ECOM40006/90013 Econometrics 3

Week 10 Extras (Part 2)

Question 1: Some Likelihood Ratio Test Precursors

The derivation of the null distribution of the Likelihood Ratio (LR) statistic can be quite involved, and makes use of some mathematical properties that may be difficult to get one's head around. This question aims to set up some of those properties that we use in the derivation that you see on lectures.

(a) Suppose that the MLE $\hat{\theta}_n$ is consistent for the true value θ_0. Approximations of functions involving the parameter vector θ sometimes rely on results such as the Mean Value Theorem, whereby an intermediate value θ_n^* between $\hat{\theta}_n$ and θ_0 is used as part of the approximation process.

Argue, informally, that if $\hat{\theta}_n \xrightarrow{p} \theta_0$ then $\theta_n^* \xrightarrow{p} \theta_0$ also, for any θ_n^* between $\hat{\theta}_n$ and θ_0. What would this suggest about the Hessian evaluated at $\theta = \theta_n^*$?

(b) Positive definite matrices have some useful features, although they can be a bit unintuitive at first glance. One such feature is that these matrices can be broken down into "matrix square roots" – e.g. a positive definite matrix A may have another matrix $A^{1/2}$ such that $A^{1/2}A^{1/2} = A$.

We can illustrate this property using diagonal matrices where all the diagonal elements are strictly positive. To see what those characteristics are, consider the following diagonal matrix:

$$A = \text{diag}(2, 3) = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Let's proceed with the following quick calculations to verify the following results:

(i.) The inverse matrix A^{-1} is calculated by taking the reciprocal of all the diagonal elements of A. Illustrate this example using the matrix A by showing that $A^{-1}A = I$.[1]

(ii.) Diagonal matrices with strictly positive diagonal elements also admit "square roots" in the sense that there exists a matrix $A^{1/2}$ such that

$$A^{1/2}A^{1/2} = A.$$

[1]You could also use the matrix inverse formula for a 2×2 matrix – but in my view this way's faster.

For diagonal matrices A, $A^{1/2}$ is explicitly given as the square root of all diagonal elements. In this example this would be

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

Verify that for this example, $A^{1/2}A^{1/2} = A$.

(iii.) Similarly, one can also find a matrix $A^{-1/2}$ such that $A^{-1/2}A^{-1/2} = A^{-1}$. For diagonal matrices this is another diagonal matrix where every element is the reciprocal square root. In the case of A above this would be

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

Verify that for this matrix A, one has $A^{-1/2}A^{-1/2} = A^{-1}$.

(iv.) Lastly, we can also combine the properties above. Using the matrix A above, illustrate the property that

$$A^{-1/2}A^{1/2} = I.$$

Question 2: The Null Distribution for the Likelihood Ratio Statistic

Consider the standard set of hypotheses concerning the $p \times 1$ vector of parameters θ_0:

$$\begin{aligned} H_0 &: \theta_0 = \theta_H \\ H_1 &: \theta_0 \neq \theta_H. \end{aligned}$$

Suppose that the statistic used to evaluate this hypothesis is the *Likelihood Ratio* (LR) test statistic

$$LR = 2\left[\log L_n(\hat{\theta}_n) - \log L_n(\theta_H)\right]$$

and that we were interested in figuring out what the distribution of this test statistic is under the null hypothesis.

To do so, we'll need some information to keep in mind. One such piece of information is the asymptotic distribution of the MLE:

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

Keep in mind that we also have the *Information Equality* available:

$$H(\theta_0) = \mathbb{E}\left(\left.\frac{\partial^2 \log f(Y_i;\theta)}{\partial\theta\partial\theta'}\right|_{\theta=\theta_0} \right) = -\mathscr{I}^{-1}.$$

From here we can begin the derivation of the null distribution of the LR statistic.

(a) Let's first get an intermediate result out of the way. Namely: with appeal to the weak law of large numbers, show that

$$\frac{1}{n}\left(\frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}\right) \xrightarrow{p} \mathbb{E}\left(\frac{\partial^2 \log f(Y_i;\theta)}{\partial\theta\partial\theta'}\right).$$

(b) Consider the log-likelihood under the null, i.e. that $\theta_0 = \theta_H$. A second-order Taylor approximation with a Lagrange form of the remainder around $\theta = \hat{\theta}_n$ (note where the approximation is being centered around) gives

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

Show that this expression converges in distribution to a chi-squared distribution with p degrees of freedom i.e. $LR \xrightarrow{d} \chi^2_p$ under the null.

This expression doesn't look great to begin with, but we can simplify it down.

- First, you can cancel out one of the terms by appealing to the solution of maximizing the log-likelihood. The question then is: which term would it be...?
- Try taking out a common factor of (-1) from the $\hat{\theta}_n - \theta_0$ expressions in the second-order part of the approximation.
- A multiplication by one trick would work, specifically n/n on the second-order part. The n on top can be written as $\sqrt{n}\sqrt{n}$ and then sent to some of the other expressions...
- The second-order partial expression is a Hessian at θ_n^*. Surely with the dot point above, that would converge to something in probability as $n \to \infty$...
- The information matrix $\mathscr{I}$ is positive definite. Perhaps this means the results from Question 1 can be used to great effect...
- A chi-squared distribution with p degrees of freedom is given as the sum of p squared independent standard normals, i.e.

$$\sum_{i=1}^{p} z_i^2 \sim \chi^2_p,$$

  where each $z_i \sim N(0,1)$ and is independent from the other z_i expressions.