ECOM40006/90013 Econometrics 3

Week 4 Extras: Solutions

Question 1: Projection matrices and residual makers

Note: in lectures, $u = y - X\beta$ denotes the error term, while $e = y - X\hat{\beta}$ denotes the OLS residuals. Because this notation makes it extremely difficult for me to track typos, I will use $\hat{u}$ to denote residuals in place of e. The TeX code can switch between these at will so if you want notation consistent with lectures get in touch with me and I'll send you the code.

The first five parts are straight derivations.

(a) By definition of M_X

$$\begin{aligned} M_X X &= (I - X(X'X)^{-1}X')X \\ &= X - X(X'X)^{-1}X'X \\ &= X - X \\ &= 0 \end{aligned}$$

as required.

(b) All we need to do is expand out:

$$\begin{aligned} M_X y &= (I - P_X)y \\ &= y - P_X y \\ &= y - X(X'X)^{-1}X'y \\ &= y - X\hat{\beta} \\ &= \hat{u} \end{aligned}$$

where the second last line follows by recognizing that $\hat{\beta} = (X'X)^{-1}X'y$ is the OLS estimator. Observe that this also implies that $P_X y = X\hat{\beta} = \hat{y}$, which we will use later in part (e).

(c) Using the fact that $u = y - X\beta$,

$$\begin{aligned} M_X u &= M_X(y - X\beta) \\ &= M_X y - M_X X\beta \end{aligned}$$

but as we have already found from parts (a) and (b), $M_X y = \hat{u}$ and $M_X X = 0$ so

$$M_X u = \hat{u} - 0\beta = \hat{u}.$$

(d) As in the previous parts, just apply the definitions and expand:

$$P_X(I - P_X) = P_X - P_X P_X = P_X - P_X = 0$$

since P_X is idempotent. Also observe that if $P_X M_X = 0$, then $(P_X M_X)' = 0'$ also. Using the rule that $(AB)' = B'A'$, we also have that $M_X P_X = 0$.

(e) Here we use two results: (i) that $\hat{y} = P_X y$ and (ii) $\hat{u} = M_X u$. Expanding out the left-hand side we get

$$\begin{aligned}\hat{y}'\hat{u} &= (P_X y)' M_X u \\ &= y' P_X M_X u \\ &= 0\end{aligned}$$

using the result that $P_X M_X = 0$ from part (d).

Now the remaining parts, as the question suggests, are a bit less straightforward. Let's see how we go with them!

(f) To see how this works, first expand everything out:

$$P_X P_Z = \underbrace{X(X'X)^{-1}X'}_{P_X}\underbrace{Z(Z'Z)^{-1}Z'}_{P_Z} = [X(X'X)^{-1}X'Z](Z'Z)^{-1}Z'.$$

Note that nothing actually changes in the last equality. The addition of square brackets here is for emphasis. In particular, it'd be quite nice to try and get rid of stuff in the square brackets to get us closer to our main goal.

It turns out we can! A trick can be used here: let

$$Z = X \begin{bmatrix} I \\ 0 \end{bmatrix} \qquad \text{with } I, 0 \text{ assumed comformable.}$$

Why is this important? Well, if we substitute $X = \begin{bmatrix} Z & W \end{bmatrix}$ in:

$$X \begin{bmatrix} I \\ 0 \end{bmatrix} = \begin{bmatrix} Z & W \end{bmatrix} \begin{bmatrix} I \\ 0 \end{bmatrix} = ZI + W0 = Z$$

Therefore, we have

$$\begin{aligned}P_X P_Z &= \left(X(X'X)^{-1}X' X \begin{bmatrix} I \\ 0 \end{bmatrix} \right) (Z'Z)^{-1}Z' \\ &= X \begin{bmatrix} I \\ 0 \end{bmatrix} (Z'Z)^{-1}Z' \\ &= Z(Z'Z)^{-1}Z' \\ &= P_Z\end{aligned}$$

by repeatedly using the definition of Z. This is the result that we were after!

(g) In a linear regression framework, the first column of X is a vector of ones, which represents the intercept term. Hence, we begin by noting that ℓ is a subset of X. Then,

$$\begin{aligned} M_1 M_X &= (I - P_1)(I - P_X) \\ &= I - P_X - P_1 + P_1 P_X \\ &= I - P_X - P_1 + P_1 \qquad \because \ \ell \text{ is a subset of } X \\ &= I - P_X \\ &= M_X. \end{aligned}$$

(h) Use the fact that $M_X u = \hat{u}$:

$$M_1 \hat{u} = M_1 M_X u = M_X u = \hat{u}.$$

Notice that we had to use the result from part (g) to do this particular series of steps.

