PS598 - Mathematics for Political Scientists - Section Notes (Extremely Abridged!)

Jason S. Davis

Fall 2014

Contents

1 Welcome/Introduction

- Jason Davis - E-mail: jasonsd@umich.edu - Office: Haven Hall 7730 - Telephone: (write in class)
- Most of you know me from math camp.
- Will have office hours immediately after sections generally, though not today!
- Besides textbooks, one other useful resource for a lot of this math is Martin Osborne's (a game theorist at University of Toronto) math tutorial. It is references by a lot of economists and political scientists, and his game theory textbooks are also super popular:
  http://www.economics.utoronto.ca/osborne/MathTutorial/

2 LaTeX

- Good idea to get LaTeX.
- Can post some notes on how to get used to it. But a bit of an overview...
- Lots of different editors. MiKTeX is the most common implementation.
- WinEDT used by some. TeXnicCenter. Various others.
- Not WYSIWYG. Over time, easy to get used to the typsetting.
- Very good for math. Good for other stuff too.
- Useful resources for learning LaTeXand R can be found (courtesy of Rochester's political science department) at the following link:
  http://www.rochester.edu/college/psc/thestarlab/resources

3 L'Hopital's rule

- We learned a number of diferent ways of computing derivatives.
- Can also use derivatives to help evaluate limits. L'hopital's rule allows this
- Say you have two functions that tend to zero, or tend to positive infinity, as they approach either a constant or infinity.
- Examples: $lim_{x\to c}\ f(x)=\infty$ and $lim_{x\to c}\ g(x)=\infty$
- More examples: $lim_{x\to\infty}\ f(x)=0$ and $lim_{x\to\infty}\ g(x)=0$
- L'hopital's rule tells us that $lim_{x\to c}\ \frac{f(x)}{g(x)}=\lim_{x\to c}\ \frac{f'(x)}{g'(x)}$
- So long as $g'(x)\neq 0$ for all $x\neq c$
- Examples: $\lim_{x\to 1}\ \frac{5x^4-4x^2-1}{10-x-9x^3}$
- $lim_{x\to\infty}\ \frac{e^x}{x^2}$. Apply twice!

3.1 More practice with l'hopital's rule

- L'hopital's rule comes up a bunch when you talk about convergence, because we're dealing with limits. So let's recall what l'hopital's rule tells us and do a couple more examples.
- L'hopital's rule: if $lim_{x\to c}\ f(x) = lim_{x\to c}\ g(x) = 0$ or $= \pm\infty$, then $\lim_{x\to c}\ \frac{f(x)}{g(x)} = \lim_{x\to c}\ \frac{f'(x)}{g'(x)}$
- $lim_{x\to\infty}\ \frac{x+x^2}{1-2x^2}$
- $lim_{x\to 0}\ \frac{e^x-1-x-\frac{1}{2}x^2}{x^3}$
- More complicated: $lim_{x\to\infty}\ (1+\frac{1}{x})^x$

4 Strategies for Integration

- Integration by parts: $\int f'(x)g(x) = f(x)g(x) - \int f(x)g'(x)$
- Integration by substitution: reverse chain rule.
- Integrals can "converge" if as you take them to infinity or negative infinity they approach zero or some other number. This will allow a definite integral with bounds that include infinity or negative infinity to be evaluated.

4.1 Expected Utility Functions

- Requires in some sense a cardinal utility function, as relative "amounts" of utility matter.
- Simple discrete example with p chance of getting $x = 10$ and $1-p$ chance of getting $x = 5$ would lead to $EU(x) = pu(10) + (1-p)u(5)$
- For continuous case, just integrate over any random variables. E.g. if x is distributed uniformly from zero to two, then $p(x) = 1/2$ for $x \in [0, 2]$ and 0 otherwise. If $u(x) = \sqrt{x}$ then $EU(x) = \int_0^2\ \sqrt{x}(1/2)dx = |_0^2\ \frac{x^{3/2}}{2(3/2)} = \frac{2^{3/2}}{3}$
- Certainty equivalent: figure out what amount of x, when given with certainty, produces same utility as the gamble. In this case, solve:
  $U(x) = \sqrt{x} = \frac{2^{3/2}}{3} \leftrightarrow x = (\frac{2^{3/2}}{3})^2 = \frac{8}{9}$
- To find risk premium, figure out the amount that the risky asset's expected return has to exceed the certain asset in order to generate equivalent utilities.
- In this case, compute $\int_0^2\ x(1/2)dx - \frac{8}{9} = |_0^2 \frac{x^2}{4} - \frac{8}{9} = 1 - \frac{8}{9} = \frac{1}{9}$
- Keep in mind that having a concave utility function (like $\sqrt{x}$) is equivalent to having decreasing returns which is equivalent to risk aversion.
- Can also be risk loving if utility function is convex, or risk neutral if utility function is linear.

