1 Asymptotics

Asymptotic theory is concerned about the behavior of statistics when the sample size is arbitrarily large. It is a useful approximation technique to simplify extremely complicated finite-sample analysis.

1.1 Modes of Convergence

Convergence of a deterministic sequence means that for any $c > 0$, there exists an $N(c)$ such that for all $n > N(c)$, we have $|z_n - z| < c$. We say z is the limit of z_n, and denote as $z_n \to z$.

In contrast to the convergence of a deterministic sequence, we are interested in the convergence of random variables. There are various modes of convergence. We introduce three of them.

- Convergence in probability: $\lim_{n\to\infty} P(\omega : |Z_n(\omega) - z| < c) = 1$.
- Squared-mean convergence: $\lim_{n\to\infty} E\left[(z_n - z)^2\right] = 0$.

Example 1. $z_n(\omega) \in \{0,1\}$ is a binary variable. $z_n = \sqrt{n}$ with probability $1/n$, and $z_n = 0$ with probability $1 - 1/n$. Then $z_n \stackrel{p}{\to} 0$ but $z_n \stackrel{m.s.}{\nrightarrow} 0$.

Convergence in probability does not take care what happens on a set with small probability. Convergence in mean square takes care the average over the entire probability space. Even at a small probability the value of the random variables deviates too much, it may blow away the convergence in squared-mean. On the contrary, such deviation does not undermine convergence in probability.

- Convergence in distribution: $x_n \stackrel{d}{\to} x$ if $F(x_n) \to F(x)$ for each x on which $F(x)$ is continuous.

Convergence in distribution is the convergence of CDF, not the random variables themselves.

Example 2. Let $x \sim N(0,1)$. If $z_n = x + 1/n$, then $z_n \stackrel{p}{\to} x$ and of course $z_n \stackrel{d}{\to} x$. However, if $z_n = -x + 1/n$, or $z_n = y + 1/n$ where $y \sim N(0,1)$ is independent of x, then $z_n \stackrel{d}{\to} x$ but $z_n \stackrel{p}{\nrightarrow} x$.

1.2 Law of Large Numbers

The law of large numbers (LLN) is a collect of probability theorems about the convergence in probability of sample means. The basic form of the LLN is: the sample average of $(z_1, \ldots, z_n)$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n} z_i - E\left(\frac{1}{n}\sum_{i=1}^{n} z_i\right) \stackrel{p}{\to} 0.$$

Various versions of LLN work under different assumptions about the random variables.

The simplest version of LLN utilizes Chebyshev inequality, a simple fact from probability theory. For any random variable x with finite $E\left[x^2\right]$, we have

$$P(|x| > c) \leq E\left[x^2\right]/c^2$$

for any positive constant c.

Proof. $E\left[x^2\right] = \int_{|x|>c} x^2 dF_X + \int_{|x|\leq c} x^2 dF_X \geq \int_{|x|>c} x^2 dF_X \geq c \int_{|x|>c} dF_X = c^2 P\left(|x| > c\right).$ □

- Chebyshev LLN
- Kolmogorov LLN

Not revised yet. Don't read beyond this point.

1.3 Central Limit Theorem

The central limit theorem (CLT) is a collect of probability theorems about the convergence in distribution to a normally distributed random variable. The basic form of the CLT is: For a sample $(z_1, \ldots, z_n)$ of *zero-mean* random variables, the sample mean scaled up by $\sqrt{n}$ satisfies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i \xrightarrow{d} N\left(0, \sigma^2\right).$$

Various versions of CLT work under different assumptions about the random variables.

Let $\varphi_x(t) = E\left[\exp(ixt)\right]$ be the characteristic function. If $E[X] < \infty$ for a positive integer k, then

$$\varphi_x(t) = 1 + itE[X] + \frac{(it)^2}{2} E\left[X^2\right] + \ldots \frac{(it)^k}{k!} E\left[X^k\right] + o\left(t^k\right).$$

Therefore, if x_i's mean is zero and variance is σ^2, and its third moment is finite, then $\varphi_{x_i}(t) = 1 - \frac{\sigma^2}{2} t^2 + o\left(\frac{\sigma^2}{2} t^2\right)$. If the observations in the sample are independent, then

$$\varphi_{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i}(t) = \left(1 - \frac{\sigma^2}{2n} t^2 + o\left(\frac{\sigma^2}{2n} t^2\right)\right)^n \to \exp\left(-\frac{\sigma^2}{2} t^2\right),$$

where the limit is exactly the characteristic function of $N\left(0, \sigma^2\right)$.