Question 2: More block matrices

(a) Suppose that A is a $n \times n$ symmetric and idempotent matrix of rank r. Since it is symmetric it admits the spectral decomposition

$$A = H \Lambda H'$$

where $H' = H^{-1}$ so that $H'H = I_n$, and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Now, because A is also idempotent we can write

$$\Lambda = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

by appropriately rearranging the eigenvalues in Λ as well as their associated eigenvectors in H.[1] This can then be broken up as

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_r \\ 0 \end{bmatrix} \begin{bmatrix} I_r & 0 \end{bmatrix}$$

so if we define

$$C = H \begin{bmatrix} I_r \\ 0 \end{bmatrix}$$

then we can write $A = CC'$ and hence we have demonstrated the desired decomposition of A.

(b) Here, the residual maker M_X is a symmetric and idempotent matrix that can be shown to be of rank $n - k$.

[1] Yet another property of diagonalizing matrices. Best to refer to the math camp, session 5 for some more details. This was also addressed to some extent in the week 2 extras, specifically question 3.

Aside. If you want to show this explicitly, you need an auxiliary result, which is that the rank of an idempotent matrix equals its trace. Furthermore, you will also need the property that $\text{tr}(AB) = \text{tr}(BA)$. Now, let X be a full rank $n \times k$ matrix with $n > k$. Then,

$$\begin{aligned}
\text{rank}(M_X) &= \text{rank}(I_n - P_X) \\
&= \text{tr}(I_n - P_X) \\
&= \text{tr}(I_n) - \text{tr}(P_X) && \because \text{tr}(A+B) = \text{tr}(A) + \text{tr}(B) \\
&= n - \text{tr}(X(X'X)^{-1}X') \\
&= n - \text{tr}(X'X(X'X)^{-1}) && \because \text{tr}(AB) = \text{tr}(BA) \\
&= n - \text{tr}(I_k) \\
&= n - k
\end{aligned}$$

Hence, the rank of M_X is $n-k$. Perhaps we may derive this properly again in later classes!

Now back to the question. We can write

$$C'M_X X = C'0 = 0,$$

since $M_X X = 0$. But also we have

$$C'M_X X = C'CC'X$$

using the decomposition $M_X = CC'$. Furthermore, we can also show that[2]

$$\begin{aligned}
C'C &= \begin{bmatrix} I_{n-k} & 0 \end{bmatrix} \underbrace{H'H}_{=I} \begin{bmatrix} I_{n-k} \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} I_{n-k} & 0 \end{bmatrix} \begin{bmatrix} I_{n-k} & 0 \\ 0 & I_k \end{bmatrix} \begin{bmatrix} I_{n-k} \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} I_{n-k} & 0 \end{bmatrix} \begin{bmatrix} I_{n-k} \\ 0 \end{bmatrix} \\
&= I_{n-k}.
\end{aligned}$$

So if $C'C = I_{n-k}$ and $CC'C'X = 0$ then we can write

$$I_{n-k}C'X = 0 \implies C'X = 0,$$

as required.

(c) As per the question, let $y \sim N(X\beta, \sigma^2 I_n)$. Consider the block matrix

$$G = \begin{bmatrix} (X'X)^{-1}X' \\ C' \end{bmatrix}.$$

[2]For these derivations you might also like to convince yourself that an identity matrix can be re-partitioned into a block diagonal matrix of smaller identity matrices like we're about to do here.

Then multiplying it with y gives

$$Gy = \begin{bmatrix} (X'X)^{-1}X'y \\ C'y \end{bmatrix} = \begin{bmatrix} \hat{\beta} \\ C'y \end{bmatrix} \sim N(GX\beta, \sigma^2 GG')$$

where we note that since Gy is an affine transformation of y, Gy is also normally distributed. To see where the mean and variance comes from we can also derive them explicitly. In particular

$$\mathbb{E}(Gy) = GX\beta = \begin{bmatrix} (X'X)^{-1}X'X\beta \\ C'X\beta \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \end{bmatrix} \because C'X = 0.$$

