsilon_i \sim$ iid with $E\{\epsilon_i\} = 0$ and $\text{var}(\epsilon_i) = \sigma^2$

$\lim_{n \to \infty} \frac{1}{n} \sum_1^n x_i^2 = Q > 0$ and finite

This is a linear regression model with 1 RHS regressor.

OLS gives

$$\hat{\beta} = (x'x)^{-1} x' y = \left(\sum_1^n x_i^2 \right)^{-1} \sum_1^n x_i y_i$$

$$= \left(\sum_1^n x_i^2 \right)^{-1} \sum_1^n x_i \left(x_i \beta + \epsilon_i \right)$$

$$= \beta + \left(\sum_1^n x_i^2 \right)^{-1} \sum_1^n x_i \epsilon_i$$

$$\Rightarrow \quad \hat{\beta} - \beta = \left(\sum_1^n x_i^2 \right)^{-1} \sum_1^n x_i \epsilon_i \qquad (*)$$

Consistency of $\hat{\beta}$

Multiply the RHS of (*) by $\frac{n}{n}$ to give

$$\hat{\beta} - \beta = \left(\frac{1}{n} \Sigma_1^n x_i^2 \right)^{-1} \frac{1}{n} \Sigma_1^n x_i \epsilon_i$$

By assumption

$$\frac{1}{n} \Sigma_1^n x_i^2 \longrightarrow Q > 0 \quad \text{as } n \to \infty$$

and by the continuity of (matrix) inversion

$$\left(\frac{1}{n} \Sigma_1^n x_i^2 \right)^{-1} \longrightarrow Q^{-1} \quad \text{as } n \to \infty$$

Next

$$\frac{1}{n} \Sigma_1^n x_i \epsilon_i = \frac{1}{n} \Sigma_1^n w_i \quad , w_i = x_i \epsilon_i$$

and

$$E[w_i] = x_i E[\epsilon_i] = 0$$

$$\text{Var}(w_i) = x_i^2 \text{var}(\epsilon_i) = x_i^2 \sigma^2 < \infty$$

and is uniformly bounded.

Applying Markov's LLNs gives

$$\frac{1}{n} \sum_{i=1}^{n} w_i \xrightarrow{P} 0$$

Hence

$$\hat{\beta} - \beta \xrightarrow{P} Q^{-1} \cdot 0 = 0$$

and so

$$\hat{\beta} \xrightarrow{P} \beta$$

<u>Asymptotic Normality</u>

Multiply both sides of $(*)$ by $\sqrt{n}$ to give

$$\sqrt{n}\,(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_1^n x_i^2\right)^{-1} \frac{1}{\sqrt{n}} \sum_1^n x_i \cdot \epsilon_i$$

As shown above

$$\left(\frac{1}{n} \sum_1^n x_i^2\right)^{-1} \longrightarrow Q^{-1} \quad \text{as} \quad n \to \infty$$

Next

$$\frac{1}{\sqrt{n}} \sum_1^n x_i \epsilon_i = \frac{1}{\sqrt{n}} \sum_1^n w_i \quad , \quad w_i = x_i \epsilon_i$$

$$= \sqrt{n} \left(\frac{1}{n} \sum_1^n w_i\right)$$

$$= \sqrt{n}\ \bar{w}$$

To apply the Lindeberg - Feller CLT we need that the asymptotic variance of $\sqrt{n}\,\bar{w}$ be finite. Now

$$\operatorname{var}(\sqrt{n}\,\bar{w}) = \operatorname{var}\left(\frac{1}{\sqrt{n}} \Sigma_1^n x_i \epsilon_i\right)$$

$$= \frac{1}{n} \Sigma_1^n \operatorname{var}(x_i \epsilon_i)$$

$$= \frac{1}{n} \Sigma_1^n x_i^2 \sigma^2$$

$$= \frac{\sigma^2}{n} \Sigma_1^n x_i^2$$

and

$$\lim_{n \to \infty} \frac{\sigma^2}{n} \Sigma_1^n x_i^2 = \sigma^2 \cdot Q$$

= asymptotic variance of $\sqrt{n}\,\bar{w}$

$< \infty$ by ass.

Applying the Lindeberg - Feller CLT

$$\sqrt{n}\,\bar{w} \xrightarrow{d} N(0, \sigma^2 \cdot Q)$$

Hence

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \Sigma_1^n x_i^2\right)^{-1} \frac{1}{\sqrt{n}} \Sigma_1^n x_i \epsilon_i$$

$$\xrightarrow{d} Q^{-1} \cdot N(0, \sigma^2 \cdot Q) = N(0, \sigma^2 Q^{-1}$$

Alternatively,

$$\hat{\beta} \overset{A}{\sim} N\left(\beta, \sigma^2 \cdot \frac{1}{n} Q^{-1} \right)$$

and

$$\frac{\sigma^2}{n} \cdot Q^{-1} = \text{asymptotic variance of } \hat{\beta}.$$

<u>Remarks</u>

1. The asymptotic variance of $\hat{\beta}$ depends on σ^2 and Q and these values are not known in general. To get a practically useful asymptotic result we require a consistent estimate of the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta)$

2. It can be shown (see homework problem) that

$$\hat{\sigma}^2 = \frac{1}{n} \Sigma_1^n (y_i - \hat{\beta} x_i)^2 \xrightarrow{p} \sigma^2 .$$

Given the assumption that

$$\frac{1}{n} \Sigma_1^n x_i^2 \longrightarrow Q \quad \text{as } n \to \infty$$

a consistent estimate of $\sigma^2 Q^{-1}$ is given by

$$\hat{\sigma}^2 \cdot \left(\frac{1}{n} \Sigma_1^n x_i^2\right)^{-1} \xrightarrow{p} \sigma^2 Q^{-1}$$

Therefore, a consistent estimate of the asymptotic variance of $\hat{\beta}$ is given by

$$\frac{1}{n} \hat{\sigma}^2 \left(\frac{1}{n} \Sigma_1^n x_i^2\right)^{-1}$$

$$= \hat{\sigma}^2 \left(\Sigma_1^n x_i^2\right)^{-1}$$

$$= \hat{\sigma}^2 (x'x)^{-1} \quad \text{with} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

which is the usual OLS ~~est~~ variance estimate for $\hat{\beta}$.