5 Rules of Variance and Covariance

- $Var(a + bX) = b^2 Var(X)$
- $Var(a + bX + cY) = b^2 Var(X) + c^2 VAR(Y) + 2bcCov(X, Y)$
- $Var(c) = 0$
- $Cov(X, Y) = E(XY) - E(X)E(Y)$. Note, this is zero if X and Y are independent, as in this case $E(XY) = E(X)E(Y)$
- $Cov(X + c, Y + b) = Cov(X, Y)$
- $Cov(cX, bY) = cbCov(X, Y)$
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
- $Cov(X, X) = Var(X)$

6 Systems of equations: numbers of solutions

- When we reduce a coefficient matrix to RREF, we can use the resulting matrix to determine whether we have one solution, infinite solutions, or no solutions (i.e. system is inconsistent).
- Consider an example (including different solutions $\boldsymbol{b}$):
$\begin{bmatrix} 1 & b \\ 2 & c-1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \boldsymbol{b}$
- For what values of c is this of full rank? For what values of b and c does this have infinite solutions? How about no solutions?

7 Orthogonality

- When dot product of vectors equals zero, i.e. $\boldsymbol{u} \cdot \boldsymbol{v} = 0$
- Also can be written $\boldsymbol{u}'\boldsymbol{v}$ (if column vectors)
- Consider following set of vectors.
$\begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$
- Are these vectors linearly independent? (Can use RREF to find out. If linearly independent, the rank of the matrix should equal the number of vectors.)
- What is the full set of vectors that is orthogonal to all of these vectors?

8 Laplace expansions and Cramer's rule

- Cramer's rule is a way to use determinants to solve systems of equations.
- Recall that Laplace (or cofactor) expansion looks like:
  $|A| = \Sigma_{k=1}^{n} a_{in} C_{in} = a_{i1}C_{i1} + a_{i2}C_{i2} ... a_{in}C_{in}$ where $C_{ij} = (-1)^{i+j} M_{ij}$
- C_{ij} is ijth cofactor.
- Cramer's rule states that x_i of system is equal to $\frac{|A_i|}{|A|}$
- Where A_i is coefficient matrix but with ith column replaced by solution values.
- Simple example:

$$
\begin{aligned}
&4x + 3y - 2z = 7 \\
&x + y = 5 \\
&3x + z = 4
\end{aligned}
$$

- Solve using Cramer's rule.
- Now recall that $adj(A) = C'$, where C is matrix of cofactors.
- Inverse is $\frac{1}{|A|} Adj(A)$
- Now solve using matrix inverse.

9 Spanning sets, dimension

- Let $p = \begin{bmatrix} x_1 \\ x_2 \\ x_2 \end{bmatrix}$ such that $-2x_1 + x_2 - 3x_3 = 0$. Is this a subspace?
- Let $\boldsymbol{v_1} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \boldsymbol{v_2} = \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$ Are these vectors linearly independent?
- Are there any $\boldsymbol{v} \in \mathbb{R}^3$ that you could add to $\boldsymbol{v_1}, \boldsymbol{v_2}$ and still have a linearly independent set? **Yes. Because would need 3 vectors to span $\mathbb{R}^3$.**
- Let's show that $\boldsymbol{v_1}, \boldsymbol{v_2}$ fall on the above plane, and span the plane.
- Given this, is there any vector on the plane which could be added to the set and still be linearly independent? **No. Otherwise you would have one basis that is two dimensions and one basis that is three dimensions.**
- Note that we have a plane, and it is two dimension.

10 Practice with matrix algebra

- What is $(\boldsymbol{y} - \boldsymbol{X\beta})'$?
- Important matrix transpose properties: $(\boldsymbol{A}')' = \boldsymbol{A}$
- Additive: $(\boldsymbol{A} + \boldsymbol{B})' = \boldsymbol{A}' + \boldsymbol{B}'$

- $(\boldsymbol{AB})' = \boldsymbol{B'A'}$

- $(\boldsymbol{cA})' = \boldsymbol{cA'}$

- $(\boldsymbol{A^{-1}})' = (\boldsymbol{A'})^{-1}$

- What is $\boldsymbol{ee'}$ versus $\boldsymbol{e'e}$? (Think, $n \times 1$ and $1 \times n$ versus $1 \times n$ and $n \times 1$).

- Recall from math camp we did this:
  A matrix is idempotent if multiplying it by itself returns the same matrix (i.e. $\boldsymbol{AA} = \boldsymbol{A}$). Prove that $\boldsymbol{I} - \boldsymbol{X(X'X)^{-1}X'}$ is idempotent.
  Ans:
  $(\boldsymbol{I} - \boldsymbol{X(X'X)^{-1}X'})(\boldsymbol{I} - \boldsymbol{X(X'X)^{-1}X'})$
  $= \boldsymbol{II} - \boldsymbol{2IX(X'X)^{-1}X'} + \boldsymbol{X(X'X)^{-1}X'X(X'X)^{-1}X'}$
  $= \boldsymbol{I} - \boldsymbol{2X(X'X)^{-1}X'} + \boldsymbol{X(X'X)^{-1}X'}$
  $= \boldsymbol{I} - \boldsymbol{X(X'X)^{-1}X'}$

- The above matrix is called the "residual maker", often denoted $\boldsymbol{M}$. Why might this be the case?

- Recall that the regression equation gives us a best fit for $\boldsymbol{y} = \boldsymbol{X\beta}$ by choosing $\boldsymbol{\beta}$ to minimize the sum of squares.

- The OLS estimate of $\boldsymbol{\beta}$ is $\boldsymbol{(X'X)^{-1}X'y}$, as you will show in question 1 of the problem set. So what happens if you compute $\boldsymbol{My}$?