As for the variance, we have

$$\text{Var}(Gy) = G\text{Var}(y)G' = \sigma^2 GI_n G' = \sigma^2 GG'.$$

For the expressions inside the variance term, it expands out in partitioned matrix form to read

$$GG' = \begin{bmatrix} A \\ B \end{bmatrix} \begin{bmatrix} A' & B' \end{bmatrix}$$

using the property of block matrix transposes. In this case we have $A = (X'X)^{-1}X'$ and $B = C'$. Putting it together we can derive

$$\begin{aligned} GG' &= \begin{bmatrix} (X'X)^{-1}X' \\ C' \end{bmatrix} \begin{bmatrix} X(X'X)^{-1} & C \end{bmatrix} \\ &= \begin{bmatrix} (X'X)^{-1}X'X(X'X)^{-1} & (X'X)^{-1}X'C \\ C'X(X'X)^{-1} & C'C \end{bmatrix} \\ &= \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & I_{n-k} \end{bmatrix}. \end{aligned}$$

Note that in the second line we used a few results.

- On the off-diagonal block elements we already know from before that $C'X = 0$, which gives us 0 in the lower left corner of GG'. But using this we can also write

  $$(C'X)' = X'C = (0)'$$

  which is also a zero matrix. For notational purposes we just abbreviate it to read 0.
- The lower right block element is $C'C$, which we showed was equal to I_{n-k} in part (b) above.

Finally we can put it all together. We can write

$$Gy = \begin{bmatrix} \hat{\beta} \\ C'y \end{bmatrix} \sim N\left[\begin{pmatrix} \beta \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \sigma^2 I_{n-k} \end{bmatrix} \right].$$

Since $\hat{\beta}$ and $C'y$ are jointly normally distributed, zero covariance (as seen on the block diagonal elements on the covariance matrix of Gy) implies that $\hat{\beta}$ and $C'y$ are also independent.

Aside. Actually, there is a much more straightforward way to show a very similar result without using any block matrices at all. The point of this exercise is basically to show that the OLS estimator is uncorrelated with the residuals $\hat{u}$. To see how, first observe that the residuals can be written as

$$\begin{aligned}\hat{u} = y - X\hat{\beta} &= y - X(X'X)^{-1}X'y \\ &= X\beta + u - X(X'X)^{-1}X'(X\beta + u) \\ &= X\beta + u - X(X'X)^{-1}X'X\beta - X(X'X)^{-1}X'u \\ &= X\beta + u - X\beta - X(X'X)^{-1}X'u \\ &= u - X(X'X)^{-1}X'u \\ &= (I - P_X)u \\ &= M_X u.\end{aligned}$$

At the same time we have also shown in our working above that

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

Then, we can write

$$\begin{aligned}\text{cov}(\hat{\beta}, \hat{u}) &= \text{cov}(\beta + (X'X)^{-1}X'u, M_X u) \\ &= \text{cov}((X'X)^{-1}X'u, M_X u) \\ &= \text{cov}(Au, Bu)\end{aligned}$$

where $A = (X'X)^{-1}X'$ and $B = M_X$. Using our rules for covariances we can write

$$\begin{aligned}\text{cov}(Au, Bu) &= A\text{cov}(u, u)B' \\ &= \sigma^2 AB' \\ &= \sigma^2(X'X)^{-1}X'M_X \\ &= 0\end{aligned}$$

since we know that $M_X X = 0$, implying that $X'M_X = (M_X X)' = 0'$ giving us our desired result of 0.[3] Hence this particular result shows that $\hat{\beta}$ and the OLS residuals u are independent as they are jointly normal with zero covariance.

Question 3: Probability distributions

(a) Since we don't need to show that linear combinations of X are also normally distributed, this question literally boils down to finding the mean and variance of the expression $aX+b$. Doing so for the mean gives

$$\begin{aligned}\mathbb{E}(aX + b) &= a\mathbb{E}(X) + b \qquad \text{(Linearity of expectation)} \\ &= a\mu + b\end{aligned}$$

[3] Don't forget that M_X is symmetric so $M_X' = M_X$.

and for the variance

$$\begin{aligned}\text{Var}(aX+b) &= \text{Var}(aX) && \text{(Variance ignores constants)}\\ &= a^2\text{Var}(X)\\ &= a^2\sigma^2\end{aligned}$$

Therefore $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

(b) Just like before, we just need to calculate the mean and variance (but we need to be more careful with the variance). For the mean we have

$$\begin{aligned}\mathbb{E}(\bar{X}_n) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i)\\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_1) && \text{(Identical distributions)}\\ &= \frac{1}{n}\sum_{i=1}^{n}\mu\\ &= \frac{1}{n}n\mu\\ &= \mu\end{aligned}$$

For the variance, we cannot just say that $\text{Var}(A+B) = \text{Var}(A) + \text{Var}(B)$ since we must also acknowledge the covariance between A and B. However, we can invoke the fact that all of the X_i terms are i.i.d. to claim that all of the covariances are zero. Hence,

$$\begin{aligned}\text{Var}(\bar{X}_n) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\\ &= \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) && \text{(Independence)}\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 && \text{(Identical distributions)}\\ &= \frac{1}{n^2}n\sigma^2\\ &= \frac{\sigma^2}{n}\end{aligned}$$

Hence, we have $\bar{X}_n \sim (\mu, \sigma^2/n)$.

