ECOM40006/90013 Econometrics 3

Week 10 Extras (Part 2 Solutions)

Question 1: Some Likelihood Ratio Test Precursors

(a) The idea is that every time an approximation is taken around $\hat{\theta}_n$ (note that this is the centering point, not θ_0) the value of θ_n^* exists and lies between $\hat{\theta}_n$ and θ_0.

Why is this important? The approximation itself also depends on the number of observations n. Adding more observations updates the approximation and hence the values of θ_n^*. The idea here is that no matter how many times it updates, this value always sits between the *current* value of $\hat{\theta}_n$ and θ_0.

Now, θ_0 is a static expression and doesn't change. If we keep picking a value θ_n^* between $\hat{\theta}_n$ and θ_0 each time and $\hat{\theta}_n$ continues to converge in probability towards θ_0, the range of possible intermediate values also shrinks (in probability), implying that θ_n^* is also converging in probability towards θ_0.

Note that this is not a particularly rigorous argument in any way – the long story short is just that "distance between $\hat{\theta}_n$ and θ_0 narrows so θ_n^* gets closer to θ_0 in probability too."

This argument, when coupled with the Continuous Mapping Theorem, also suggests that the Hessian at θ_n^* converges towards the Hessian at θ_0, a result that we will find quite useful for the null distribution of the LR statistic.

(b) (i.) The reciprocal of all the diagonal elements of A gives a new matrix

$$A^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$$

and multiplying this matrix with A gives

$$\begin{aligned} A^{-1}A &= \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} \times 2 & 0 \\ 0 & \frac{1}{3} \times 3 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

which is the identity matrix, as required.

(ii.) In this case one has

$$\begin{aligned} A^{1/2}A^{1/2} &= \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{2} \times \sqrt{2} & 0 \\ 0 & \sqrt{3} \times \sqrt{3} \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \\ &= A \end{aligned}$$

as required.

(iii.) In a similar set of steps to before, we have

$$\begin{aligned} A^{-1/2}A^{-1/2} &= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} \times \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{3}} \times \frac{1}{\sqrt{3}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} \\ &= A^{-1} \end{aligned}$$

as required.

(iv.) Here, we have

$$\begin{aligned} A^{-1/2}A^{1/2} &= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} \times \sqrt{2} & 0 \\ 0 & \frac{1}{\sqrt{3}} \times \sqrt{3} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

which is the identity matrix, as required.

Aside. Note that all of these examples are quite simple, and are supposed to illustrate how these rules work for a simple 2×2 matrix. These results extend to positive definite matrices that are generally not always diagonal (covariance matrices being one such example), and much more difficult to show. The hope here is that by showing a simpler example where it works, you can be convinced that these results at least do hold *somewhere* and that one can extend it outwards.

Question 2: The Null Distribution for the Likelihood Ratio Statistic

(a) In order to invoke the weak law of large numbers, we'd at least need some sort of summation in there. Fortunately, that's where the definition of the log-likelihood comes in:

$$\begin{aligned}\frac{1}{n}\left(\frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}\right) &= \frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}\sum_{i=1}^{n}\log f(Y_i;\theta) \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log f(Y_i;\theta)}{\partial\theta\partial\theta'} \\ &\xrightarrow{p} \mathbb{E}\left(\frac{\partial^2 \log f(Y_i;\theta)}{\partial\theta\partial\theta'}\right),\end{aligned}$$

noting that derivatives can move inside and outside of summation operators, and with the last line appealing to the weak law of large numbers.[1]

(b) Don't worry too much about where the second-order expression comes from. The idea is that now that it's here, we want to work with it.