- Answer: You get the residuals!

11 Eigenvalues and eigenvectors

- Recall from in class that we find eigenvalues by solving $\boldsymbol{Av} = \lambda\boldsymbol{v}$ for λ.

- Eigenvalues and eigenvectors have a lot of useful applications, from the factor analysis that Iain talked about, to determining whether or not a matrix is positive or negative semidefinite.

- Determining whether a matrix is positive of negative semidefinite allows you to determine whether a multivariable function is concave or convex, which in turn allows you to determine whether or not a critical point is a minimum, maximum, or neither.

- If all eigenvalues of a Hessian matrix are negative, for instance, this means the matrix is negative definite, which is equivalent to saying the function is strictly concave, which would suggest that a critical point is a maximum.

- Let's do an example with a 3×3 matrix, where we find the eigenvalues of the following matrix $\boldsymbol{A}$.
  $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

- Recall that to solve $\boldsymbol{Av} = \lambda\boldsymbol{v}$ for λ, we rearrange to get $(\boldsymbol{A} - \boldsymbol{I}\lambda)\boldsymbol{v} = 0$.

- To avoid only getting the trivial solution of $\boldsymbol{v} = \boldsymbol{0}$, we want $det(\boldsymbol{A} - \boldsymbol{I}\lambda) = 0$, such that the matrix is singular and we'll get a set of vectors that satisfy $\boldsymbol{v}$. We compute:
  $det(\boldsymbol{A} - \boldsymbol{I}\lambda) = \left| \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \right| = \begin{vmatrix} 1-\lambda & 2 & 0 \\ 0 & 1-\lambda & 0 \\ 1 & 1 & 1-\lambda \end{vmatrix}$

- Which we can do easily by doing a Laplace/cofactor expansion on row 2, which has a bunch of zeros: $(0)(-1)^{2+1}\begin{vmatrix}2 & 0\\ 1 & 1-\lambda\end{vmatrix} + (1-\lambda)(-1)^{2+2}\begin{vmatrix}1-\lambda & 0\\ 1 & 1-\lambda\end{vmatrix} + (0)(-1)^{2+3}\begin{vmatrix}1-\lambda & 2\\ 1 & 1\end{vmatrix}$
$= 0 + (1-\lambda)(1)(1-\lambda)^2 + 0 = (1-\lambda)^3$

- Which equals zero if $\lambda = 1$. Thus, $\lambda = 1$ is our one and only eigenvalue. To determine eigenvectors, we substitute in $\lambda = 1$ into $(\boldsymbol{A} - \boldsymbol{I}\lambda)\boldsymbol{v} = 0$ and solve.

$$\begin{bmatrix}1-\lambda & 2 & 0\\ 0 & 1-\lambda & 0\\ 1 & 1 & 1-\lambda\end{bmatrix}\boldsymbol{v} = \begin{bmatrix}0 & 2 & 0\\ 0 & 0 & 0\\ 1 & 1 & 0\end{bmatrix}\boldsymbol{v} = 0$$

- Which gets us $x_1 = -x_2 = 0$, $x_3 = x_3$. So x_3 can take on any value, but the other two entries need to be zeros if $\boldsymbol{v}$ is an eigenvector. We can check this with an example, to see if it satisfies $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$.

$$\boldsymbol{A}\boldsymbol{v} = \begin{bmatrix}1 & 2 & 0\\ 0 & 1 & 0\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix}0\\ 0\\ \pi\end{bmatrix} = \begin{bmatrix}0+0+0\\ 0+0+0\\ 0+0+\pi\end{bmatrix} = \begin{bmatrix}0\\ 0\\ \pi\end{bmatrix}$$

$$\lambda\boldsymbol{v} = (1)\begin{bmatrix}0\\ 0\\ \pi\end{bmatrix} = \begin{bmatrix}0\\ 0\\ \pi\end{bmatrix}$$

- So we have $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$ as desired.

12 ZOMG Regression!!!!11

- Intuition from the basics: what are we doing when we look at a single-variable regression, i.e. a model of form $y = \beta_0 + \beta_1 x_1 + e$?
- Question: So if all we are interested in is the effect of x_1 on y, why don't we just do this all the time?
- Model: Drowning deaths $= \beta_0 + \beta_1$ice-cream sales $+ e$ What's the issue?
- "Lurking" variables, or omitted variable bias. Classic case: where a dependent variable of interest is related to some other dependent variable *and* the independent variable.
- Question: Suppose the true model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$. Say we initially estimate $y = \beta_0 + \beta_1 x_1$. Will including x_2 reduce the bias in our estimate of β_1?
- Follow up question: what are the conditions under which the above has different answers?
- Say A causes B causes C. Should we include both A and B?
- In any event, we need a mechanism of "controlling" for omitted variables that we feel may be confounding our analysis.
- Multiple regression does this, in some sense. We "partial out" the effects of other independent variables in order to isolate the effect of a single variable.
- Note: we can obtain estimates for β_1 in a two variable model in a way that illustrates the partialling out interpretation well. Regress x_1 on x_2, obtain the residuals r_1, then regress y on thse residuals r_1. This will give the estimate of β_1 when the effects of x_2 have been partialled out, and is the same as what we would have gotten from doing multiple regression in the first place (though not the same standard errors). See code at end of notes to try it out.