- Lindeberg-Levy CLT: zero-mean, finite σ^2.
- Lindeberg-Feller CLT
- Lindeberg condition: Let $(x_i)_{i=1}^n$ be a sample of zero-mean independent observations. For any fixed $\varepsilon > 0$,

  $$\frac{1}{s_n^2} \sum_{i=1}^{n} \int_{|x_i| > \varepsilon s_n} x_i^2 dP x_i \to 0$$

  where $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$.
- Lyapunov CLT

1.4 Tools for Transformations

The original forms of LLN or CLT only deal with sample means. However, most of the econometric estimators of interest are functions of sample means. Therefore, we need tools to handle transformations.

- Small op: $x_n = o_p(r_n)$ if $x_n/r_n \xrightarrow{p} 0$.
- Big Op: $x_n = O_p(r_n)$ if for any $\varepsilon > 0$, there exists a $c > 0$ such that $P(x_n/r_n > c) < \varepsilon$.
- Continuous mapping theorem 1: If $x_n \xrightarrow{p} a$ and $f(\cdot)$ is continuous at a, then $f(x_n) \xrightarrow{p} f(a)$.
- Continuous mapping theorem 2: If $x_n \xrightarrow{d} x$ and $f(\cdot)$ is continuous almost surely on the support of x, then $f(x_n) \xrightarrow{d} f(x)$.
- Delta method

2 Apply Asymptotics to OLS

Convergence in probability: $\frac{1}{n}\sum_{i=1}^n x_i x_i' \xrightarrow{p} E[x_i x_i']$, and $\frac{1}{n}\sum_{i=1}^n x_i e_i \xrightarrow{p} 0$. By the continuous mapping theorem, $\widehat{\beta} - \beta \xrightarrow{p} 0$.

Asymptotic distribution: $n^{-1/2}\sum_{i=1}^n x_i e_i \xrightarrow{d} N\left(0, E\left[x_i x_i' e_i^2\right]\right)$.

2.1 Estimation of the Variance

To show the finiteness of the variance, $\Omega = E\left[x_i x_i' e_i^2\right]$. Let $z_i = x_i e_i$, so $\Omega = E[z_i z_i']$. Because of the Cachy-Schwarz inequality,

$$\|\Omega\|_\infty = \max_{k=1,\ldots,K} E\left[z_{ik}^2\right].$$

For each k, $E\left[z_{ik}^2\right] = E\left[z_{ik}^2 e_i^2\right] \leq \left(E\left[z_{ik}^4\right] E\left[e_i^4\right]\right)^{1/2}$.

For the estimation of variance, homoskedastic,

$$\frac{1}{n}\sum_{i=1}^n \left(e_i + x_i'\left(\widehat{\beta}-\beta\right)\right)^2 = \frac{1}{n}\sum_{i=1}^n e_i^2 + \left(\frac{1}{n}\sum_{i=1}^n e_i x_i\right)'\left(\widehat{\beta}-\beta\right) + \frac{1}{n}\sum_{i=1}^n e_i^2\left(\widehat{\beta}-\beta\right)' x_i x_i'\left(\widehat{\beta}-\beta\right)$$

The third term

$$\left(\widehat{\beta}-\beta\right)\left(\frac{1}{n}\sum_{i=1}^n e_i^2 x_i x_i'\right)\left(\widehat{\beta}-\beta\right) = o_p(1)\, O_p(1)\, o_p(1) = o_p(1).$$

The second term

$$\left(\frac{1}{n}\sum_{i=1}^n e_i x_i\right)'\left(\widehat{\beta}-\beta\right) = o_p(1)\, o_p(1) = o_p(1).$$

As $\frac{1}{n}\sum_{i=1}^n \widehat{e}_i^2 = \frac{1}{n}\sum_{i=1}^n \widehat{e}_i^2 + o_p(1)$ and $\frac{1}{n}\sum_{i=1}^n e_i^2 = \sigma_e^2 + o_p(1)$, we have $\frac{1}{n}\sum_{i=1}^n \widehat{e}_i^2 = \sigma_e^2 + o_p(1)$. In other words, $\frac{1}{n}\sum_{i=1}^n \widehat{e}_i^2 \xrightarrow{p} \sigma_e^2$.