Let's first start by dealing with the (transposed) score, which is the expression

$$\left.\frac{\partial \log L_n(\theta)}{\partial\theta'}\right|_{\theta=\hat{\theta}_n}$$

which we claim is equal to zero. The reason for this is because this is the score (albeit it's transposed, but still the score), and one of the features of the score is that it is equal to zero at the MLE $\hat{\theta}_n$.[2] This allows us to reduce the Taylor approximation in the question down to

$$\begin{aligned}\log L_n(\theta_0) &= \log L_n(\hat{\theta}_n) + \frac{1}{2}(\theta_0-\hat{\theta}_n)'\left.\frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}\right|_{\theta=\theta_n^*}(\theta_0-\hat{\theta}_n) \\ \implies \log L_n(\theta_0) - \log L_n(\hat{\theta}_n) &= \frac{1}{2}(\hat{\theta}_n-\theta_0)'\left.\frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}\right|_{\theta=\theta_n^*}(\hat{\theta}_n-\theta_0),\end{aligned}$$

where $\log L_n(\hat{\theta}_n)$ was moved to the left hand side and a factor of (-1) was taken from each of the $(\theta_0-\hat{\theta}_n)$ expressions (so that we can reverse the signs in the brackets without changing the overall sign at the front).[3] If we then multiply by -2 on both sides we then

[1] Usually with these things there are also a bunch of other 'regularity conditions' that need to be assumed to guarantee that this works. For now we're just after the rough idea of it, and if you want to travel deeper down the rabbit hole of theory, you're likely to be at a level where the supplementals aren't that much use in that case!

[2] This is naturally assuming that the MLE is the same as the unrestricted estimator under the alternative hypothesis. For general applications this is usually the case. I wouldn't be surprised to hear that there are some contrived versions that exist where this is not true, but those would be some pretty niche situations to say the least.

[3] That is: generally one can write

$$(a-x)^2 = [(-1)(x-a)]^2 = (-1)^2(x-a)^2 = (x-a)^2.$$

In a quadratic expression you can reverse the signs in the brackets. For a second-order approximation like this one, the idea is no different.

have

$$\underbrace{2(\log L_n(\hat{\theta}_n) - \log L_n(\theta_0))}_{LR} = -(\hat{\theta}_n - \theta_0)' \frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\theta_n^*} (\hat{\theta}_n - \theta_0).$$

Notice that on the left hand side we now have the LR test statistic, so all that is left is to manipulate the right hand side. From here we can proceed to multiply the right hand side by n/n. For these steps I'll deliberately extend them just so it's really clear as to where things are going. Notice where the expressions go now:

$$\begin{aligned}
LR &= -(\hat{\theta}_n - \theta_0)' \frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\theta_n^*} (\hat{\theta}_n - \theta_0) \times \frac{n}{n} \\
&= (\hat{\theta}_n - \theta_0)' \left(- \frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\theta_n^*} \right) (\hat{\theta}_n - \theta_0) \times \frac{\sqrt{n}\sqrt{n}}{n} \\
&= \sqrt{n}(\hat{\theta}_n - \theta_0)' \left(- \frac{1}{n} \frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\theta_n^*} \right) \sqrt{n}(\hat{\theta}_n - \theta_0)
\end{aligned}$$

We can deal with each of these expressions on a term-by-term basis. We'll start by handling the biggest term first, which is the Hessian in the middle. Let's first define it as

$$\frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\theta_n^*} = H_n(\theta_n^*).$$

Now, in a similar fashion to what we did in part (a), observe that

$$\begin{aligned}
\frac{1}{n} H_n(\theta_n^*) &= \frac{1}{n} \left(\frac{\partial^2}{\partial\theta\partial\theta'} \sum_{i=1}^n \log f(Y_i;\theta)\bigg|_{\theta=\theta_n^*} \right) \\
&= \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^2 \log f(Y_i;\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\theta_n^*} \right) \\
&\xrightarrow{p} \mathbb{E} \left(\frac{\partial^2 \log f(Y_i;\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\theta_0} \right) \\
&= H(\theta_0) \\
&= -\mathscr{I}
\end{aligned}$$

where in this working, we invoked a combination of what we did in part (a) of this question and also part (a) of Question 1 – namely: that since $\theta_n^* \xrightarrow{p} \theta_0$, the average Hessian at θ_n^* would converge to the expected Hessian at θ_0. The last line then invokes the Information Equality.