- Predicting bias direction with two variables:

	$Corr(x_1, x_2) > 0$	$Corr(x_1, x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

- Things get more complicated when the true model includes more than two variables. For instance, say you have x_3 which is uncorrelated with x_1 but is correlated with x_2. Does not including x_3 induce bias in our coefficient for x_1?
- Answer is yes, if x_2 is correlated with x_1. Doesn't matter that x_3 is not directly correlated with x_1.
- Say our correct model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. Say we start off with just x_1, and say that we know both x_2 and x_3 are correlated with both x_1 and y. Does our bias decrease when we go from including only including x_1 to including x_1 and x_2?
- The answer: not necessarily! This is the subject of Kevin Clarke's wonderfully-titled paper: *The Phantom Menace: Omitted Variable Bias in Econometric Research.*
- If x_2 introduces negative bias and x_3 introduces positive bias, then including only one and not the other means you could be further from the truth than with neither.
- As a result, unless we have the fully specified model, we can't even know if including a variable that belongs in the model with increase or decrease the bias on the coefficient estimate of interest.
- Summary question: You are interested in the effect of x_1 on y. x_2 is also part of the true model. Should you include it?

12.1 OLS estimator derivation

Ordinary least squares linear regression is based on minimizing the squared differences between your regression "line" (hyperplane) and your observed data. Same deal as what we did earlier with least squares estimators for the mean. So, want to minimize $e'e$ where $e = y - XB$ (can you see why this equation holds?)

$$\begin{aligned} min_B \ (y - XB)'(y - XB) &= (y' - B'X')(y - XB) \\ &= y'y - B'X'y - y'XB + B'X'XB \\ &= y'y - 2B'X'y + B'X'XB \end{aligned}$$

taking derivative with respect to B and setting to zero returns

$$\begin{aligned} -2X'y + 2X'XB &= 0 \\ \leftrightarrow X'XB &= X'y \text{ (Note, this is sometimes called the normal equation(s))} \\ \leftrightarrow (X'X)^{-1}X'XB &= (X'X)^{-1}X'y \\ \leftrightarrow B &= (X'X)^{-1}X'y \end{aligned}$$

12.2 Omitted variable bias and partitioned regression

Let's see how to mathematically represent omitted variable bias using partitioned matrices. First, let's consider the initial "normal equation" for regression, but with a data matrix I will label $\boldsymbol{X_1}$ for reasons that will become obvious after.

$$\boldsymbol{X_1'X_1b_1 = X_1'y}$$

Which leads to the regression equation to solve for b:

$$\boldsymbol{b_1} = (\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'y}$$

Now, let's say that $\boldsymbol{X_1}$ includes all the variables except one, denoted $\boldsymbol{X_2}$. Now say we want to add this in. We can represent this in a partioned matrix like so.

$$\begin{bmatrix}\boldsymbol{X_1} & \boldsymbol{X_2}\end{bmatrix}'\begin{bmatrix}\boldsymbol{X_1} & \boldsymbol{X_2}\end{bmatrix}\begin{bmatrix}\boldsymbol{b_1}\\ \boldsymbol{b_2}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X_1} & \boldsymbol{X_2}\end{bmatrix}'\boldsymbol{y}$$

Components of partitioned matrices can mostly be treated the same way as you would treat elements of regular matrices, while keeping in mind a few things when you do things like transposing or finding inverses.

$$\begin{aligned}
&\leftrightarrow \begin{bmatrix}\boldsymbol{X_1'}\\ \boldsymbol{X_2'}\end{bmatrix}\begin{bmatrix}\boldsymbol{X_1} & \boldsymbol{X_2}\end{bmatrix}\begin{bmatrix}\boldsymbol{b_1}\\ \boldsymbol{b_2}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X_1'}\\ \boldsymbol{X_2'}\end{bmatrix}\boldsymbol{y}\\
&\leftrightarrow \begin{bmatrix}\boldsymbol{X_1'X_1} & \boldsymbol{X_1'X_2}\\ \boldsymbol{X_2'X_1} & \boldsymbol{X_2'X_2}\end{bmatrix}\begin{bmatrix}\boldsymbol{b_1}\\ \boldsymbol{b_2}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X_1'y}\\ \boldsymbol{X_2'y}\end{bmatrix}
\end{aligned}$$

Let's start to multiply out to try to solve for $\boldsymbol{b_1}$, solving for the part of the left hand side equal to $\boldsymbol{X_1'y}$

$$\begin{aligned}
\boldsymbol{X_1'X_1b_1} + \boldsymbol{X_1'X_2b_2} &= \boldsymbol{X_1'y}\\
\leftrightarrow \boldsymbol{X_1'X_1b_1} &= \boldsymbol{X_1'y} - \boldsymbol{X_1'X_2b_2}\\
\leftrightarrow (\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'X_1b_1} &= (\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'y} - (\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'X_2b_2}\\
\leftrightarrow \boldsymbol{b_1} &= (\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'y} - (\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'X_2b_2}
\end{aligned}$$

What do you notice about $(\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'y}$? It's the same as the initial regression estimator when we didn't have the omitted variable! So when will our estimate for $\boldsymbol{b_1}$ be the same with the added variable as it was without it? When $(\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'X_2b_2}$ is equal to zero. Let's unpack the components of this. $(\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'X_2}$ is a regression of $\boldsymbol{X_2}$ on $\boldsymbol{X_1}$, and $\boldsymbol{b_2}$ is a measure of the effect of $\boldsymbol{X_2}$ on $\boldsymbol{y}$. This is all very similar to our initial bias table. Note that $\boldsymbol{X_2}$ can be generalized to a set of omitted variables instead of just one omitted variable.

