ECOM40006/90013 Econometrics 3

Week 4 Extras

Question 1: Projection matrices and residual makers

Let X be a $n \times k$ full rank matrix. Then, consider the linear model

$$y = X\beta + u.$$

In lectures, we defined the projection matrix and residual maker P_X and M_X as

- Projection/hat matrix: $P_X = X(X'X)^{-1}X'$
- Residual maker: $M_X = (I - X(X'X)^{-1}X') = I - P_X$

where both of these matrices are **idempotent** and **symmetric**.[1] We can do quite a few things with these matrices that will let us arrive at the same results that we would have come across in earlier courses on linear regression. In particular, take as given the following formula for the OLS estimator:[2]

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y, \tag{1}$$

along with other features such as $y = X\hat{\beta} + e$ or $e = y - X\hat{\beta}$. Now, using these features, show that the following properties hold.

(a) $M_X X = 0$ **(b)** $M_X y = e$ **(c)** $M_X u = e$

(d) $P_X M_X = M_X P_X = 0$ **(e)** $\hat{y}'e = 0$

These ones are less straightforward, but we will need to know these for tutorials and later work with linear regression:

(f) Let Z be a subset of X so that we can rewrite $X = [\,Z \;\; W\,]$ with W denoting all other variables in X. Show that

$$P_X P_Z = P_Z P_X = P_Z.$$

(g) Show that $M_1 M_X = M_X$, where $M_1 = I = \ell(\ell'\ell)^{-1}\ell' = I - P_1$ is the residual maker associated with ℓ, a $N \times 1$ vector of ones.

(h) Show that $M_1 e = e$.

[1] A matrix A is idempotent if $AA = A$, and symmetric if $A' = A$. These properties can be shown for both matrices, and I strongly recommend you do so first if you find these unfamiliar.

[2] We'll derive the OLS estimator itself in the following pre-tute.

Question 2: More block matrices

This question will clarify some points from lectures. Consider the linear model

$$y = X\beta + u,$$

where as usual, y and u are $n \times 1$, X is $n \times k$ and β is $k \times 1$. Take as given that the OLS estimator is

$$\hat{\beta} = (X'X)^{-1}X'y.$$

***Optional**: Derive the OLS estimator from the minimization problem* $\arg\min_\beta u'u$. *You will need the following vector calculus rules*

$$\frac{\partial a\beta}{\partial \beta} = a', \qquad \frac{\partial \beta' A \beta}{\partial \beta} = 2A\beta$$

for this to work, where a is $1 \times k$, β is $k \times 1$ and A is $k \times k$.

(a) First consider Question 3 part (c) from the Week 2 extra questions. Using this result, show, using spectral decomposition, that a symmetric and idempotent matrix A admits the decomposition

$$A = CC'.$$

Furthermore, take as given that $C'C = I$.[3]

(b) Taking into account that the residual maker M_X is symmetric and idempotent, decompose it into the form

$$M_X = I - X(X'X)^{-1}X' = CC',$$

and use this to show that

$$C'M_XX = 0 \qquad \text{and} \qquad C'X = 0.$$

(c) Suppose that $y \sim N(X\beta, \sigma^2 I_n)$. Show that $\hat{\beta}$ and $C'y$ are independent, where C comes from part (b) above.

[3]This is a property of spectral decomposition: when the matrix is symmetric the decomposition takes the form $A = H\Lambda H'$ with the property $HH' = H'H = I$. Basically C is a subset of H but it still has the same property.

Question 3: Probability distributions

In this question, we explore the background of the properties of distributions that may be used in class. These properties are very useful, but it's often quite difficult to use them properly when you're not sure about *why* they work. Most of the proofs of these properties make use of techniques that we should be comfortable with, so let's go through these properties in a bit more detail.

Considering that proofs by themselves can be difficult to get started on, each question will give you a few pointers as to where you should get started and what you should expect to get, if you wish to do them by yourself first.

Part 1. *The normal distribution (revision)*

(a) Let $X \sim N(\mu, \sigma^2)$. Take as given that any transformation of the form $aX + b$ is normally distributed.[4] Show that $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

- You'll find the properties of expectations and variance useful (specifically: mean and variance of a sum).