(c) The working for this is identical to part (a); we just need to use matrix rules for variance instead. For the mean we have

$$\mathbb{E}(Ay + b) = A\mathbb{E}(y) + b = A\mu + b$$

and for the variance

$$\text{Var}(Ay+b) = \text{Var}(Ay+b) = \text{Var}(Ay) = A\text{Var}(y)A' = A\Sigma A'$$

as variance ignores constants. Putting the two together implies that $Ay + b \sim N(A\mu + b, A\Sigma A')$ as required.

(d) For this one we use the definition of covariance:

$$\begin{aligned}
\text{cov}(Ay, By) &= \mathbb{E}[(Ay - \mathbb{E}(Ay))(By - \mathbb{E}(By))'] \\
&= \mathbb{E}[A(y - \mathbb{E}(y))(y'B' - \mathbb{E}(y)'B')] \quad \because (AB)' = B'A',\ (A+B)' = A' + B' \\
&= A\underbrace{\mathbb{E}[(y - \mathbb{E}(y))(y - \mathbb{E}(y))']}_{=\Sigma}B' \\
&= A\Sigma B'
\end{aligned}$$

as required. The steps above skip a few of the algebraic steps, so make sure you are comfortable with what is going on.

(e) Note that $u = (u_1, u_2, \ldots, u_N)'$, where each u_i is i.i.d. $N(0,1)$ by construction. So if we calculate $u'u$ we have

$$\begin{aligned}
u'u &= \begin{bmatrix} u_1 & u_2 & \ldots & u_N \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \\
&= u_1^2 + u_2^2 + \cdots + u_N^2 \\
&= \sum_{i=1}^{N} u_i^2
\end{aligned}$$

where each u_i^2 is a squared standard normal random variable. All of these expressions are independent of each other as the original u_i expressions are i.i.d. Hence in this case $u'u$ is a sum of N squared independent standard normals, so $u'u \sim \chi^2_N$ as required.

(f) Since A is symmetric, we can write

$$A = Q\Lambda Q'$$

but this by itself is not going to be enough, since we want a very specific expression for Q. In particular, we derived in Question 1 that the eigenvalues of a symmetric idempotent matrix are either zero or one. We are also aware that the matrix Q is made up out of the eigenvectors of the matrix and the positioning of the eigenvectors in Q determine where the eigenvalues appear in Λ.

Why is this important? The reason why this comes in useful is because we can *order* the eigenvectors in Q such that

- The first diagonal part of Λ consists of only the eigenvalues that equal 1
- The remainder of the diagonal consists of the eigenvalues that equal 0

How do we know how many ones appear in Λ? The reason for this is that $\text{rank}(A) = q$. If that isn't enough to convince you, then consider the following: Remember that the rank of an idempotent matrix is equal to its trace, so

$$\begin{aligned}
\text{rank}(A) &= \text{tr}(A) \\
&= \text{tr}(Q\Lambda Q') \\
&= \text{tr}(Q'Q\Lambda) && \because \text{tr}(AB) = \text{tr}(BA) \\
&= \text{tr}(\Lambda)
\end{aligned}$$

But $\text{rank}(A) = q$ so $\text{tr}(\Lambda) = q$. Since eigenvalues are either zero or one, there must be q ones in Λ. Appropriate placement of the eigenvectors means we can come up with

$$\Lambda = \begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix}$$

We will need this for later. For the time being, let $y = Q'u$ (we also need this for later, since we do need to show that something turns into a sum of squared standard normals at some stage!). Notice that

$$\mathbb{E}(y) = Q'\mathbb{E}(u) = 0$$

and

$$\text{Var}(y) = Q'\text{Var}(u)Q = Q'Q = I_N.$$

Hence $y \sim N(0, I_N)$ – that is: y is standard multivariate normal. Therefore,

$$\begin{aligned}
u'Au &= u'Q\Lambda Q'u \\
&= y'\Lambda y \\
&= y' \begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix} y \\
&= \sum_{i=1}^{q} y_i^2 \sim \chi_q^2
\end{aligned}$$

since each of the y_i^2 elements are independent of each other. Note that the summation goes up to q only because there are q ones in the diagonal of Λ.

