Large Sample Properties & Simulation

Quantitative Microeconomics

R. Mora

Department of Economics
Universidad Carlos III de Madrid

Outline

The Notion of Large Sample Properties

- large sample properties look at how an estimate $\hat{\theta}$ of a parameter θ "behave" as the sample size gets larger and larger:
  1. how far is $\hat{\theta}$ from the true parameter θ as $n \to \infty$?
  2. how does the distribution of $\hat{\theta}$ look as $n \to \infty$?
- accordingly, we look at two notions of "large sample behavior":
  1. convergence in probability
  2. convergence in distribution

Convergence in Probability

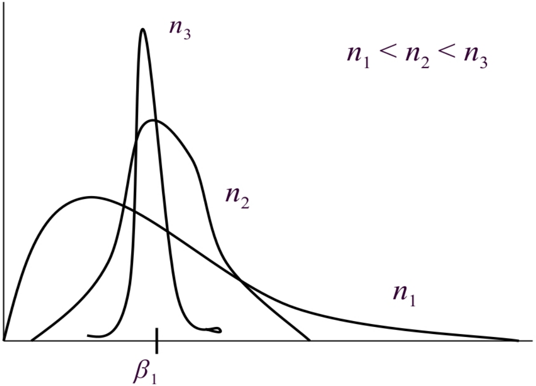
Definition

As the sample size grows, any arbitrarily small difference between $\hat{\theta}$ and θ becomes arbitrarily unlikely

- Technically: $\Pr\left(\left|\hat{\theta}-\theta\right|>\varepsilon\right)\to 0$ as $n\to\infty$

- θ is probability limit of $\hat{\theta}$
- $\hat{\theta}$ converges in probability to θ
- $plim(\hat{\theta})=\theta$

A Graphical Interpretation of Consistency



Source: Wooldrigde (2003)

The Law of Large Numbers: Some Basic Info

- in its simplest version, first proved by Bernoulli in 1713: it took him 20 years to get the actual proof
- it essentially states that the average converges in probability to the expected value
- the LLN is important because it "guarantees" stable long-term results for random events

A Law of Large Numbers

For any random variable y with an expected value μ define the average of a sample of size n as $\overline{y}_n$. Then

$$plim(\overline{y}_n) = \mu$$

Example 1

$$plim(\hat{cov}_n(y,x)) = cov(y,x)$$

Example 2

$$plim(\hat{var}_n(x)) = var(x)$$

plim Properties

Continuous Mapping Theorem

For every continuous function $g(\cdot)$ and random variable x:

$$plim(g(x)) = g(plim(x))$$

Example 1 $plim(x+y) = plim(x) + plim(y)$

Example 2 $plim\left(\frac{x}{y}\right) = \frac{plim(x)}{plim(y)}$ if $plim(y) \neq 0$

Example 1: The Fundamental Theorem of Statistics

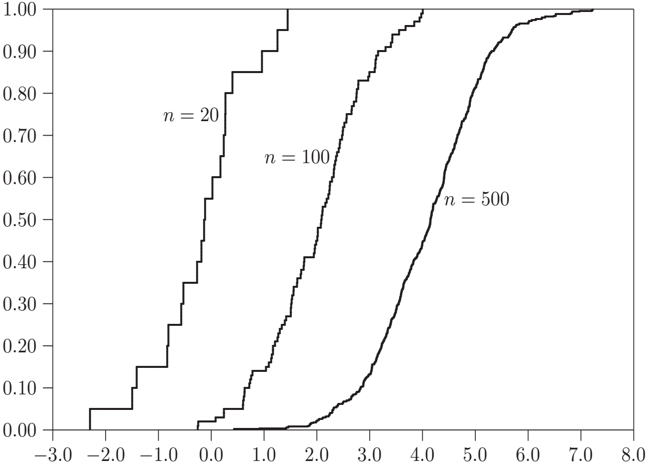
- Suppose that X is a random variable with CDF $F(X)$ and that we obtain a random sample of size n where typical element x_i is an independent realization of X
- The **empirical distribution** is the discrete distribution that puts a weight of $\frac{1}{n}$ at each of the x_i , $i = 1, ..., n$
- The **EDF** is the distribution function of the empirical distribution:

$$\hat{F}(x) \equiv \frac{1}{n} \sum_{i=1}^{n} I(x_i \leq x)$$

where $I(\cdot)$ is the indicator function.