(b) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d $N(\mu, \sigma^2)$ random variables. Let $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ be the average of these RVs. Show that

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

What does this imply for the expression $\lim_{n\to\infty} \bar{X}_n$? That is: what happens to your answer as n becomes very large?

- The sum of i.i.d. normal RVs is also normal with mean $\mathbb{E}(\bar{X}_n)$ and variance $\mathrm{Var}(\bar{X}_n)$. Be careful with the last point in particular and *state explicitly* what properties you use to make this work!

Part 2. *The multivariate random normal distribution.* Suppose that y is a $n \times 1$ random vector with mean vector μ and covariance matrix Σ. We denote this as $y \sim N(\mu, \Sigma)$.

(c) Show that for non-random conformable matrices A and b, $Ay + b \sim N(A\mu + b, A\Sigma A')$.

- The result for linear transformations of RVs in part (a) extends to the multivariate format, so your derivations should look very similar.

(d) Prove that if $y \sim N(0, \Sigma)$, then for some non-random conformable matrices A and B, Ay and By are independent if $A\Sigma B' = 0$.

- This logical statement is in the form "p if q", so you should start by assuming that $A\Sigma B' = 0$ and then deriving a result that implies independence.

[4]The easiest way to prove this property is via the use of *moment generating functions*, which are mentioned briefly in the week 1 extras.

- Remember that independence $\not\Rightarrow$ zero covariance generally. However, an exception exists in the case of the normal distribution...
- Try and derive $\text{cov}(Ay, By)$ and go from there!

Part 3. *The chi-squared distribution.* The proofs relating to chi-squared distributions can be quite troublesome. We're going to work through these properties here. Recall that a chi-squared distribution is a sum of independent squared standard normal RVs – that is:

$$\sum_{i=1}^{N} z_i^2 \sim \chi_N^2 \qquad \text{where } z_i \overset{\text{i.i.d.}}{\sim} N(0,1) \quad \text{for all } i = 1, 2, \dots, N.$$

(e) If $u \sim N(0, I_N)$, then $u'u \sim \chi_N^2$.

- This is a special case of part (b) below.
- Note that u is a $N \times 1$ random vector, which should tell you something about $u'u$...

(f) Show that if A is a $N \times N$ symmetric idempotent matrix with rank q, then $u'Au \sim \chi_q^2$.

- Spectral decomposition is required. The trick to this one is that the elements in Λ depend on the ordering of eigenvectors used in Q, so you can rearrange Q such that Λ becomes a block matrix. What kind of block matrix? Question 1c) might help...
- You should also define a new RVec $y = Q'u$ and its distribution. If you derive the appropriate block matrix form for Λ, you'll find out very quickly why this comes in useful.

(g) Show that if $z \sim N(0, \Omega)$ where Ω is an $N \times N$ nonsingular matrix, then $z'\Omega^{-1}z \sim \chi_N^2$.

- Note that because Ω is a covariance matrix, it is automatically p.s.d.
- The nonsingular property means that Ω is actually p.d.
- The rough steps to this proof are as follows:

  (i.) use spectral decomposition on Λ

  (ii.) define a new matrix $\Lambda^{-1/2}$ so that $\Lambda^{-1/2}\Lambda^{-1/2} = \Lambda^{-1}$

  (iii.) find out what Ω^{-1} is in terms of Q and Λ

  (iv.) define a new matrix $H = Q\Lambda^{-1/2}Q'$ and let $y = Hz$.

  From here, the steps will be similar to those in part (f).

Question 4: Convergence in Probability

Prove the *Weak Law of Large Numbers* (WLLN), using the definition of convergence in probability. The theorem itself is reproduced below for you:

- **Theorem (Weak Law of Large Numbers).** Let $X_1, X_2, \ldots, X_n$ be i.i.d. RVs, each with finite mean and variance $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$ respectively. Then, define

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \text{for } n = 1, 2, \ldots.$$

  As $n \to \infty$, $S_n \xrightarrow{p} \mu$.

You may find the following theorem useful:

- **Theorem (Chebyshev's Inequality).** $\mathbf{P}(|X - \mu| \geq k\sigma^2) \leq \frac{1}{k^2}$, for any $k > 0$.

