ECO375 Tutorial 6
Asymptotics

Matt Tudball

University of Toronto Mississauga

November 1, 2017

Review: Types of Convergence of Random Variables

- **Convergence in Probability:** Consider a sequence of n real-valued random variables $\{X_1, X_2, ..., X_n\}$ and some constant θ. Then the sequence converges in probability to θ if, for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|X_n - \theta| > \epsilon) = 0$.
- **Convergence in Distribution:** Consider a sequence of n real-valued random variables $\{X_1, X_2, ..., X_n\}$, each X_i with a distribution F_i. Consider also some random variable X with distribution F. Then the sequence converges in distribution to X if $\lim_{n\to\infty} F_n(x) = F(x)$ for every x in $\mathbb{R}$ such that F is continuous.

Review: LLN and CLT

- **(Weak) Law of Large Numbers:** Consider n random variables $X_1, X_2, ..., X_n$ drawn independently from the same distribution with expected value $\mathbb{E}(X_i) = \mu$. Then the Law of Large Numbers says that the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converges in probability to μ as $n \to \infty$.
- **(Lindeberg–Lévy) Central Limit Theorem:** Consider n random variables $X_1, X_2, ..., X_n$ drawn independently from the same distribution with expected value $\mathbb{E}(X_i) = \mu$ and variance $\text{Var}(X_i) = \sigma^2 < \infty$. Then the Central Limit Theorem says that as n converges to infinity $\sqrt{n}(\bar{X}_n - \mu)$ converges in distribution to $\mathcal{N}(0, \sigma^2)$.

Review: What are asymptotics?

- The basic idea behind asymptotic or "large sample" theory is that estimators will have certain desirable properties as the sample size n increases to infinity.
- What are some of these properties?
- Under MLR.1 to MLR.5,
  - **Consistency:** $\hat{\beta}_j$ converges to β_j as n goes to infinity.
    This intuition here is that the variance of $\hat{\beta}_j$ is going to 0 and the value of $\hat{\beta}_j$ is going to the true parameter β_j.
  - **Known Sampling Distribution:** $\sqrt{n}(\hat{\beta}_j - \beta_j)$ follows a Normal distribution as n goes to infinity (by the Central Limit Theorem).
  - **Asymptotic Efficiency:** $\hat{\beta}_j$ is asymptotically efficient, meaning that $\hat{\beta}_j$ has the lowest asymptotic variance within some class of estimators.

Consistency vs. Unbiasedness

- How does consistency relate to unbiasedness?
- Perhaps surprisingly, neither one implies the other. You can have estimators that are unbiased and inconsistent or biased and consistent.
- Consider a sample of n independently and identically distributed random variables $X_1, X_2, ..., X_n$ with mean μ and variance $\sigma^2 < \infty$. Suppose we want to estimate μ. We know that an unbiased and consistent estimator is $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- Example 1 (Unbiased and inconsistent): Consider the estimator $\hat{\mu}' = X_1$. Since the data is IID, $\mathbb{E}(\hat{\mu}') = \mathbb{E}(X_1) = \mu$, so this estimator is unbiased. But note that the variance is $\text{Var}(\hat{\mu}') = \text{Var}(X_1) = \sigma^2$ which does not go to 0 as n goes to ∞.
- Example 2 (Biased and consistent): Consider the estimator $\hat{\mu}'' = \hat{\mu} + \frac{1}{n}$. Since $\hat{\mu}$ is consistent and $\frac{1}{n} \to 0$, $\hat{\mu}''$ is consistent. But note that $\mathbb{E}(\hat{\mu}'') = \mathbb{E}(\hat{\mu} + \frac{1}{n}) = \mu + \frac{1}{n}$ which is not equal to μ.

Review: Relaxing MLR.6

- Notice how, in the previous slide, we did not have to assume MLR.6 for $\sqrt{n}(\hat{\beta}_j - \beta_j)$ to follow a Normal distribution.
- Recall the MLR.6 assumption:

MLR.6 The error u_i is independent of the explanatory variables $x_{i1}, x_{i2}, ..., x_{ik}$ and is normally distributed with mean 0 and variance σ^2
$u_i \sim \mathcal{N}(0, \sigma^2)$

- By appealing to the Central Limit Theorem instead of MLR.6, we can develop confidence intervals and test statistics for $\hat{\beta}_j$ based on a "good approximation" to the sampling distribution of u no matter what form that distribution actually takes.

Review: Test Statistics

- By the Central Limit Theorem,
  1. $\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}\hat{u}_i^2$ is a consistent estimator of $\sigma^2 = \text{Var}(u)$.
  2. For each j, $(\hat{\beta}_j - \beta_j)/\text{se}(\hat{\beta}_j) \xrightarrow{a} \mathcal{N}(0,1)$
     where $\text{se}(\hat{\beta}_j)$ is the usual OLS standard error.
- In practice, however, when doing hypothesis testing over a single parameter, we will typically still draw critical values from the t_{n-k-1} distribution.
- This is because t_{n-k-1} approaches $\mathcal{N}(0,1)$ as n goes to infinity. For u_i with distributions that are "close" to normal, t_{n-k-1} will serve as a better finite sample approximation to the t-statistic than the normal distribution.

Possible Problem: Finite Sample Approximation

- But what if u_i is not even close to normally distributed? What problems does this cause?
- We know that our estimators will still be consistent and efficient and $\sqrt{n}(\hat{\beta}_j - \beta_j)$ will still be normally distributed for $n \to \infty$.
- The problem occurs in how well the t-distribution approximates the distribution of the t-statistic in finite sample.
- The results of our hypothesis tests could potentially be very misleading.