(g) Begin by decomposing $\Omega = Q\Lambda Q'$ as in the previous questions. Since Ω is positive definite, it also follows that the eigenvalues of Ω are strictly positive. Hence, there exists a matrix $\Lambda^{1/2}$ such that $\Lambda^{1/2}\Lambda^{1/2} = \Lambda$. In actual fact the matrix itself behaves according to the

usual power rules *as long as* you do it with Λ.[4] Similarly, we have a matrix $\Lambda^{-1/2}$ such that

$$\Lambda^{-1/2}\Lambda^{-1/2} = \Lambda^{-1}.$$

From here we do two things. The first thing to note is that

$$\Omega^{-1} = (Q\Lambda Q')^{-1} = Q\Lambda^{-1}Q'$$

using the matrix rule $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$. Furthermore, let $y = \Lambda^{-1/2}Q'z$ and observe that

$$\begin{aligned} z'\Omega^{-1}z &= z'Q\Lambda^{-1}Q'z \\ &= z'Q\Lambda^{-1/2}\Lambda^{-1/2}Q'z \\ &= y'y \end{aligned}$$

So all we need to do is show that y is multivariate standard normal and then we are done. The mean is straightforward as

$$\mathbb{E}(y) = \Lambda^{-1/2}Q'\mathbb{E}(z) = 0$$

and as for the variance,

$$\begin{aligned} \text{Var}(y) &= \text{Var}(\Lambda^{-1/2}Q'z) \\ &= \Lambda^{-1/2}Q'\text{Var}(z)Q\Lambda^{-1/2} \\ &= \Lambda^{-1/2}Q'\Omega Q\Lambda^{-1/2} \\ &= \Lambda^{-1/2}Q'Q\Lambda Q'Q\Lambda^{-1/2} && \text{(Spectral decomposition)} \\ &= \Lambda^{-1/2}\Lambda\Lambda^{-1/2} \\ &= I_N. \end{aligned}$$

Hence, after a bit of algebraic work,

$$y \sim N(0, I_N)$$

and using what we derived in part (e), $y'y \sim \chi^2_N$.

Question 4: Convergence in Probability

Firstly, it helps if we write out what we want to show, which in this case is

$$\lim_{n\to\infty} \mathbf{P}(|S_n - \mu| \geq \varepsilon) = 0,$$

for any $\varepsilon > 0$. The proof proceeds as follows.

[4] If you are not convinced, we can derive it explicitly. Since Λ is a diagonal matrix this makes it straightforward, since $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n})$. Since Ω is p.d., all of those elements are positive and therefore the square roots exist. Implicitly it is also symmetric as it is a diagonal matrix.

Proof. Let $X_1, X_2, \ldots, X_n$ be i.i.d. RVs with finite expectation $\mathbb{E}(X_i) = \mu$ and finite variance $\text{Var}(X_i) = \sigma^2 < \infty$. Now, let $\varepsilon > 0$. The i.i.d. property of each X_i implies that

$$\begin{aligned}
\text{Var}(S_n) = \sigma^2_{S_n} &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(X_i) && \text{(Independence)} \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 && \text{(Identical distributions)} \\
&= \frac{n\sigma^2}{n^2} \\
&= \frac{1}{n}\sigma^2.
\end{aligned}$$

The reason why this is useful is because we can use Chebyshev's Inequality to write

$$\mathbf{P}(|S_n - \mu| \geq k\sigma_{S_n}) \leq \frac{1}{k^2} \implies \mathbf{P}\left(|S_n - \mu| \geq \frac{k}{n}\sigma^2\right) \leq \frac{1}{k^2},$$

where $k > 0$. If we let $k = \dfrac{\varepsilon n}{\sigma^2}$, we have

$$\mathbf{P}(|S_n - \mu| \geq \varepsilon) \leq \frac{\sigma^4}{n^2\varepsilon^2}.$$

Now, observe that[5]

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{P}(|S_n - \mu| \geq \varepsilon) &\leq \lim_{n\to\infty} \frac{1}{n^2}\frac{\sigma^4}{\varepsilon^2} \\
&= 0.
\end{aligned}$$

Hence, $S_n \xrightarrow{p} \mu$, as required.