Question 5: Convergence in Distribution

Suppose that $\{X_n\}$ represents a sequence of RVs. For the following distributions, say whether or not any convergence in distribution takes place. For reference, the CDF of an $\text{Exp}(\lambda)$ distribution with parameter λ is given by

$$F_X(x) = \mathbf{P}(X \leq x) = 1 - e^{-\lambda x} \qquad \text{for } x \in \mathbb{R}_+.$$

and the CDF of a uniform distribution on $[a, b]$ is

$$F_X(x) = \frac{x - a}{b - a} \qquad \text{for } a \leq x \leq b \text{ with } a < b.$$

(a) $X_n \sim \text{Exp}(n^{2/n})$. (hint: find the limit of $n^{2/n}$ first. Try using the fact that $n = e^{\log n}$).

(b) $X_n \sim U[-\sqrt{n}, n]$ (hint: intuitively, what happens to this distribution when n increases?)

(c) Now suppose that $X_i \sim U[0, 1]$ and all draws are i.i.d. Find the limiting distribution of $X_n = n \min\{X_1, X_2, \ldots, X_n\}$. Some hints:

- Try checking the convergence of the tails of the distribution and backing out the answer from there. Recall that the tail of a distribution is simply $\mathbf{P}(X > x) = 1 - \mathbf{P}(X \leq x)$.
- For a uniform RV, the draw being bigger than a certain value is the same thing as saying that it falls within an interval. But what does that interval look like?
- You can also use the result that $\lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n = e^x$.

Question 6: The Monte Carlo method [R Exercise]

Note: This is a coding exercise that requires the use of supplementary code (provided).

In econometrics, the Monte Carlo method is a very useful numerical tool that complements the theory that one picks up during the course. The basic idea behind it is very simple: use averages to approximate theoretical quantities that we'd like to know about. Behind the scenes, a large number of random[5] draws from our specified distributions are needed, along with some logical thinking to get the code up and running.

(a) First, let's examine the concept of a *random seed*. Type the following commands in the order provided:

`set.seed(42)`, then `rnorm(5)`, then `set.seed(41)`, then `rnorm(5)`, then `set.seed(42)` again followed by `rnorm(5)`.

What happens here, and why is this useful for replication in particular?

(b) The Monte Carlo method can be used to illustrate both convergence in *probability* and convergence in *distribution*. Consider for example the simple regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i,$$

where $(\beta_0, \beta_1)' = (0, 2)$ is the vector of *true parameters*, and $x_i, e_i \overset{\text{i.i.d.}}{\sim} N(0, 1)$, where e_i is the usual econometric disturbance term. Furthermore, $y_i | x_i \sim N(\beta_0 + \beta_1 x_i, 1)$. Take as given that for a fixed sample size N, $\hat{\beta}_1 \xrightarrow{d} N(2, \frac{1}{n})$. We will illustrate this using Monte Carlo. The R-script `Week4ExtraQ6.R` contains code that will conduct the simulations for you. Within the **Part 1** code, explain the following:

(i.) `beta1.hat = matrix(nrow = N, ncol = 1)` – what does this give and what do you think this will be used for?

(ii.) The role of the `for` loop (i.e. what's actually happening when we write `for (i in 1:M)`?)

(iii.) `y = rnorm(N, mean = beta0 + beta1*x1, sd = sigma)`

(iv.) `OLS = coeftest(lm(y ~ x1))`, particularly what the `coeftest` command does

(v.) `beta1.hat[i] = OLS[2,1]`

(c) Run the code for part 1. In a discussion with a colleague, they tell you that "because the average value of $\hat{\beta}_1$ is 1.999 for $n = 100$, it doesn't equal 2 and therefore $\hat{\beta}_1$ doesn't converge to $\hat{\beta}$ in probability." Point out any problems you see in this statement.

(d) The code for part 1 provides a histogram of our estimated $\hat{\beta}_1$ values. Overlaid in green is the actual theoretical distribution of $\hat{\beta}_1$ given the sample size of 100 in this case. Does the histogram seem close to its theoretical density?

(e) Run the code for part 2. Does the variance of the sampling distribution seem consistent with $\hat{\beta}_1 \xrightarrow{d} N(2, \frac{1}{n})$?

[5] Technically *pseudo-random*: there is quite the philosophical debate about the possibilities of drawing *truly* random numbers from a computer program, but that's not important for what we want to do.