12.3 Quick additional note on partioned matrices

- For PS4, you had two variables, and one of the variables was a scalar multiple of the other. Many interpreted this as have a 2x2 data matrix.
- In fact, $k = 2$ but n can be anything. It is straightforward to show the result we want (that making one of the variables a scalar multiple means we can't invert $\boldsymbol{X'X}$) if we use a simple partioned matrix.
- $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{x_1} & r\boldsymbol{x_2}\end{bmatrix}$, $\boldsymbol{X'} = \begin{bmatrix}\boldsymbol{x_1'}\\ \boldsymbol{x_2'}\end{bmatrix}$
- If $\boldsymbol{x_2} = r\boldsymbol{x_1}$ then $\boldsymbol{X'X} = \begin{bmatrix}\boldsymbol{x_1'}\\ r\boldsymbol{x_1'}\end{bmatrix}\begin{bmatrix}\boldsymbol{x_1} & r\boldsymbol{x_1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{x_1'x_1} & \boldsymbol{x_1'}r\boldsymbol{x_1}\\ r\boldsymbol{x_1'x_1} & r\boldsymbol{x_1'}r\boldsymbol{x_1}\end{bmatrix}$
- Which means that $det(\boldsymbol{X'X}) = \boldsymbol{x_1'x_1}r^2\boldsymbol{x_1'x_1} - r^2\boldsymbol{x_1'x_1x_1'x_1} = 0$, and thus it's not invertible.

13 Definiteness of matrices, etc.

- Let's do some examples:
- $\begin{bmatrix}-1 & 1\\ 1 & -3\end{bmatrix}$
- Leading principals minors: First: -1. Second: $3 - 1 = 2$. Thus, the leading principals alternate according to $(-1)^k$ for kth order leading principal, i.e. start negative than alternate positive and negative. This implies the matrix is negative definite.

- $\begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}$
- Leading principal minors: First: 1. Second: 0. So we can't say that it's positive definite, but it might be positive semi-definite, if *all* principal minors are nonnegative. So in this case, check the other 1st order principal minor, which is 9. Thus, in this case, the matrix is positive semidefinite, but not positive definite.
- $\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$
- Leading principal minors: First: 2. Second: 3. Third: $4(4-1) = 12$. All leading principal minors are positive, so the matrix is positive definite.
- Now let's apply this to some problems we dealt with in math camp.
- Find the critical point of $f(x, y, z) = 2x^2 + y^2 + z^2 + 8x + 6y + 8z$, and determine whether it's a minimum or maximum.
- First order conditions: $f_x = 4x + 8 = 0$, $f_y = 2y + 6 = 0$, $f_z = 2z + 8 = 0$, implies critical point $(x, y, z) = (-2, -3, -4)$
- Now let's take a look at the Hessian, taking derivatives of each of these derivatives with respect to each of the variables (i.e. constructing the matrix of second derivatives and cross partials).
  $\begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$
- Leading principal minors: First: 4 Second: 8 Third: 16. All are positive, which means the matrix is positive definite, which means the critical point is a minimum.
- Find the critical point of $f(x, y) = -x^2 - y^2 + xy + 3y$ and identify if it's a minimum or maximum.
- First order conditions: $f_x = -2x + y$, $f_y = -2y + x + 3$. Solve to get $(x, y) = (1, 2)$.
- Hessian: $\begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$
- Leading principal minors: First: -2 Second: 3. They alternate starting with negative, so this matrix is negative deifnite, which means the critical point is a maximum.

14 Comparative statics with arbitrary function

- Now let's look at a problem where we don't know the explicit functional form, but we want to figure out how a choice variable changes when we adjust the value of a parameter. We will look at a classic problem from economics, where a firm is trying to maximize their profits, with a revenue function $F(L, K)$, where L is units of labour (it might be hours of work) and K is units of capital. For each unit of each, a cost must be paid, with w the wage being the cost for a unit of labour, and r the rent being the cost for a unit of capital. Thus, the profit maximization problem is:
- $max_{L,K}\ \pi = F(L, K) - wL - rK$
- Our approach is familiar, in that we start by taking first order conditions (FOCs):
  $\frac{\partial \pi}{\partial L} = F_L - w = 0$
  $\frac{\partial \pi}{\partial K} = F_K - r = 0$

- However, in this case we cannot explicitly solve for the optimal L^* and K^*, as we do not know the explicit form of $F(K, L)$. If we could solve it, then to find the comparative statics, we could do what we did in an earlier problem set, and just find L^* and then take the derivative with respect to a parameter. For instance, if we wanted to figure out whether the optimal choice of labour decreases with an increase in the wage w, we would just solve for L^* and take the derivative with respect to w. However, given that we can't solve it, we will need to do a bit more work.