2.2 Asymptotic Distribution of Functional of Parameters

A linear restriction that leads to the Wald test.

Hansen's textbook provides a few examples why we are interested in a function of parameters. For example, the optimal experience level of a worker. We derive the distribution by the delta method. It is particularly easy if the parameter of interest is of one dimensional.

Besides point estimator, we might be interested in the interval estimator C_n. What want to construct $\widehat{C}_n$ such that

$$\lim_{n\to\infty} P\left(\beta_0 \in \widehat{C}_n\right) = 1-\alpha.$$

If we construct $\widehat{C}_n = \left[\widehat{\beta} - 1.96\widehat{\sigma}_\beta, \widehat{\beta} + 1.96\widehat{\sigma}\right]$, then we have

$$P\left(\widehat{\beta} - 1.96\widehat{\sigma} \le \beta_0 \le \widehat{\beta} + 1.96\widehat{\sigma}\right) = P\left(1.96 \le \frac{\widehat{\beta}-\beta_0}{\widehat{\sigma}} \le 1.96\right) = F_n\left(1.96\right) - F_n\left(-1.96\right) \to \Phi\left(1.96\right) - \Phi\left(-1.96\right).$$

If multiple dimension, then one way to construct the interval is

$$P\left(\theta : \left(\widehat{\beta}-\beta\right) V^{-1} \left(\widehat{\beta}-\beta\right) \le q_{\chi^2}\right)$$

A Appendix

For general heteroskedasticity,

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \left(e_i + x_i'\left(\widehat{\beta}-\beta\right)\right)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' e_i^2 + \frac{1}{n}\sum_{i=1}^{n} x_i x_i e_i x_i' \left(\widehat{\beta}-\beta\right) + \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \left(\left(\widehat{\beta}-\beta\right)' x_i\right)^2$$

The third term is bounded by

$$\begin{aligned} & \operatorname{trace}\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \left(\left(\widehat{\beta}-\beta\right)' x_i\right)^2\right) \\ \le\ & K \max_k \frac{1}{n}\sum_{i=1}^{n} x_{ik}^2 \left[\left(\widehat{\beta}-\beta\right)' x_i\right]^2 \le K \left\|\widehat{\beta}-\beta\right\|_2^2 \max_k \frac{1}{n}\sum_{i=1}^{n} x_{ik}^2 \left\|x_i\right\|_2^2 \le K \left\|\widehat{\beta}-\beta\right\|_2^2 \frac{1}{n}\sum_{i=1}^{n} \left\|x_i\right\|_2^2 \left\|x_i\right\|_2^2 \\ =\ & K \left\|\widehat{\beta}-\beta\right\|_2^2 \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{K} x_{ik}^2\right)^2 \le K \left\|\widehat{\beta}-\beta\right\|_2^2 K \sum_{k=1}^{K} \frac{1}{n}\sum_{i=1}^{n} x_{ik}^4 = o_p\left(1\right) O_p\left(1\right) = o_p\left(1\right). \end{aligned}$$

where the third inequality follows by $(a_1 + \cdots + a_K)^2 \leq K\left(a_1^2 + \cdots + a_K^2\right)$. The second term is bounded by

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^{n} x_{ik}x_{ik'}e_i x_i'\left(\widehat{\beta}-\beta\right)\right| &\leq \max_k \left|\widehat{\beta}_k - \beta_k\right| K \max_{k,k',k''}\left|\frac{1}{n}\sum_{i=1}^{n} e_i x_{ik}x_{ik'}x_{ik''}\right| \\
&\leq \left\|\widehat{\beta}-\beta\right\|_2 \left(\frac{1}{n}\sum_{i=1}^{n} e_i^4\right)^{1/4} K \max_{k,k',k''}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{ik}x_{ik'}x_{ik''}\right)^{4/3}\right)^{3/4} \\
&\leq \left\|\widehat{\beta}-\beta\right\|_2 K \max_k \left(\frac{1}{n}\sum_{i=1}^{n} x_{ik}^4\right)^{3/4} = o_p(1)\, O_p(1)
\end{aligned}$$

where the second and the third inequality hold by the Holder's inequality.