The Fundamental Theorem of Statistics

$$plim \, \hat{F}(x) = F(x)$$



EDFs for three samples of sizes 20, 100, and 500 drawn from three normal distributions, each with variance 1 and with means 0, 2, and 4, respectively

Example 2: OLS under Classical Assumptions

Gauss-Markov Assumptions

- A1: Linearity: $y = \beta + \beta_1 x_1 + \ldots + \beta_k x_k + v$
- A2: Random Sampling
- A3: Conditional Mean Independence: $E[y\,|\mathbf{x}] = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$
- A4: Invertibility of Variance-covariance Matrix
- A5: Homoskedasticity: $Var[v\,|\mathbf{x}] = \sigma^2$

Normality

- A6: Normality: $y\,|\mathbf{x} \sim N(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \sigma^2)$

OLS Consistency

Theorem

Under Gauss-Markov A1-A4, OLS is consistent

Example: $wages = \beta_0 + \beta_1 educ + u$ with $cov(educ, u) = 0$

- $\hat{\beta}_1 = \beta_1 + \frac{\hat{cov}(educ_i, u_i)}{\hat{var}(educ_i)}$
- $plim\left(\hat{\beta}_1\right) = plim(\beta_1) + \frac{plim(\hat{cov}(educ_i, u_i))}{plim(\hat{var}(educ_i))} = \beta_1 + \frac{cov(educ, u)}{var(educ)}$
- Since $cov(educ, u) = 0 \Rightarrow plim\left(\hat{\beta}_1\right) = \beta_1$

An Example of Inconsistency

True Model: $wages = \beta_0 + \beta_1 educ + \beta_2 IQ + v$

- $cov(educ, v) = cov(IQ, v) = 0$
- $cov(educ, IQ) \neq 0$, $\beta_2 \neq 0$

- Estimated equation by OLS: $wages = \hat{\gamma}_0 + \hat{\gamma}_1 educ + \hat{u}_{educ}$

$$\hat{\gamma}_1 = \hat{\beta}_1 + \hat{\beta}_2 \frac{\hat{cov}(educ, IQ)}{\hat{var}(educ)} \Rightarrow plim(\hat{\gamma}_1) = \beta_1 + \beta_2 \frac{cov(educ, IQ)}{var(educ)}$$

- $plim(\hat{\gamma}_1) \neq \beta_1$ if
  - intelligence is relevant: $\beta_2 \neq 0$
  - education is correlated to intelligence: $cov(educ, IQ) \neq 0$

Asymptotic Normality

Definition

As the sample size grows, the distribution of $\hat{\beta}_j$ gets arbitrarily close to the normal distribution

- Technically: $\Pr(\hat{\beta}_j \leq z) \to \Phi(z)$ as $n \to \infty$

The Central Limit Theorem: Some History

- arguably, one of the most interesesting laws in maths, the proof is astonishing simple (but I must admit that I cannot intuitively understand the result)
- the first who thought about it was a French mathematician, de Moivre, in 1733
- Pierre Simon Laplace, another French, got the simplest version right in 1812
- the Russian Aleksander Liapunov was the guy who proved the general case in 1901

Rats and the Central Limit Theorem

- go to the sewage system in Madrid
- capture 100 rats, measure their tails,standardize the measures, and compute the average times square root of 100
- now capture more rats, say 500 rats, do as before using the square root of 500 instead of the square root of 100
  - if the distribution of the second average is closer to the standard normal, that's basically it
  - otherwise, instead of 500, try with a larger number of rats, say 10000
- the point is: if you keep increasing the sample size, you will CERTAINLY get closer to the normal

Stars and the Central Limit Theorem

- look at the brightness of 100 stars
- yeah, that's it! you will CERTAINLY get as close as you want to the normal after averaging by increasing sample size
- what is remarkable about the CLT is that the random distributions of star's brightness and rat's tails have nothing to do with each other
- the crucial issue is the averaging carried out over the measures after measurement

The Central Limit Theorem

For any random variable y with an expected value μ and variance σ^2 define the average of a sample of size n as $\overline{y}_n$. Then

$$n^{1/2}\frac{\overline{y}_n - \mu}{\sigma} \rightarrow N(0,1) \text{ as } n \rightarrow \infty$$

Under Gauss-Markov A.1 to A.5

$$n^{1/2}\frac{\hat{\beta}_j - \beta_j}{\sigma/a_j} \rightarrow N(0,1) \text{ as } n \rightarrow \infty \text{ where } a_j{}^2 = plim\left(\frac{1}{n}\sum_i r\hat{e}s_{ji}{}^2\right)$$

Moreover: $plim(\hat{\sigma}^2) = plim\left(\frac{SSR}{n-k-1}\right) = \sigma^2$

Asymptotic Normality for OLS Estimators

- As sample size n increases, the OLS estimators—conveniently scaled up—get as close as we want to a normal distribution

$$n^{1/2}\hat{\beta}_j \approx N\left(\beta_j, \sigma^2 a_j{}^2\right)$$

- the larger the sample, the more accurate the estimates
- important: the expression for the asymptotic variance depends on the homoskedasticity assumption