Since the Fisher information matrix is positive definite, this means that the inverse $\mathscr{I}^{-1}$ admits a decomposition of the form

$$\mathscr{I} = \mathscr{I}^{1/2}\mathscr{I}^{1/2}.$$

Notice that on either side of this are terms that we know the asymptotic distribution of, namely

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \mathscr{I}^{-1}).$$

We can invoke Slutsky's Theorem on the LR test statistic. Restating it we begin with

$$\begin{aligned}
LR &= \underbrace{\sqrt{n}(\hat{\theta}_n - \theta_0)'}_{\xrightarrow{d} N(0,\mathscr{I}^{-1})'} \underbrace{\left(-\frac{1}{n}\frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\theta_n^*} \right)}_{\xrightarrow{p} -H(\theta_0) \;=\; \mathscr{I}^{-1}} \underbrace{\sqrt{n}(\hat{\theta}_n - \theta_0)'}_{\xrightarrow{d} N(0,\mathscr{I}^{-1})} \\
&= N(0,\mathscr{I}^{-1})'\mathscr{I}N(0,\mathscr{I}^{-1}) \\
&= \underbrace{N(0,\mathscr{I}^{-1})'\mathscr{I}^{1/2}}_{Z'} \underbrace{I^{1/2}N(0,\mathscr{I}^{-1})}_{Z} \\
&= Z'Z,
\end{aligned}$$

where we note that

$$\begin{aligned}
Z &= \mathscr{I}^{1/2}N(0,\mathscr{I}^{-1}) \\
&= N(0,\mathscr{I}^{1/2}\mathscr{I}^{-1}\mathscr{I}^{1/2}) \\
&= N(0,\mathscr{I}^{1/2}\mathscr{I}^{-1/2}\mathscr{I}^{-1/2}\mathscr{I}^{1/2}) \\
&= N(0, I_p).
\end{aligned}$$

As an aside: while we have done this a number of times before, just in case you're not comfortable with moving expressions inside normal distribution notation, you can always get from the first to the second line by taking the variance directly: namely

$$\begin{aligned}
\text{Var}(Z) &= \text{Var}(\mathscr{I}^{1/2}N(0,\mathscr{I}^{-1})) \\
&= \mathscr{I}^{1/2}\text{Var}(N(0,\mathscr{I}^{-1})\mathscr{I}^{1/2} \\
&= \mathscr{I}^{1/2}\mathscr{I}^{-1}\mathscr{I}^{1/2},
\end{aligned}$$

which is what we see in the working above. Note that in all of this, we use the rule for variance of a random vector: namely

$$\text{Var}(AX) = A\text{Var}(X)A'.$$

You can also calculate that Z has zero mean as well:

$$\mathbb{E}(Z) = \mathbb{E}(\mathscr{I}^{1/2}N(0,\mathscr{I}^{-1})) = \mathscr{I}^{1/2} \times 0 = 0.$$

Be sure to get some practice at moving things in and out of normal distributions if you're not comfortable with them! They're quite useful to know.

In any case, we now have that $Z \sim N(0, I_p)$. That is: the random vector Z is multivariate standard normal, or that each of the p elements of Z are distributed as $N(0, 1)$ and are independent of all the other elements of Z due to having zero covariance.[4] Finally, note that if we write out

$$Z = \begin{bmatrix} z_1 & z_2 & \dots & z_p \end{bmatrix}'$$

then

$$\begin{aligned} Z'Z &= \begin{bmatrix} z_1 & z_2 & \dots & z_p \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{bmatrix} \\ &= z_1^2 + z_2^2 + \cdots + z_p^2 \\ &= \sum_{i=1}^{n} z_i^2 \end{aligned}$$

which is a sum of independent squared standard normals. This fits the definition of a chi-squared distribution with p degrees of freedom, so we can conclude the derivation and note that under the null hypothesis,

$$LR \xrightarrow{d} \chi^2_p.$$

[4] Generally zero covariance does not imply independence, with one major exception being the multivariate normal distribution, where this is actually true.