Day 15 Notes

Friday, August 27, 2021 8:59 AM

Asymptotic Statistics

Properties of statistics and estimators when the sample size is "large".

→ especially when small sample properties are unknown (almost always the case in social science).

- convergence definitions
- limiting distributions
- asymptotic distributions and variance
- consistency
- Central Limit Theorem
- Delta method

Jeffrey Wooldridge
Follow @jmwooldridge
9 May, 5 tweets, 2 min read

Bookmark | Save as PDF | My Authors

A year ago on Facebook, at the request of a former MSU student, I made this post. I used to say in class that econometrics is not so hard if you just master about 10 tools and apply them again and again. I decided I should put up or shut up.

#metricstotheface

I cheated by combining tools that are connected, so there are actually more than 10 ....

1. Law of Iterated Expectations, Law of Total Variance
2. Linearity of Expectations, Variance of a Sum
3. Jensen's Inequality, Chebyshev's Inequality
4. Linear Projection and Its Properties

5. Weak Law of Large Numbers, Central Limit Theorem
6. Slutksy's Theorem, Continuous Convergence Theorem, Asymptotic Equivalence Lemma
7. Big Op, Little op, and the algebra of them.

8. Delta Method
9. Frisch-Waugh Partialling Out
10. For PD matrices A and B, A - B is PSD if and only if B^(-1) - A^(-1) is PSD.

The list hasn't changed in a year.

* read this for any questions about asymptotics

→ $\boxed{\text{Greene, Appendix D}}$

---

t distributions are exact distributions of estimators when the sample has a Normal dist. (in most cases this isn't true)

Also, as $n \to \infty$ (>100), $t \xrightarrow{d} N(0,1)$

So for large n, just replace t with Normal.

Convergence

3 main types of convergence

1) convergence in distribution $X_n \xrightarrow{d} X$ (weak convergence)
2) " in probability $X_n \xrightarrow{p} X$ (plim)
3) almost sure convergence $X_n \xrightarrow{a.s.} X$ (strong convergence)

$3 \Rightarrow 2 \Rightarrow 1$

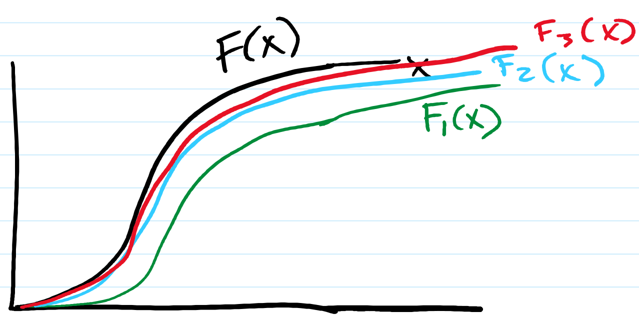
A second form of convergence is **convergence in distribution.** Let x_n be a sequence of random variables indexed by the sample size, and assume that x_n has cdf $F_n(x_n)$.

> **DEFINITION D.9 Convergence in Distribution**
> *x_n converges in distribution to a random variable x with cdf $F(x)$ if $\lim_{n\to\infty} |F_n(x_n) - F(x)| = 0$ at all continuity points of $F(x)$.*

$X_n \xrightarrow{d} X$

X_n has "same" distrib. as X in the limit

but maybe X_n is not close to X in any particular realization → not correlated (necessarily)



> **DEFINITION D.6 Convergence in Probability to a Random Variable**
> *The random variable x_n* ***converges in probability*** *to the random variable x if $\lim_{n\to\infty} \text{Prob}(|x_n - x| > \varepsilon) = 0$ for any positive ε.*

each realization of X_n must be close to the realizations of X

> **DEFINITION D.7 Almost Sure Convergence to a Random Variable**
> *The random variable x_n converges almost surely to the random variable x if and only if $\lim_{n\to\infty} \text{Prob}(|x_i - x| > \varepsilon \text{ for all } i \geq n) = 0$ for all $\varepsilon > 0$.*

We will make frequent use of a special case of convergence in probability, **convergence in mean square** or **convergence in quadratic mean.**

> **THEOREM D.1 Convergence in Quadratic Mean**
> *If x_n has mean μ_n and variance σ_n^2 such that the ordinary limits of μ_n and σ_n^2 are c and 0, respectively, then x_n converges in mean square to c, and*
> $$\text{plim } x_n = c.$$

conv. in quadratic mean $\Rightarrow$ conv. in prob.

$E(X_n) \to c$ (constant)

$Var(X_n) \to 0$

Limiting distributions

An estimator is said to have a limiting distribution of F if $X_n \xrightarrow{d} X$ [i.e. $F_n \to F$]

Asymptotic distributions

A large sample distribution which is used to approximate a small sample distribution.

Ex. Interested in studying y, but the limiting distribution of y is not useful or difficult to find, then we might take the limiting distribution of $X = g(y)$, which is easier to find, and then use $g^{-1}(X)$ to find an asymptotic dist of y.

g in this case is called a stabilizing transformation.