The t Test

- from the CLT (and a LLN), it can also be shown that

$$t = \frac{\hat{\beta}_j - \beta_j}{se\left(\hat{\beta}_j\right)} \to N(0,1) \text{ as } n \to \infty$$

- this result can be used to test whether a coefficient is significant

Small Sample Properties

- the idea is to undertand the behavior of an estimator for a given fixed sample size n

Under Gauss-Markov A.1 to A.5 AND Normality A.6

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \backsim t_{n-k-1}$$

- the problem is, normality is a very strong assumption

Monte Carlo Experiments: Definition and Motivation

Definition

- Monte Carlo experiments are a type of computational algorithms that rely on repeated random sampling to compute their results

- Monte Carlo methods are often used in simulating physical and mathematical systems
- Tend to be used when it is unfeasible or impossible to compute an exact result with a deterministic algorithm
- We are going to use them to generate observational data to study the behaviour of econometric techniques when samples are not infinite

Why Monte Carlo?

- the term was first used in the 1940s by physicists working on nuclear weapons in the US
- they wanted to solve a radiation problem but could not do it analytically, so they decided to model the experiment using chance
- they codenamed the project "Monte Carlo" in reference to the Monte Carlo Casino in Monaco where the uncle of one of them would borrow money to gamble
- it has been since used in many sciences
- for economists, Monte Carlo techniques are important for solving typical problems (integration, optimization) and also in games and in applied statistics

Monte Carlo in Applied Statistics

- there are two main uses of Monte Carlo in applied statistics
  - to compare and contrast competing statistics for small samples
  - to study how small sample behaviour translates into asymptotic results as samples become larger
- Monte Carlo techniques strike a nice balance
  - usually, they are more accurate than asymptotic results
  - not as time consuming to obtain as are exact tests

Monte Carlo vs. Simulation

- a simulation is a fictitious computer representation of reality
- a Monte Carlo study is a technique that can be used to solve a mathematical or statistical problem
- A Monte Carlo simulation uses repeated sampling of simulated data to determine the properties of some phenomenon

A Simple Simulation Exercise

1. using a computer, draw a value from a pseudo-random uniform variate from the interval [0,1]
2. If the value is less than or equal to 0.50 designate the outcome as heads
3. if the value is greater than 0.50 designate the outcome as tails

- this is a simulation of the tossing of a coin

A Simple Monte Carlo Study

the area of an irregular figure inscribed in a unit square

1. draw two values from a pseudo-random uniform variate from the interval [0,1]
2. If the point identified is within the figure, designate the outcome as "success", otherwise, as failure
3. repeat steps 1 and 2 many times
4. the proportion of successes provides the area of the figure

- this is using simple Monte Carlo techniques to compute a complex integral

A Simple Monte Carlo Experiment

1. draw one value from a pseudo-random uniform variate from the interval [0,1]
2. If the value is less than or equal to 0.50 designate the outcome as heads, otherwise tails
3. repeat steps 1 and 2 many times
4. the proportion of heads is the Monte Carlo simulation of the probability of heads

Using Monte Carlo in Applied Statistics

- we assume an econometric model and simulate it may times
- from each simulated population, we can extract a sample of size n and estimate the parameter of interest
- by looking at the descriptive statistics of the estimates across all simulated realities, we "estimate" the properties of the estimator when the sample size is fixed
- by increasing n and doing everything again, we "estimate" how the estimator behaves when the sample size increases

A Basic algorithm for a Monte Carlo experiment

A Monte Carlo experiment for a fixed sample of size N

1. assume values for the exogenous parts of the model or draw them from their respective distribution function
2. draw a (pseudo) random sample of size N for the error terms in the statistical model from their respective probability distribution functions
3. calculate the endogenous parts of the statistical model
4. calculate the value (e.g. the estimate) you are interested in
5. replicate step 1 to 4 R times
6. examine the empirical distribution of the R values

A simple Monte Carlo experiment

$log(wages) = 10 + 0.05 * D + u, \ u \sim N(0,1), \ D = 1$ with prob. 0.3

1. draw N realizations of D
2. draw N realizations of u
3. compute $log(wages)$
4. OLS $log(wages)$ on D and store $\hat{\beta}_1^r$
5. replicate step 1 to 4 R times
6. examine the empirical distribution of $\hat{\beta}_1^r$

Summary

- Large sample properties tell us how an estimator behaves as the sample size becomes arbitrarily large.
- Exact small sample properties are hard to get and sometimes they require strong assumptions.
- Monte Carlo simulations are useful in applied statistics.
- We can study small sample properties of estimators.
- We can also study how large sample properties are achieved in practice.