This version: September 28, 2016

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 complicated finite-sample analysis.

1.1 Modes of Convergence

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

In contrast to the convergence of a deterministic sequence, we are interested in the convergence of random variables. Since a random variable has an associated probability measure, we must define clearly what "convergence" means. Several modes of convergence are often encountered.

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

Example 1. z_n is a binary random 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 count what happens on a subset with small probability in the sample space. Convergence in mean square takes care the average over the entire probability space. If a random variable can take a wild value, even with small probability, it may blow away the squared-mean convergence. On the contrary, such irregularity 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

(Weak) law of large numbers (LLN) is a collection of statements about convergence in probability of the sample average to its population counterpart. The basic form of the LLN is: the sample average of $(z_1, \ldots, z_n)$ satisfies

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

as $n \to \infty$. Various versions of LLN work under different assumptions about the distributions and dependence of the random variables.

- Chebyshev LLN: if $(z_1, \ldots, z_n)$ is a sample of i.i.d. observations, $E[z_1] = \mu$ exists, and $\sigma^2 = \operatorname{var}[x_1] < \infty$, then $\frac{1}{n}\sum_{i=1}^{n} z_i - \mu \stackrel{p}{\to} 0$.

Chebyshev LLN utilizes

- *Chebyshev inequality.* For any random variable x with finite $E\left[x^2\right]$, we have $P(|x| > \varepsilon) \le E\left[x^2\right]/\varepsilon^2$ for any $\varepsilon > 0$.

Chebyshev inequality is a special case of

- *Markov inequality*: $P(|x| > \varepsilon) \le E[|x|^r]/\varepsilon^r$ for $r \ge 1$ and any $\varepsilon > 0$.

It is easy to verify Markov inequality.

$$\begin{aligned} E[x^r] &= \int_{|x|>\varepsilon} x^r dF_X + \int_{|x|\le\varepsilon} x^r dF_X \\ &\ge \int_{|x|>\varepsilon} x^r dF_X \ge \varepsilon \int_{|x|>\varepsilon} dF_X = \varepsilon^r P(|x| > \varepsilon). \end{aligned}$$

Consider a partial sum $S_n = \sum_{i=1}^{n} x_i$, where $\mu_i = E[x_i]$ and $\sigma_i^2 = \text{var}[x_i]$. We apply the Chebyshev inequality to the sample mean $\overline{x} - \bar{\mu} = n^{-1}\left(S_n - \sum_{i=1}^{n} \mu_i\right)$.

$$
\begin{aligned}
P\left(|\bar{x} - \bar{\mu}| \geq \varepsilon\right) &= P\left(\left|S_n - \sum_{i=1}^{n} \mu_i\right| \geq n\varepsilon\right) \\
&\leq (n\varepsilon)^{-2} E\left[\sum_{i=1}^{n} (x_i - \mu_i)^2\right] \\
&= (n\varepsilon)^{-2} \text{var}\left(\sum_{i=1}^{n} x_i\right) \\
&= (n\varepsilon)^{-2}\left[\sum_{i=1}^{n} \text{var}(x_i) + \sum_{i=1}^{n} \sum_{j \neq i} \text{cov}(x_i, x_j)\right].
\end{aligned}
$$

From the above derivative, as long as the right-hand side shrinks to 0 as $n \to \infty$, convergence in probability holds. It holds under much more general conditions than just under the i.i.d. assumption. The random variables in the sample do not have to be identically distributed, and they also do not have to be independent either.

Another useful LLN is called Kolmogorov LLN. Since the derivation requires advanced knowledge of mathematics, we just state the result without proof.

- Kolmogorov LLN: if $(z_1, \ldots, z_n)$ is a sample of i.i.d. observations and $E[z_1] = \mu$ exists, then $\frac{1}{n}\sum_{i=1}^{n} z_i - \mu \xrightarrow{p} 0$.

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 $\varphi_x(t) = E[\exp(ixt)]$ be the characteristic function. If $E\left[|x|^k\right] < \infty$ 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, 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: iid, zero-mean, finite σ^2.
- Lindeberg-Feller CLT: inid, and Lindeberg condition: for any fixed $\varepsilon > 0$,
$$\frac{1}{s_n}\sum_{i=1}^n \int_{x_i^2 > \varepsilon s_n} x_i^2 dPx_i \to 0$$
where $s_n = \sum_{i=1}^n \sigma_i^2$.
- Lyapunov CLT: inid, finite $E\left[|x|^3\right]$.

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 $x_n \stackrel{p}{\to} a$ and $f(\cdot)$ is continuous at a, then $f(x_n) \stackrel{p}{\to} f(a)$.
- Continuous mapping theorem 2: If $x_n \stackrel{d}{\to} x$ and $f(\cdot)$ is continuous almost surely on the support of x, then $f(x_n) \stackrel{d}{\to} f(x)$.
- Delta method: if $\sqrt{n}\left(\widehat{\theta}-\theta_0\right) \stackrel{d}{\to} N(0,\Omega)$, and $f(\cdot)$ is continuously differentiable at θ_0, then

$$\sqrt{n}\left(f\left(\widehat{\theta}\right)-f(\theta_0)\right) \stackrel{d}{\to} N\left(0, \frac{\partial f}{\partial \theta'}(\theta)\,\Omega\left(\frac{\partial f}{\partial \theta}(\theta)\right)'\right).$$

2 Apply Asymptotics to OLS

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

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

$$\sqrt{n}\left(\widehat{\beta}-\beta\right) \stackrel{d}{\to} N\left(0, Q^{-1}\Sigma Q^{-1}\right).$$

2.1 Estimation of the Variance

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

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

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

For the estimation of variance, homoskedastic,

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

The second term

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

The third term

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

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

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

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

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

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

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