- Before we move on, let's write down the Hessian, and keep track of it for later:
$H = \begin{bmatrix} F_{LL} & F_{LK} \\ F_{KL} & F_{KK} \end{bmatrix}$
(Recall that this notation represents $\frac{\partial^2 F}{\partial L^2} = F_{LL}$ and $\frac{\partial^2 F}{\partial L \partial K} = F_{LK}$

- Which means our principal minors will be: First order: F_{LL} (leading) and F_{KK} (not leading). Second: $F_{LL}F_{KK} - F_{LK}F_{KL}$

- Now let's say we want to get the comparative statics with respect to w. Let's return to our FOCs are implicitly differentiate both sides with respect to w to get:
$F_{LL}\frac{\partial L^*}{\partial w} + F_{LK}\frac{\partial K^*}{\partial W} - 1 = 0$
$F_{KL}\frac{\partial L^*}{\partial w} + F_{KK}\frac{\partial K^*}{\partial W} = 0$

- We now have two equations in two unknowns, if we think of the unknowns as $\frac{\partial L^*}{\partial w}$ and $\frac{\partial K^*}{\partial}$. We can put these together in a coefficient matrix and use what we know about matrices and inverses to solve for the comparative statics.
$\begin{bmatrix} F_{LL} & F_{LK} \\ F_{KL} & F_{KK} \end{bmatrix} \begin{bmatrix} \frac{\partial L^*}{\partial w} \\ \frac{\partial K^*}{\partial W} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
$\leftrightarrow \begin{bmatrix} \frac{\partial L^*}{\partial w} \\ \frac{\partial K^*}{\partial W} \end{bmatrix} = \frac{1}{F_{LL}F_{KK} - F_{LK}F_{KL}} \begin{bmatrix} F_{KK} & -F_{LK} \\ -F_{KL} & F_{LL} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{F_{KK}}{F_{LL}F_{KK} - F_{LK}F_{KL}} \\ \frac{-F_{KL}}{F_{LL}F_{KK} - F_{LK}F_{KL}} \end{bmatrix}$

- These comparative statics are completely general, and phrased in terms of the partial derivatives. But now we can imagine how we might use these by putting some structure on the question. For instance, what if we assumed that the revenue function is concave, which is equivalent to saying that the Hessian is negative semidefinite? Then we know that $F_{LL} \leq 0$, $F_{KK} \leq 0$ and $F_{LL}F_{KK} - F_{LK}F_{KL} \geq 0$, from the principal minors found above.

- Now look at the expression for $\frac{\partial L^*}{\partial w}$ (i.e. the partial representing how the optimal labour choice changes with the wage). Let's assume as well that all the inequalities derived from concavity are strict (which amounts, in part, to saying that the function is strictly concave in each choice variable). Then because $F_{KK} < 0$ and $F_{LL}F_{KK} - F_{LK}F_{KL} > 0$, we know that $\frac{\partial L^*}{\partial w} < 0$, or the optimal choice of labour is decreasing in the wage. This makes intuitive sense! We would expect someone to hire fewer workers or employ them for less time if it is more costly to do so. However, this approach also makes it clear exactly what assumptions are required for that conclusion to be true.

15 Unbiasedness, convergence in quadratic mean, discounting, etc.

- Unbiasedness just means that the expected value of the estimator is equal to quantity we want.

- We spoke earlier of omitted variable bias. This is an example of conditions under which OLS will not be unbiased, i.e. our estimates of the relevant quantites with not be on average correct.

- $plim(x) = c$ if $lim_{n\to\infty} Pr(|x - c| > \epsilon) = 0$. In words, if as n approaches infinity the probability that x will be different than c approaches zero, the probability limit of x is c.

- Convergence in quadratic mean requires convergence in the first moment (i.e. the expected value) as well as the second moment (i.e. the variance). If $E(x)$ converges to c and $Var(x)$ converges to 0, then $plim(x) = c$

- With discounted payoff y over an infinite period of time, you have an infinite geometric series, which converges to value $\frac{y}{1-\delta}$ for $\delta \in (0,1)$. This can be shown using the telescoping series stuff Iain talked about, but you can just take it as a given for the problem set (and, most likely, generally in life).

16 Constrained Optimization

- We've seen in lecture that equality-constrained optimization can be handled pretty straightforwardly by constructing the Lagrangian with the constraint added in, and then taking an extra derivative with respect to λ and solving the new system. From forever ago (math camp) you may recall that I gave a snake fighting problem on a problem set, which is reproduced here:

  > Now, assume that you derive positive utility from shrinking the size of the snake by working on your dissertation (think of this as negative negative utility), and that this positive utility has decreasing marginal returns in the number of hours per day x you spend working on your dissertation (completing ANYTHING shrinks the snake a lot, but improvements afterwards have a higher work to snake-shrinking ratio). This utility will be captured in this problem by $12\sqrt{x}$. Now assume that you could also spend time honing your snake-battling skills, by practicing snake-fighting, trapping, learning the flute, etc. Denote the time spent on this as t, and the utility gained from this as $3t$ (so your utility from snake-battling training increases at a constant rate/does not experience decreasing marginal returns). Now assume that you have 16 hours a day to allocate between your thesis and snake-fight training, such that $t = 16 - x$. Thus, your utility function can be expressed as:
  >
  > $$U(x,t) = 12\sqrt{x} + 3t = U(x) = 12\sqrt{x} + 3(16 - x)$$

- You could solve this by just taking a derivative and setting to zero. You might notice, howver, that this is essentially an equality-constrained maximization problem! I have built the constraint into the problem, by making the number of hours spent on snake-fight training $16 - x$.