Question 5: Convergence in Distribution

(a) For this, we need to obtain an expression for the limit of $n^{2/n}$ as $n \to \infty$. In particular, we can write

$$n = e^{\log(n)} \implies n^2 = e^{2\log(n)} \implies n^{2/n} = e^{2n^{-1}\log(n)}$$

and, noting that we have a ∞/∞ indeterminate form in the first line,

$$\begin{aligned}
\lim_{n\to\infty} \frac{\log(n)}{n} &= \lim_{n\to\infty} \frac{1/n}{1} && \text{(L'Hôpital's Rule)} \\
&= \lim_{n\to\infty} \frac{1}{n} = 0
\end{aligned}$$

[5]Implicitly, we are showing that the limit of $\mathbf{P}(|S_n - \mu| \geq \varepsilon)$ is less than or equal to zero. Since probabilities are always non-negative, the only possible value that would satisfy that condition is zero.

Therefore

$$\lim_{n\to\infty} n^{2/n} = \lim_{n\to\infty} e^{2n^{-1}\log(n)} = e^{2\lim_{n\to\infty} n^{-1}\log(n)} = e^0 = 1$$

since the exponential function is continuous. Now, for any $x \geq 0$,

$$\begin{aligned} F_{X_n}(t) &= (1 - e^{-n^{2/n}x}) \\ &\to (1 - e^{-x}) \qquad\qquad \text{as } n \to \infty \end{aligned}$$

which we recognize as the DF of an Exp(1) distribution on the RHS. Hence, we conclude that X_n does indeed converge in distribution, specifically that

$$X_n \xrightarrow{d} X \sim \text{Exp}(1).$$

(b) Using the fact that the CDF of a $U[a, b]$ distribution (with $a < b$ as usual) can be written as

$$F_X(x) = \frac{x-a}{b-a},$$

we have that

$$F_{X_n}(x) = \frac{x+\sqrt{n}}{n+\sqrt{n}}.$$

But for any fixed $x \in \mathbb{R}$, if we try to take the limit as $n \to \infty$ we end up with

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} \left(\underbrace{\frac{x}{n+\sqrt{n}}}_{\to 0} + \underbrace{\frac{\sqrt{n}}{n+\sqrt{n}}}_{\to 0} \right) = 0.$$

There is no convergence in distribution; the limiting function is not a CDF since it is always zero!

(c) To begin with, let's follow the hint and check the convergence of the right tails. In this case, we have

$$\begin{aligned} \mathbf{P}(X_n > x) &= \mathbf{P}(n \min\{U_1, \ldots, U_n\} > x) \\ &= \mathbf{P}\left(\min\{U_1, \ldots, U_n\} > \frac{x}{n}\right) \\ &= \mathbf{P}\left(\bigcap_{j=1}^{n} \left\{U_j > \frac{x}{n}\right\}\right) \end{aligned}$$

but since U_j, $j = 1, \ldots, n$ is iid $N(0, 1)$ then first, by independence we can split the term above into the product of individual probabilities. From identical distributions, all of the individual probabilities are the same so we have

$$\begin{aligned} \mathbf{P}\left(\bigcap_{j=1}^{n} \left\{U_j > \frac{x}{n}\right\}\right) &= \mathbf{P}\left(U_1 > \frac{x}{n}\right)^n \\ &= \mathbf{P}\left(U_1 \in \left(\frac{x}{n}, 1\right]\right)^n \qquad \text{since } U_1 \sim U(0,1) \text{ and } U_1 > \frac{x}{n} \\ &= \left(1 - \frac{x}{n}\right)^n \\ &\to e^{-x} \quad \text{as } n \to \infty \end{aligned}$$

which is the right tail of the CDF of a Exp(1) distribution, as required. Hence we conclude that $X_n \xrightarrow{d} X \sim \text{Exp}(1)$.

Question 6: The Monte Carlo method [R Exercise]

(a) These are the results that I get when I type the commands on the extras sheet:

```
> set.seed(42)
> rnorm(5)
[1]  1.3709584 -0.5646982  0.3631284  0.6328626  0.4042683
> set.seed(41)
> rnorm(5)
[1] -0.7943683  0.1972575  1.0017043  1.2888254  0.9057534
> set.seed(42)
> rnorm(5)
[1]  1.3709584 -0.5646982  0.3631284  0.6328626  0.4042683
```

These may change depending on your version of R, but the most important thing is: notice that the five random draws that are taken after `set.seed(42)` are the **same**. It seems weird that you're guaranteed to get draws that should be otherwise random, but the biggest implication of knowing this is that setting the seed allows you to easily replicate your work on anyone else's computer.