Ex) $\text{plim}\, \hat{\theta} = \theta \rightarrow$ distribution of $\hat{\theta}$ approaches a constant w/ zero variance

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} F(0, \sigma^2)$$

$$\downarrow$$

asymptotic variance

$$\text{Then } \hat{\theta} \sim F\left(\theta, \overbrace{\frac{\sigma^2}{n}}\right)$$

asymptotic distribution

Consistency

Sort of like "asymptotic unbiasedness".

Unbiased: $E(\hat{\theta}) = \theta$

Consistent: $\text{plim}\ \hat{\theta} = \theta$ ☆

Central Limit Theorem

(Greene App. D has more variants of CLT)

The Central Limit Theorem (Lindeberg and Lévy) for the Sample Mean. As in Section 5.6, we shall let Φ denote the d.f. of the standard normal distribution.

If the random variables $X_1, \ldots, X_n$ form a random sample of size n from a given distribution with mean μ and variance σ^2 $(0 < \sigma^2 < \infty)$, then for each fixed number x,

$$\lim_{n\to\infty} \Pr\left[\frac{n^{1/2}(\overline{X}_n - \mu)}{\sigma} \le x\right] = \Phi(x). \quad ■ \quad (5.7.1)$$

any distribution

$$\underbrace{\frac{n^{1/2}(\overline{X}_n - \mu)}{\sigma}}$$

a stabilizing transformation of $\overline{X}_n$ which a standard normal limiting distribution

The Central Limit Theorem (Liapounov) for the Sum of Independent Random Variables

We shall now state a central limit theorem that applies to a sequence of random variables $X_1, X_2, \ldots$ that are independent but not necessarily identically distributed. This theorem was first proved by A. Liapounov in 1901. We shall assume that $E(X_i) = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2$ for $i = 1, \ldots, n$. Also, we shall let

$$Y_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}}. \quad (5.7.6)$$

divide these all by N for clarity

Then $E(Y_n) = 0$ and $\text{Var}(Y_n) = 1$. The theorem that is stated next gives a sufficient condition for the distribution of this random variable Y_n to be approximately a standard normal distribution.

Theorem 5.7.2 Suppose that the random variables $X_1, X_2, \ldots$ are independent and that $E(|X_i - \mu_i|^3) < \infty$ for $i = 1, 2, \ldots$. Also, suppose that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n E(|X_i - \mu_i|^3)}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} = 0. \quad (5.7.7)$$

Finally, let the random variable Y_n be as defined in Eq. (5.7.6). Then, for each fixed number x,

$$\lim_{n\to\infty} \Pr(Y_n \le x) = \Phi(x). \quad ■ \quad (5.7.8)$$

regularity conditions
→ all R.V.'s need a finite third moment
→ always the case in survey research

For Liapounov's CLT, the R.V.'s

For Liapounov's CLT, the R.V.'s don't need to come from the same distribution.

* Look at Wikipedia article CLT
  → CLT converges quickly

• There are multivariate CLT's (see App. D of Greene)
  → these converge to multivariate normal distributions

Delta Method

→ used to approximate the variance of an asymptotic distribution

THEOREM D.12 Slutsky Theorem
For a continuous function $g(x_n)$ that is not a function of n,

$$\text{plim } g(x_n) = g(\text{plim } x_n). \quad \textbf{(D-6)}$$

D.2.7 THE DELTA METHOD

At several points in Appendix C, we used a linear Taylor series approximation to analyze the distribution and moments of a random variable. We are now able to justify this usage. We complete the development of Theorem D.12 (probability limit of a function of a random variable), Theorem D.16 (2) (limiting distribution of a function of a random variable), and the central limit theorems, with a useful result that is known as the **delta method.** For a single random variable (sample mean or otherwise), we have the following theorem.

→ ex. p. 1012

THEOREM D.21 Limiting Normal Distribution of a Function
If $\sqrt{n}(z_n - \mu) \xrightarrow{d} N[0, \sigma^2]$ and if $g(z_n)$ is a continuous function not involving n, then

$$\sqrt{n}[g(z_n) - g(\mu)] \xrightarrow{d} N[0, \{g'(\mu)\}^2\sigma^2]. \quad \textbf{(D-18)}$$