- The Lagrange approach intuition has some similarities. This is what Iain means when he says that each constraint "nails down" a particular variable; in this case, one constraint nails down the y in terms of x

- The answer in this case was $x = 4$ hours on the dissertation and 12 hours spent training to fight snakes. We can get the same answer by setting up the Lagrangian.

- $\mathcal{L}(x,t,\lambda) = 12\sqrt{x} + 3t + \lambda(16 - x - t)$
  $\frac{\partial \mathcal{L}}{\partial x} = 6x^{-1/2} - \lambda = 0$
  $\frac{\partial \mathcal{L}}{\partial t} = 3 - \lambda = 0$
  $\frac{\partial \mathcal{L}}{\partial \lambda} = 16 - x - t = 0$

- Which we can solve, noting that the second expression gives us $\lambda = 3$. Which means that $6x^{-1/2} = 3 \leftrightarrow 2 = \sqrt{x} \leftrightarrow x = 4$. We then plug into the third expression to get t, i.e. $16 - 4 - t = 0 \leftrightarrow t = 12$

- Inequality constraints, and Kuhn-Tucker (KT) conditions, are grosser looking, but are essential just a way of formalizing the idea of checking whether or not each constraint is binding or not.

- Sometimes we impose assumptions on the problem to ensure that we know which constraints will bind and which won't without having the check each case. For instance, if everyone always wants more of every good (monotonic preferences) then the budget constraint will bind.

- So let's go back to the snake-fighting problem, make the constraint an inequality one $x + t < 16$ and add in another inequality constraint that $t > 1$. Perhaps you live in a snake-infested area and are likely to be attacked by snakes for an hour a day irrespective of your choices. We can set up the Lagrangian: $\mathcal{L}(x,t) = 12\sqrt{x} + 3t + \lambda_1(16 - x - t) + \lambda_2(t - 1)$

- Where we only take derivatives with respect to variables and binding constraints.
$\frac{\partial \mathcal{L}}{\partial x} = 6x^{-1/2} - \lambda_1 = 0$
$\frac{\partial \mathcal{L}}{\partial t} = 3 - \lambda_1 + \lambda_2 = 0$

- This gives us 6 Kuhn-Tucker conditions in addition to the original three first order conditions:
$\lambda_1 \geq 0$, $16 - x - t \geq 0$, $\lambda_1(16 - x - t) = 0$
$\lambda_2 \geq 0$, $t - 1 \geq 0$, $\lambda_2(t - 1) = 0$

- Any points that satisfying the first order conditions and KT conditions are fair game for minima and maxima. As mentioned earlier, we might be able to rule out certain constraints binding or not binding based on how we've structured the problem. Failing that, we need to check each condition, which can give us a number of "cases" that are candidates for solutions. See below:

- Assume $t = 1$ which implies $\lambda_2 \geq 0$. Two possible options: (1) $16 - x - 1 = 15 - x = 0$ and $\lambda_1 \geq 0$ or (2) $\lambda_1 = 0$ and $15 - x \geq 0$. Looking at (2), if we then look at the first FOC, we get $6x^{-1/2} = 0$, which cannot hold for any x, so this is not a candidate. Option (1) gives us $x = 15$, which with the first FOC gives us approximately $\lambda_1 = 1.5$. Plugging into the second FOC, we get $3 - 1.5 + \lambda_2 = 0 \leftrightarrow \lambda_2 = -1.5$, which contradicts the our initial assumption that $\lambda_2 \geq 0$. So this is also not a candidate. So we have ruled out $t = 1$, and can move on to evaluating the converse.

- Assume $t - 1 \geq 0$ and $\lambda_2 = 0$. Two possible options: (1) $16 - x - t = 0$ and $\lambda_1 \geq 0$ or (2) $\lambda_1 = 0$ and $15 - x \geq 0$. Once again, let's start with option 2. If $\lambda_1 = 0$, we have from the second FOC that $3 - \lambda_1 + \lambda_2 = 3 - 0 + 0 = 0$, which cannot be true, so this cannot be a candidate. So we are left with only the last candidate, which is where the constraint $16 - x - t$ holds with equality. This gives us $x = 4$, $t = 12$ as before, and we can check and see that all KT conditions are satisfied.

- Essentially, every KT condition gives us two possible options, which have to be checked against every other option in the other KT conditions. So we get 2^r cases for each r inequality constraints.

- Note, finally, that all of these conditions essentially correspond to assuming a constraint is binding or not, and then checking the implications for contradictions with the FOCs or the other KT conditions.

17 Game theory

17.1 Nash Equilibrium

- Games consist of a set of N players, S strategies, and U payoffs over strategy profiles (a vectors of each player's strategies).

- Nash equilibrium: All strategies are elements of each player's best response correspondence, given everyone else's strategies.

- Note that each strategy profile is a candidate for equilibrium, and Nash allows us to restrict our attention to a subset of strategy profiles.

- Implies: no player has an incentive to deviate from their current strategy, given the other player's strategies.