This is especially relevant for anything that involves random numbers: running the same code does not guarantee that you will get exactly the same numbers as someone else, since the numbers you get will rely on the random draws that you obtain. Setting the seed fixes these draws no matter where you run it from, so you'll always have the same results. Ultimately, it makes it much easier to follow someone else's simulation work!

(b) (i.) This command gives you a $N \times 1$ column vector of `NA` values, which can be replaced later. The role of this is to set up the matrix so that we can modify it later during the loop that we're about to run. In terms of what we're going to use it for, the matrix `beta1.hat` will store all of the $\hat{\beta}_1$ estimates that we get from the Monte Carlo simulations.

(ii.) One way you can think about a `for` loop is: R runs the code inside the loop `M` times **in a row**, but every time it runs it, the value of `i` changes.

- For example: the first time it runs, `i` is equal to 1.
- The second time it runs, `i` is equal to 2
- ...and so on.

To see where the value of this comes from (pun not intended), take a look at where it appears inside the loop. It appears in the command `beta1.hat[i] = OLS[2,1]`, which we'll talk about soon.

(iii.) The command here is a very rapid way of generating random draws according to the conditional mean of y_i given x_i. Since $y_i|x_i \sim N(\beta_0 + \beta_1 x_1, 1)$, this means that

$$\mathbb{E}(y_i|x_i) = \beta_0 + \beta_1 x_i = 0 + 2x_i,$$

which is a linear function. Taking this straight line as given, the values of i are then normally distributed *around that line*, with a standard deviation of `sigma` $= 1$. Note that if this standard deviation depends on x_i, you would have the problem of *heteroskedasticity*, or non-constant variance, which we'll cover later in the semester.

To get to the point, though: this command generates a $N \times 1$ vector of y_i values, where each y_i value has a (conditional) mean of $2x_i$ and a (conditional) standard deviation of 1.

(iv.) The `coeftest` command generates a matrix of regression statistics. Here is one such example:

```
> coeftest(lm(y~x1))

t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)
(Intercept) -0.39067    0.18017 -2.1683   0.03879 *
x1           2.13328    0.19991 10.6712 2.251e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The real value of `coeftest` comes from this specific part of the output above:

```
-0.39067    0.18017 -2.1683   0.03879
 2.13328    0.19991 10.6712 2.251e-11
```

Notice that this is a 2×4 matrix, with the benefit that we can pull out *individual* values by using standard R indexing. We can name this command `OLS`, which has been done inside the code.

(v.) Using the `coeftest` command from before, typing

