Probability and Statistics Review
Econometrics discussion section 1

John Green

Spring 2025

Housekeeping

- Introductions: name, major, thoughts about econometrics
- For now, office hour in Brody Cafe, Tuesdays, 3-4PM
    - May change to Wyman Park 601A
    - Timing poll: bit.ly/metricsOH
- What will we do in sections?
    - Cover important topics
    - Practice problems
    - Stata practice
    - Slides and resources in GitHub repo: github.com/JohnRGreen/EconometricsSpring25
- Sign in for attendance each week

Recap

What have we covered so far?

- Review of probability and statistics (Stock & Watson chapters 2 and 3)
- Random variables and their first 2 moments
- Marginal, joint and conditional distributions
- Independence
- Covariance and correlation
- Law of Iterated Expectations

Recap

Still to come (probably):

- Normal, Fisher (F), χ^2 distributions
- LLN and CLT
- Estimators and their properties: consistency, unbiasedness, normal approximation
- One variable t-test
- Two variable t-test

Random variables

- A random variable is a *function* from a space of possible outcomes to (usually) some subset of real numbers.
- May be discrete or continuous
- Roll two dice and add up the numbers (discrete)
  - Sample space is all the possible rolls of the two dice (how many?)
  - Outcome space is all the possible sums of those rolls (how many?)
  - The sum is a random variable
- Height of a person (continuous, but bounded)
  - Sample space is all people; might think about characterizing them by some covariates such as weight, age, etc.
  - Outcome space is all possible heights (what would this be?)
- Thinking about probability is different for discrete and continuous cases

		FIRST DICE					
		1	**2**	**3**	**4**	**5**	**6**
SECOND DICE	**1**	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Mean

- The first moment of a random variable is its mean, aka average or expected value
- What is the mean value of the result from rolling one die?
- What is the mean value of the sum of rolling both dice?

Mean

- The first moment of a random variable is its mean, aka average, aka expected value
- What is the mean value of the result from rolling one die?
  - $E[X] = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = 3.5$
- What is the mean value of the sum of rolling both dice?
  - $E[X] =$ $\frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \frac{4}{36} \times 5 + \frac{5}{36} \times 6 + \frac{6}{36} \times 7 + \frac{5}{36} \times 8 + \frac{4}{36} \times 9 + \frac{3}{36} \times 10 + \frac{2}{36} \times 11 + \frac{1}{36} \times 12 = 7$
  - (Could also write out using $\frac{1}{36}$ weight everywhere but easier to consolidate terms)

Variance

- The second moment of a random variable is its variance
  - Think of this as the mean distance from the mean: it measures the spread of our data
  - $Var(X) = E\left[(X - E[X])^2\right]$
  - Why do we need to square the spread?
- What is the variance of rolling one die?
- What is the variance of the sum of rolling both dice?

Variance

- The second moment of a random variable is its variance
  - Think of this as the mean distance from the mean: it measures the spread of our data
  - $Var(X) = E\left[(X - E[X])^2\right]$
  - Why do we need to square the spread?
- What is the variance of rolling one die?
  - $Var(X) = \frac{1}{6} \times (1 - 3.5)^2 + \frac{1}{6} \times (2 - 3.5)^2 + \cdots + \frac{1}{6} \times (6 - 3.5)^2 = \frac{35}{12} \approx 2.92$
- Similar idea for the sum of 2 dice.
- Often we will work with the *standard deviation* which is the square of the variance since the units are more interpretable.

Higher moments

- Third moment is *skewness* and tells us how symmetric our distribution is
  - What is the skewness of our example?
- Fourth moment is the *kurtosis* which measures the mass of the tails
  - Gives us an idea of the likelihood of large values

Covariance

- The covariance of two random variables tells us the strength of their *linear* (careful!) relationship
- If two random variables are independent, covariance is 0 (though the reverse is not true)
  - What is the covariance in our example?
- Often we will look at the *correlation* instead of the covariance, $\frac{cov(X,Y)}{(var(X)var(Y))^{1/2}}$
  - Unlike covariance, correlation is scaled between -1 and 1 and so is easily interpretable

Rules on joint distributions

- If two variables are independent, $P(X = x, Y = y) = P(X = x) * P(Y = y)$
- Conditional probability: $P(X = x|Y = y) = \frac{P(X=x,Y=y)}{P(Y=y)}$
    - What happens if the events are independent?
- Law of iterated expectations says the mean of Y can be written as a weighted average of the mean of $Y|X$: $E[Y] = E[E[Y|X = x]]$
    - $E[Y] = \sum_x E[Y|X = x]P(X = x)$ in discrete case
    - $E[Y] = \int_x E[Y|X = x]f_X(x)dx$ for continuous case
- Other topics in textbook: Bayes' law, law of total probability, etc.