→ $g'(\mu)^2\sigma^2$ is the univariate version of $\nabla g(\mu)' \Sigma \nabla g(\mu)$

$$\downarrow$$

$$g'(\mu) \cdot \sigma^2 \cdot g'(\mu)$$
$$= g'(\mu)^2 \cdot \sigma^2$$

$$g(x) \approx g(\mu) + g'(\mu)(x-\mu)$$ → ~~first order~~ linear approximation

$$E(g(x)) \approx E(g(\mu) + g'(\mu)[E(x) - \mu])$$
$$= g(\mu) + g'(\mu)[\mu - \mu]$$ → 0
$$= g(\mu)$$

$$\text{Var}(g(x)) \approx E((g(x) - \underbrace{g(\mu)}_{E(g(x))})^2)$$

$$\mathrm{Var}(g(X)) \approx E\big((g(X) - \underbrace{g(\mu)}_{E(g(X))})^2\big)$$

$$\approx E\left(\left[g'(\mu)(X-\mu)\right]^2\right)$$

→ lets switch to multivariate so $g' \to \nabla g$, $X, \mu \in \mathbb{R}^n$

$$= E\left[\left(\nabla g(\mu)'(X-\mu)\right)^2\right]$$

$$= E\left[\left(\frac{\partial g(\mu)}{\partial x_1}\cdot(x_1-\mu_1) + \dots + \frac{\partial g(\mu)}{\partial x_n}(x_n-\mu_n)\right)^2\right]$$

$$= E\left[\sum_i \sum_j \frac{\partial g}{\partial x_i}\cdot\frac{\partial g}{\partial x_j}(x_i-\mu_i)(x_j-\mu_j)\right]$$

expanding the square

→ this is a quadratic form

$$= \sum_i \sum_j \frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j}\underbrace{E\left[(x_i-\mu_i)(x_j-\mu_j)\right]}_{\mathrm{cov}(x_i, x_j)}$$

$$= \nabla g(\mu)' \, \Sigma \, \nabla g(\mu)$$

$$\Sigma = \mathrm{COV}(X) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots \\ \vdots & \sigma_2^2 & & \\ \vdots & & \ddots & \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

In conclusion, if we approximate g with its linear approximation $\hat{g}$,

then $E(\hat{g}) = g(\mu)$

and $\mathrm{Var}(\hat{g}) = \nabla g(\mu)' \, \Sigma \, \nabla g(\mu)$

Notice that the mean and variance of the limiting distribution are the mean and variance of the linear Taylor series approximation:

$$g(z_n) \simeq g(\mu) + g'(\mu)(z_n - \mu).$$

The multivariate version of this theorem will be used at many points in the text.

THEOREM D.21A Limiting Normal Distribution of a Set of Functions

If $\mathbf{z}_n$ is a $K \times 1$ sequence of vector-valued random variables such that $\sqrt{n}(\mathbf{z}_n - \boldsymbol{\mu}) \xrightarrow{d} N[\mathbf{0}, \boldsymbol{\Sigma}]$ and if $\mathbf{c}(\mathbf{z}_n)$ is a set of J continuous functions of $\mathbf{z}_n$ not involving n, then

$$\sqrt{n}[\mathbf{c}(\mathbf{z}_n) - \mathbf{c}(\boldsymbol{\mu})] \xrightarrow{d} N[\mathbf{0}, \mathbf{C}(\boldsymbol{\mu})\boldsymbol{\Sigma}\mathbf{C}(\boldsymbol{\mu})'], \tag{D-19}$$

where $\mathbf{C}(\boldsymbol{\mu})$ is the $J \times K$ matrix $\partial \mathbf{c}(\boldsymbol{\mu})/\partial \boldsymbol{\mu}'$. The jth row of $\mathbf{C}(\boldsymbol{\mu})$ is the vector of partial derivatives of the jth function with respect to $\boldsymbol{\mu}'$.

$$\rightarrow \underbrace{C(\mu)\Sigma C(\mu)'}_{\nabla g(\mu)'\Sigma \nabla g(\mu)}$$

THEOREM D.22 Asymptotic Distribution of a Nonlinear Function

If $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N[0, \sigma^2]$ and if $g(\theta)$ is a continuous function not involving n, then $g(\hat{\theta}_n) \overset{a}{\sim} N[g(\theta), (1/n)\{g'(\theta)\}^2\sigma^2]$. If $\hat{\boldsymbol{\theta}}_n$ is a vector of parameter estimators such that $\hat{\boldsymbol{\theta}}_n \overset{a}{\sim} N[\boldsymbol{\theta}, (1/n)\mathbf{V}]$ and if $\mathbf{c}(\boldsymbol{\theta})$ is a set of J continuous functions not involving n, then $\mathbf{c}(\hat{\boldsymbol{\theta}}_n) \overset{a}{\sim} N[\mathbf{c}(\boldsymbol{\theta}), (1/n)\mathbf{C}(\boldsymbol{\theta})\mathbf{V}\mathbf{C}(\boldsymbol{\theta})']$, where $\mathbf{C}(\boldsymbol{\theta}) = \partial \mathbf{c}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$.

This expands on the Delta method to allow us to find asymptotic distributions of function of X, not just limiting distributions.

The algebra I did above was a linear approximation, so the $E(g(x))$ and $Var(g(x))$ were approximations. But these are the actual (not approx.) mean and variance of the limiting distributions.

Why? Because as $n \to \infty$, $X_n \to \mu$, so X_n is always super close to μ, so the linear approx. is a really good approximation.