```
OLS[2,1]
```

pulls out the number in row 2, column 1 of the matrix generated by `coeftest`. Now, it turns out that this is exactly the OLS estimate $\hat{\beta}_1$ that we need.

What we've done here is to write `beta1.hat[i] = OLS[2,1]`, which means:

- Take the OLS estimate $\hat{\beta}_1$
- Put it in element i of the vector `beta1.hat`.

Notice that because the loop goes over `i = 1:M`, eventually all of the elements in the original `beta1.hat` matrix will be filled with their respective OLS estimates.

From a heuristic point of view, what the loop does is:

- Draw $N = 30$ observations on y_i and x_i.
- Run a regression of y_i on x_i.
- Take out the coefficient estimate on x_1 and store it.
- Chuck out the data and draw another $N = 30$ observations.
- ...then continue to repeat the same steps 1000 times.

Eventually, you'll have a whole dataset of $\hat{\beta}_1$ values, which you can then create a histogram with (in fact, we're going to do that right now!).

(c) The code for part 1 conducts a Monte Carlo experiment using $M = 1000$ repetitions, with $N = 30$ observations per repetition. The classic OLS estimator is known to have the property of *consistency*, so our hope is that for increasing sample sizes, we would observe the *Law of Large Numbers* coming into effect:

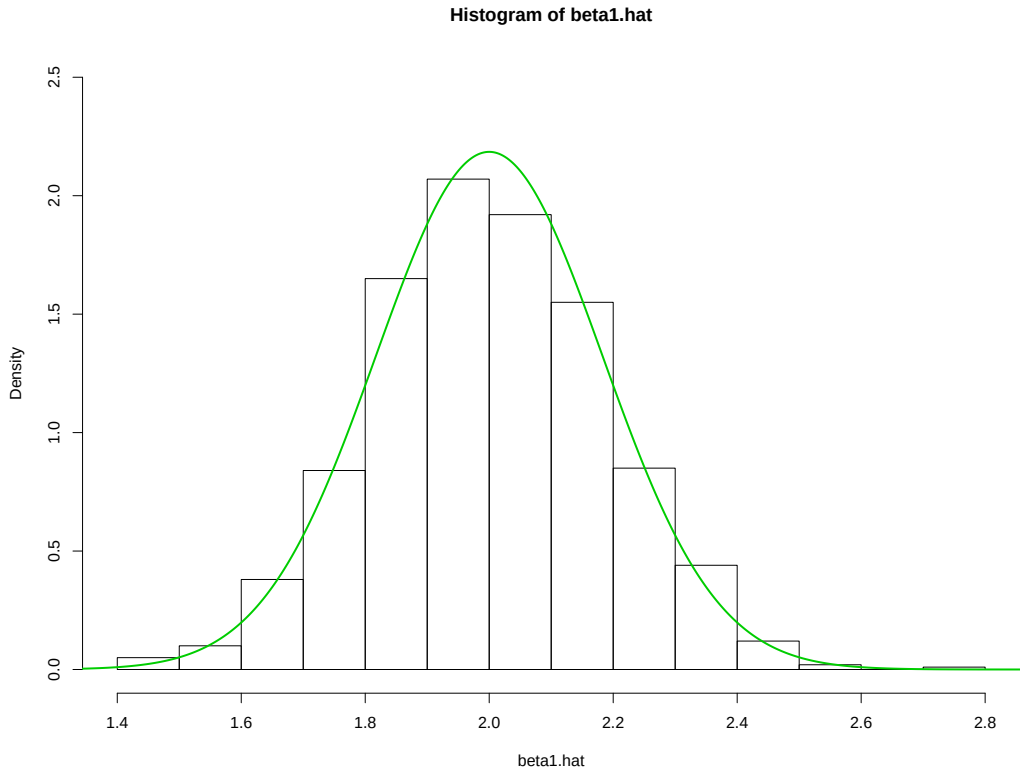
$$\frac{1}{M}\sum_{i=1}^{M}\hat{\beta}_{1,i} \xrightarrow{p} \beta_1 = 2.$$

The biggest problem with the statement given is the statement that $1.999 \neq 2$. While this is true mathematically, the nature of random numbers means that whenever we work with them, we're not guaranteed to actually *reach* a theoretical expected value using just data. The occurrence we're seeing here is known as **simulation error**; this would actually be considered close enough to the population value!

For these kinds of experiments, you will need to make your own judgement about, say, whether 1.999 is considered "far" from 2. If the estimated value was 1.5 for example, this could be considered to be "not close" to 2 and then such an argument might have merit. Ultimately, you have to think carefully about these problems!

(Anyone who is particularly astute may also notice, among others, that the statement is made on a single value of N, in this case 100. Considering the properties of the OLS estimator, it doesn't matter too much here. But for Monte Carlo experiments involving more complicated estimators, you might want to rely on a bit more than just one set of Monte Carlo simulations to illustrate your point!)

(d) The figure is provided below.



The estimated values of $\hat{\beta}_1$ and its variance are very close to what we would expect them to be. The histogram overlays very nicely with the true theoretical density, so subject to simulation error, one would likely conclude that this looks quite good!

(e) The code for part 2 returns the following output:

```
The average beta1 estimate for each of the sample sizes is:

 n=10  n=30  n=50 n=100
1.989 1.982 1.989 1.995

The variances of the sampling distribution of beta1.hat are:

 n=10  n=30  n=50 n=100
0.142 0.033 0.022 0.010

The theoretical variances of the sampling distributions are:

[1] 0.100 0.033 0.020 0.010
```

Observe that as we increase the sample size from $n = 10$ to $n = 100$, the average $\hat{\beta}_1$ remains very close to its true value of 2. Any differences in these cases can be attributed

to simulation error.[6] Consequently, the mean estimates line up with what we would expect.

The output also returns the estimated variances of $\hat{\beta}_1$ for each of our sample sizes, along with the theoretical variance of $\hat{\beta}_1$ associated with those sample sizes. Apart from the case where $n = 10$ (which isn't particularly surprising since $n = 10$ is generally considered an abysmal number of observations), the estimated variances look very close to their theoretical counterparts as the sample size increases. Based on the results of the Monte Carlo experiment, this would be consistent with the statement

$$\hat{\beta}_1 \xrightarrow{d} N\left(2, \frac{1}{n}\right).$$

[6] If you are in doubt, try changing the random seed and seeing what happens to these results!