Second example

- New example:
  - Roll the first dice
  - If first roll ≥ 4 then we roll the second die and observe its value
  - If first roll ≤ 3 then the second value is simply set to 1
- What is the joint probability distribution?
- What is the covariance of the 2 rolls?
- What is the correlation?
- What is the marginal distribution of the two r.v.? What is the first moment of each?

		FIRST DICE						
		1	2	3	4	5	6	Marginal:
SECOND DICE	1	0.17	0.17	0.17	0.03	0.03	0.03	0.58
	2	0	0	0	0.03	0.03	0.03	0.08
	3	0	0	0	0.03	0.03	0.03	0.08
	4	0	0	0	0.03	0.03	0.03	0.08
	5	0	0	0	0.03	0.03	0.03	0.08
	6	0	0	0	0.03	0.03	0.03	0.08
Marginal:		0.17	0.17	0.17	0.17	0.17	0.17	

Second example

- New example:
  - Roll the first dice
  - If first roll ≥ 4 then we roll the second die and observe its value
  - If first roll ≤ 3 then the second value is simply set to 1
- What is the marginal distribution of the two r.v.? What are the first two moments of each? $E[X_1] = 3.5$, $var(X_1) = 2.92$, $E[X_2] = 2.25$, $var(X_1) = 3.02$
- What is the covariance of the 2 rolls? $cov(X_1, X_2) = 1.875$
- What is the correlation? $corr(X_1, X_2) = \rho_{X_1, X_2} = 0.63$

Normal distribution

- Two parameters: mean μ and variance σ^2
  - μ tells us about the location of the distribution (where is it centered?)
  - σ tells us about the shape of the distribution (what is the spread? what do the tails look like?)
- We can *standardize* a normal random variable X:
  - $Z = \frac{X-\mu}{\sigma}$ so $Z \sim N(0,1)$
  - Careful to divide by standard deviation and not by variance!
- This will be helpful in hypothesis testing because we can consider how unlikely a given value z is to be drawn from a N(0,1) distribution

Estimation

- So far we have been in the world of probability theory, dealing with random variables and their distributions
- *Statistics* is all about using actual data to make inferences about the world, and so we will often try to find *estimators* for probabilistic objects
- For example, we will often be interested in estimating the mean (average, first moment) of a distribution

Estimation

- Example: wait time for a coffee at Brody Cafe
- A natural estimator for the population mean is the sample mean:
  - Take a survey of people as they walk out of Brody and ask how long they waited
  - Average the responses
- Sample mean $\bar{Y}$ is a random variable since we taking a random sample and thus $\bar{Y}$ has a sampling distribution
  - What can we say about it?

Unbiased

- First, $\bar{Y}$ is an **unbiased** estimator of μ_Y:

$$E[\bar{Y}] = \mu_Y$$

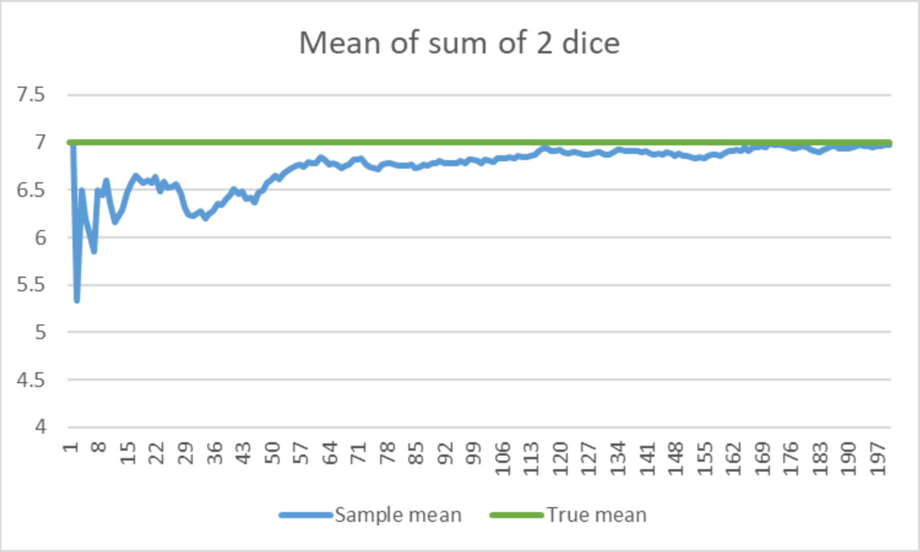
- In expectation, we get the thing we want!

Law of Large Numbers

- Second, $\bar{Y}$ is a **consistent** estimator of μ_Y
- Mathematically, this means that for any number ϵ we choose, we know that

$$\text{lim}_{n \to \text{inf}} P(|\bar{Y} - \mu_Y| > \epsilon) = 0$$

- In plain English: the larger our sample size, the better our guess for μ_Y
- As our sample gets infinitely large, we get infinitesimally close
- This follows from the **Law of Large Numbers**


Mean of sum of 2 dice
7.5
7
6.5
6
5.5
5
4.5
4
1 8 15 22 29 36 43 50 57 64 71 78 85 92 99 106 113 120 127 134 141 148 155 162 169 176 183 190 197
Sample mean
True mean

Efficiency

- Lastly, we can think about the *efficiency* of the estimator: what is its variance?
- In this case, $var(\bar{Y}) = \frac{\sigma_Y^2}{n}$
- In other words, the variance decreases as the sample gets larger, at a rate of $\frac{1}{n}$
- We can use this to compare the efficiency of two estimators, for example if we had another estimator with variance $\frac{\sigma_Y^2}{\sqrt{n}}$ we can say that the sample mean is more efficient

Now let's look at the most remarkable result about the sample mean.

Central Limit Theorem

- The central limit theorem tells us that $\bar{Y}$ is *asymptotically normal*:

$$\sqrt{n}(\bar{Y} - \mu_Y) \xrightarrow{d} N(0, \sigma_Y^2)$$

- In other words, as $n \to \infty$, $\bar{Y}$ follows a normal distribution, $\bar{Y} \sim N(\mu_Y, \frac{\sigma_Y^2}{n})$
- This means we can use the normal distribution to make inferences about sample mean
- To make it easy, we can standardize the sample mean: $Z = \frac{\bar{Y} - \mu_Y}{\sigma_Y / \sqrt{n}} \sim N(0, 1)$
- We will use the sample variance as an estimator for the population variance, just like we do for mean (but we will need to correct a small bias)

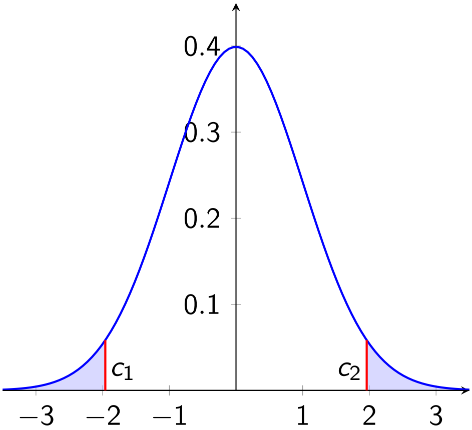
Hypothesis testing

- We can use the normal approximation to perform a *hypothesis test*
  - Set a null hypothesis, eg $\mu_Y = \mu_{Y,0}$
  - Usually we make the null the thing we "want" to disprove; if we reject the null, then we have evidence for our alternative hypothesis, eg $\mu_Y \neq \mu_{Y,0}$
- Intuition: assuming the true mean is some value $\mu_{Y,0}$, how likely is it that we would observe the sample mean $\bar{Y}$?
  - If it is "very" unlikely, we will reject the null
  - If it is "reasonably likely" then we fail to reject the null
- Alternative perspective: what would I have to see to convince myself *that my hypothesis is wrong*?
- Usually, we standardize everything and create a $t - statistic$
  - We test t under the null that $t \sim N(0, 1)$ rather than work directly with the sample mean

Hypothesis testing (cont.)