- Formally: s^* is Nash if $\forall i \in N,\ u_i(s_i^*) \geq u_i(s_i, s_{-i}^*), \forall s_i \in S_i$

- (With this notation, "$-i$" means every player outside of player i)

- Example:

		Player 2	
		A	B
Player 1	A	$1, 1$	$0, 0$
	B	$0, 0$	$0, 0$

- Two equilibria in pure strategies: (A, A) and (B, B). (B, B) is less intuitive, because it might seem that a player has nothing to lose by deviating to A in the "hopes" that the other player will change their strategy.
- This is true but irrelevant to whether something is a Nash equilibrium. Nash treats the other player's strategy as fixed and immovable, and asks, "do you have an incentive to deviate?". If you're player 1 and player 2 is playing B, you can't make yourself better off by changing.
- Nash is a theory of equilibria, and says nothing about the steps that people will take to get to that equilibrium; it is concerned with stability of strategy combinations.
- Different solution concepts can allow us to eliminate more of these strategy profiles; "trembling hand" perfect equilibria, for instance, would not include (B, B). But in each case you are imposing more implicit assumptions about behaviour, and inviting one to challenge you to justify these assumptions.

17.1.1 Brief aside on fixed points

- In lecture we will be discussing fixed point theorems, and your problem set has a question where you are required to demonstrate that an equilibrium is a fixed point of a vector of best response correspondences.
- A fixed point is when $y \in f(y)$ for a correspondence $f(y)$. The intuition behind this as far as game theory is concerned is that if $f : S \rightrightarrows P(S)$ (where S is the strategy space [i.e. set of strategy profiles] and $P(S)$ is the power set of the strategy space [i.e. set of all subsets of the strategy space]) is a vector of best response correspondences, then a fixed point of f is where each of the strategies is a best response to each other strategy. This is equivalent to saying that $y \in f(y)$ is a Nash equilibrium.

17.2 Mixed strategies

- Mixed strategies: Loosen the requirement that a strategy need be a deterministic set of actions. Now, players assigning probabilities to each pure strategy. This is subtlely but importantly different from assigning probabilities to actions (see games of imperfect recall).
- Mixed strategies are thus drawn from all convex combinations of pure strategies.
- Nash equilibrium in mixed strategies: the same deal applies with Nash equilibrium in pure strategies, in that each strategy must be a best response to every other player's strategy.
- However, keep in mind that in order to randomize, a player must be indifferent between every strategy that is assigned positive probability in equilibrium; if they were not, i.e. a particular pure strategy obtained higher expected utility, they would switch to playing that strategy all the time.
- This creates the somewhat weird result that the probabilities you play certain strategies depends on the payoffs to the *other* player of different outcomes.
- Examples of mixed strategy equilibria: rock, paper, scissors. If either person is playing anything other that $(1/3, 1/3, 1/3)$, then there is an incentive to deviate. For instance, if I play $(1/2, 1/4, 1/4)$, then you would have an incentive to always play paper. If the utility to winning is 1, the utility to tieing is 0, and the utility to losing is -1, then $EU(paper) = 0.5(1) + (0.25)0 + (0.25) - 1 = 0.25$, $EU(rock) = 0.5(0) + (0.25)(1) + (0.25)(-1) = 0$, $EU(scissors) = (0.5)(-1) + (0.25)(0) + (0.25)(1) = -0.25$.

- Let's consider as well the Bach or Stravinsky game (often referred to as the Battle of the Sexes game).

		Player 2	
		Bach	*Stravinsky*
Player 1	*Bach*	$2, 1$	$0, 0$
	Stravinsky	$0, 0$	$1, 2$

- This has two pure strategy equilibria, i.e. (B, B) and (S, S). To find the mixed strategy equilibrium, the payoffs to both strategies have to be equal for each player. Setting $p = Pr(Bach)$ for player 1 and $q = Pr(Bach)$ for player 2, we can find this as follows.

$$
\begin{aligned}
EU_1(B) &= EU_1(S) \\
2q + 0(1 - Q) &= 0q + 1(1 - q) \\
\leftrightarrow 3q &= 1 \\
\leftrightarrow q &= \frac{1}{3} \\
\leftrightarrow 1 - q &= \frac{2}{3}
\end{aligned}
$$

$$
\begin{aligned}
EU_2(B) &= EU_2(S) \\
1p + 0(1 - p) &= 0p + 2(1 - p) \\
\leftrightarrow 3p &= 2 \\
\leftrightarrow p &= \frac{2}{3} \\
\leftrightarrow 1 - p &= \frac{1}{3}
\end{aligned}
$$

- Thus, the mixed strategy equilibrium is, if $\sigma_i = (Pr(B), Pr(S))$, $\sigma_1 = (2/3, 1/3)$, $\sigma_2 = (1/3, 2/3)$. Note that the randomization probabilities of each player are determined by the *other* player's preferences; they must randomize in such a way that makes the other player indifferent.

17.3 Subgame perfection

- All strategies are Nash *in every subgame*, where a subgame can be loosely thought of as a game which, at that point, can be viewed as an independent game in it's own right (players can ignore the history).[1]
- (Time permitting at this point) draw simple game tree in class to talk about subgames, and something being an equilibrium in every subgame.

[1] Fun fact: if you look up the wikipedia article on subgames, the book cited is Jim Morrow's game theory book!