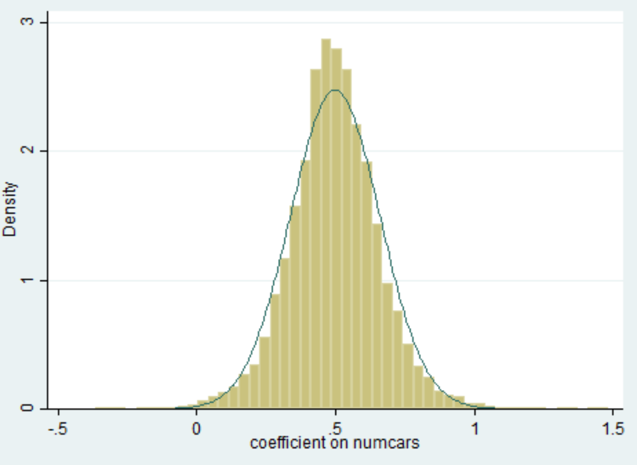
Possible Problem: Finite Sample Approximation

- Let's consider a simulation. Suppose we spend a month observing both the number of cars passing through a given intersection each day, called *numcars*, and the number of crashes occurring each day, called *crash*. We consider the model

$$crash = \beta_0 + \beta_1 numcars + u$$

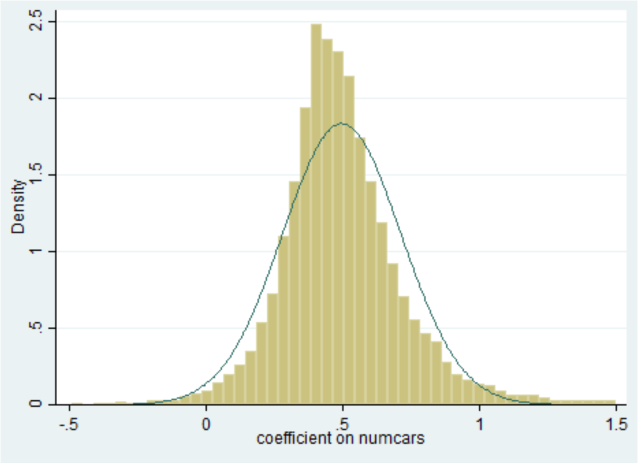
- I initially consider a data generating process where $numcars \sim \chi^2(1)$ and $u \sim \mathcal{N}(0, 1)$ which satisfies MLR.6. I assume true values $\beta_1 = 0.5$ and $\beta_0 = 0$.
- I then draw 10,000 samples from this data each of size $n = 30$ and record the estimates $\hat{\beta}_1$. We need these to be t-distributed in order to conduct accurate inference in finite sample.

Possible Problem: Finite Sample Approximation



Possible Problem: Finite Sample Approximation

- Next let's see what happens when we assume that $u \sim \chi^2(1)$ which is nowhere near normal. Keeping everything else the same, I run the simulation again.



In-Class Exercise 1

The purpose of this exercise is to determine when MLR.6 appears to be reasonable, in which case the t-distribution approximation in the previous slide will be sufficiently good in finite sample.
Download the dataset WAGE1.dta from my website (matthewtudball.com).

1. Estimate the equation

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

   Save the residuals and plot a histogram.
2. Repeat part 1 but with $\log(wage)$ as the dependent variable.
3. Would you say that assumption MLR.6 is closer to being satisfied for the level-level model or the log-level model?

In-Class Exercise 2

Using the same data as In-Class Exercise 1, suppose we are now interested in a model in which *educ* is the dependent variable.

1. How many different values are taken on by *educ* in the sample? Does *educ* have a continuous distribution?
2. Plot a histogram of *educ* with a normal distribution overlay. Does the distribution of *educ* appear anything close to normal?
3. Which of the MLR.1 - MLR.6 assumptions seems clearly violated in a model with *educ* as a dependent variable?

In-Class Exercise 3

The purpose of this exercise is to use real data to show how the standard errors of the OLS estimates decrease as n increases. Download the dataset GPA2.dta from my website (matthewtudball.com).

1. Using all 4,137 observations, estimate the equation

$$colgpa = \beta_0 + \beta_1 hsperc + \beta_2 sat + u$$

2. Re-estimate the equation in part 1 using the first 2,070 observations. You can use the command `keep in 1/2070`.
3. Find the ratio of the standard errors on *hsperc* from parts 1 and 2. What do you notice about this ratio?
   Hint: The formula for the standard error of $\hat{\beta}_j$ is $\frac{1}{\sqrt{n}} \frac{\sigma_u}{\sigma_{x_j}\sqrt{1-R_j^2}}$
   where R_j^2 is the population R-squared from regressing x_j on the other explanatory variables.

In-Class Exercise 4 (**Tough**)

Load the dataset SIMULATE.dta from my website (matthewtudball.com). Several statistics are commonly used to detect non-normality in underlying population distributions. Here we will study one that measures the amount of skewness in a distribution. Recall that any normally distributed random variable is symmetric about its mean: therefore, if we standardise a symmetrically distributed random variable, say $z = (\beta_j - \hat{\beta}_j)/\text{se}(\hat{\beta}_j)$, then z has mean 0, variance 1 and $\mathbb{E}(z^3) = 0$ (indicating no skew). We can use z to test for no skewness by calculating the statistic $\bar{z} = \frac{1}{n-1}\sum_{i=1}^{n} z_i^3$ (which is the sample analogue of $\mathbb{E}(z^3)$). The closer $\bar{z}$ is to 0, the less skewed is the distribution of z.

1. Calculate the skewness measure for *coef*1 which uses $u \sim \mathcal{N}(0, 1)$.
2. Calculate the skewness measure for *coef*2 which uses $u \sim \chi^2(1)$.
3. Using only the skewness measure, which of the two coefficients would you consider the most likely to have been calculated from a normally distributed error?