- Set a given significance α and find the *critical value(s)*
  - ie the value(s) such that the sum in the tails (or tail) of the standard normal distribution is α
- A common value is $\alpha = 0.05$, with critical values $c_1 = -1.96$ and $c_2 = 1.96$)
- In this test, the probability of getting a t less than -1.96 or greater than 1.96 is 0.05

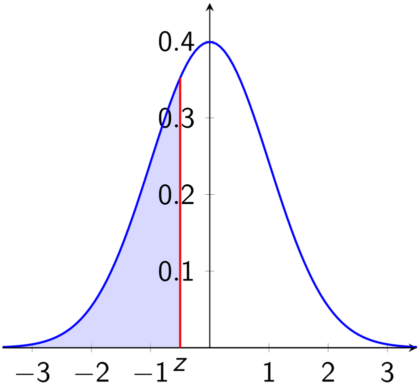
2-tailed test with $\alpha = 0.05$



Hypothesis testing (cont.)

- Set a given significance α and find the *critical value(s)*
  - ie the value(s) such that the sum in the tails (or tail) of the standard normal distribution is α
- A common value is $\alpha = 0.05$, with critical values $c_1 = -1.96$ and $c_2 = 1.96$)
- In this test, the probability of getting a t less than -1.96 or greater than 1.96 is 0.05
- Our decision procedure is that we reject (fail to accept) the null hypothesis if $|t| > 1.96$
  - Intuition: under the null, it is very unlikely we observe a t in the tails. Thus we reject the null if we do so.
- Alternatively, if $|t| < 1.96$ then we fail to reject H_0

Standard Normal Distribution



Calculating p-value

- Alternatively we might focus on the *p-value*: probability of a test statistic at least as unlikely as the one you observe (under the null)

$$P(x > |t|)$$

- So, the **p-value** tells us how unlikely the null hypothesis is:
    - $P_{H_0}(|\frac{\bar{Y}-\mu_{Y,0}}{\sigma_Y/\sqrt{n}}| \geq |\frac{\bar{Y}_{data}-\mu_{Y,0}}{\sigma_Y/\sqrt{n}}|)$
    - RHS is just a number! LHS is RV with known distribution under null
    - CLT tells us this **t-statistic** is N(0,1) so probability in tails is easy to look up
- If observed t-statistic is very large or very small, then p-value is very small and we reject the null
    - If the null had been true, it is very unlikely that we would have observed the t-statistic

Confidence interval

- Comparing p-value to α is equivalent to comparing t to c
- In either case, more rigorous to pre-set our rejection region
- We can also construct a *confidence interval* for μ_X based on our data: the set of all possible μ_X values which would not be rejected by a two-sided t-test with size α

$$\bar{X} \pm c(\alpha)\frac{\sigma_X}{\sqrt{n}} = \left(-c(\alpha)\frac{\sigma_X}{\sqrt{n}}, c(\alpha)\frac{\sigma_X}{\sqrt{n}}\right)$$

- Under the null, 95% of the time we draw a value in this interval
- What would this look like for a one-sided test?

Some terminology

- **Type-1 error**: reject a true null hypothesis
    - **Size** is probability of a type-1 error
- **Type-2 error**: fail to reject a false hypothesis
    - **Power** is probability of a type-2 error
- Which of these two mistakes is worse?
- What is the relationship between the size and the power?

Significance Level (α) and Power (1-β)

	H_0 is True	H_0 is False
Test Rejects H_0	α	1-β
Test Doesn't Reject H_0	1-α	β

Hypothesis about difference in means

- What if we want to perform a hypothesis on the difference in means between two random variables? For example:

$$H_0 : \mu_1 = \mu_2$$

- Simple: combine them to one!
    - $\mu^* = \mu_1 = \mu_2$
    - This is just a difference of two normal random variables so we can use the normal approximation, t-statistic is:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- What if we want to test hypothesis that the difference is 2?